



A Psychometric Approach to Understanding Professionalism

Thesis submitted in accordance with the requirements of the University of Liverpool for
the degree of Doctor in Philosophy by Katie Denise Cunliffe

May 2020

Declaration

No portion of this work has been submitted in support of any other application for degree or qualification at this or any other university or institute of learning.

Acknowledgements

Huge and heartfelt thanks to my personal Sunshine Band, Greg and Barney, who kept me grounded and safe throughout this process, reminding me of the important things in life. I dedicate this thesis to you both with love and gratitude; you stopped me feeling alone in isolation and weathered the chronic insomnia and omnipresent rumination truly graciously. You're my most favourite people.

There were so many other steadying hands that played parts too small for them to remember, but too large for me to forget, and a few require special mention:

My Mum, who taught me about softness in strength, kindness in combat.

My friend JP, who took the time to really understand the view from inside my own head.

My sister of shared experience, Iliana Makri, united in trauma.

I also extend eternal gratitude to Dr Emma Nelson, without whom this journey would never have started, and Dr Alex Forsythe and Prof Luke Dawson, without whom it would never have ended – each more than a supervisor, a saviour.

For some my success came with their support, others despite their best efforts, but I learned from every encounter and profited from every contribution along the way.

Contents

DECLARATION	II
ACKNOWLEDGEMENTS	III
CONTENTS.....	IV
LIST OF TABLES AND FIGURES	IX
ABSTRACT	XII
PART I: INTRODUCTION AND LITERATURE REVIEW	1
CHAPTER 1: A REVIEW OF THE LITERATURE RELATING TO PROFESSIONALISM	2
1.1 <i>Historical and Political Issues in the Understanding and Practice of Professionalism.....</i>	3
1.1.1 Phase one: the rise and partial fall of the sociological approach.....	4
1.1.2 Phase two: the prioritisation of education.....	7
1.1.3 Phase three: the two new professionalisms.	8
1.1.4 The contemporary picture.	10
1.2 <i>Defining Professionalism</i>	11
1.3 <i>Theories of Professionalism and Professional Development</i>	15
1.3.1 Theories of professional identity and socialisation.....	16
1.3.2 Educational theory.	18
1.3.2.1 Miller’s learning pyramid.	19
1.3.2.2 Kolb’s experiential learning theory.....	20
1.3.2.3 Service-learning.	23
1.3.2.4 Weaknesses of professionalism as an educational competency.	25
1.3.3 Psychological theory.....	27
1.3.3.1 Social cognitive theory.	28
1.3.3.2 Cognitive affective personality system.	31
1.3.3.2.1 Cognitive-affective units.	32
1.3.3.2.2 An integrated theory of personality and cognition.	35
1.3.3.3 Emotional intelligence.....	37
1.3.3.4 Theories of action and values.	41
1.3.4 Conclusions regarding theories of professionalism.	44
1.4 <i>Measuring Professionalism.....</i>	44
1.4.1 Contemporary practice in the educational assessment of professionalism.....	45
1.4.1.1 Observer ratings.	45
1.4.1.2 Paper-based tests.....	47
1.4.1.3 Self-assessment.	48
1.4.1.4 Critical incident data.	48
1.4.1.5 Psychological measurement tools.	49
1.4.2 The role of context.	51
1.4.3 Conclusions regarding the measurement of professionalism.	52
1.5 <i>The Literature Relating to Professionalism: Concluding Remarks</i>	53
1.6 <i>Chapter Summary.....</i>	53
CHAPTER 2: A REVIEW OF THE PSYCHOMETRIC LITERATURE	55
2.1 <i>Operationism.....</i>	56
2.2 <i>Classical Test Theory and its Core Principles.....</i>	58
2.2.1 Validity.	59
2.2.1.1 Content validity.	61
2.2.1.2 Dimensionality and factor structure.....	63
2.2.1.3 Criterion validity.	66
2.2.1.4 Construct validity.....	69
2.2.2 Reliability.	72
2.2.3 Critique of classical test theory.....	75
2.2.4 Summary of classical test theory.	77

2.3 <i>Item Response Theory</i>	77
2.3.1 A comparison of classical test theory and item response theory.	79
2.4 <i>Generalisability Theory</i>	81
2.5 <i>The Ethics of Validation</i>	83
2.6 <i>Conclusions Regarding Psychometric Best Practice</i>	84
2.7 <i>Chapter Summary</i>	85
CHAPTER 3: RESEARCH AIMS AND OBJECTIVES.....	87
3.1 <i>Aims and Objectives</i>	88
3.2 <i>Thesis Structure</i>	89
3.3 <i>Ethical Issues</i>	90
3.3.1 Interviews and focus groups.	91
3.3.2 Online data collection.	91
3.3.3 Participant information and consent.....	92
3.3.4 Data management.....	92
PART II: THE METHODOLOGICAL APPROACH.....	94
CHAPTER 4: METHODOLOGY	95
4.1 <i>Mixed Methods Research</i>	95
4.1.1 Philosophy and epistemology.....	96
4.1.2 Advantages of MMR.....	99
4.1.3 Controversies.....	101
4.1.4 Practical considerations.....	102
4.1.5 The approach of this thesis.....	103
4.2 <i>Theories, Methodologies, and Methods</i>	103
4.2.1 The lexical hypothesis.....	103
4.2.1.1 Contribution and critique.....	104
4.2.1.2 Implications for this thesis.....	107
4.2.2 Personal construct theory and associated methods.....	107
4.2.2.1 Contribution.....	110
4.2.2.2 Controversy, criticism, and critique.....	110
4.2.2.3 The repertory grid technique.....	113
4.2.2.3.1 Terminology & process.....	113
4.2.2.3.2 Analytical techniques.....	116
4.2.2.3.3 Strengths and applications.....	118
4.2.2.3.4 Criticism.....	120
4.2.2.4 Implications of PCT and the RGT for the study of professionalism.....	121
4.2.3 Q methodology.....	121
4.2.3.1 Historical issues and the statistical underpinnings of Q methodology.....	123
4.2.3.2 The theory of operant subjectivity.....	127
4.2.3.3 Concourse theory.....	128
4.2.3.4 The modern context: pragmatism and mixed methods research.....	128
4.2.3.5 The Q sort.....	130
4.2.3.5.1 Q sort terminology and procedure.....	130
4.2.3.5.2 The roles of triangulation and reflexivity in intersubjectivity.....	142
4.2.3.6 Strengths and applications.....	144
4.2.3.7 Criticism and controversies.....	145
4.2.3.8 Implications of Q methodology for the study of professionalism.....	149
4.2.4 Miscellaneous methods.....	150
4.2.4.1 Systematic and methodological reviews.....	150
4.2.4.1.1 The approach of this thesis.....	154
4.2.4.2 Consulting stakeholders.....	154
4.2.4.2.1 Member checking.....	155
4.2.4.2.2 Stakeholder consultation.....	156
4.2.4.2.3 Interacting focus groups.....	157
4.2.4.2.4 The nominal group technique.....	161
4.2.4.2.5 The approach of this thesis.....	167

4.3 Chapter Summary.....	167
PART III: EMPIRICAL RESEARCH.....	169
CHAPTER 5: THE PERENNIAL PROBLEM OF MEASURING PROFESSIONALISM: A METHODOLOGICAL REVIEW.....	170
ABSTRACT	170
5.1 Introduction.....	171
5.1.1 The case for effective measurement	171
5.1.2 The case for a methodological review	172
5.2 Methods.....	174
5.2.1 Search strategy.....	174
5.2.2 Inclusion and exclusion criteria.....	175
5.2.3 Screening and data extraction procedure	175
5.2.4 Materials.....	175
5.2.5 Analysis.....	183
5.3 Results.....	183
5.3.1 Psychometric quality.....	192
5.3.1.1 Measure rationale.....	192
5.3.1.2 Measure materials and documentation.....	193
5.3.1.3 Norms.....	193
5.3.1.4 Reproducibility.....	193
5.3.1.5 Validity.....	194
5.3.1.6 Floor/ceiling effects.....	195
5.4 Discussion.....	199
5.4.1 Psychometric quality.....	199
5.4.1.1 Measure rationale.....	199
5.4.1.2 Measure materials and documentation.....	200
5.4.1.3 Norms.....	200
5.4.1.4 Reproducibility.....	201
5.4.1.5 Floor/ceiling effects.....	203
5.4.1.6 Validity.....	204
5.4.1.7 Ethical considerations.....	205
5.4.2 The context and relevance of findings.....	205
5.4.3 Recommendations for practice and further research.....	206
5.4.4 Strengths and limitations of this review.....	207
5.5 Chapter Summary and Conclusions.....	208
CHAPTER 6: AN EXPLORATION OF SHARED CONSTRUING IN JUDGEMENTS OF PROFESSIONALISM USING THE REPERTORY GRID TECHNIQUE	210
ABSTRACT	210
6.1 Introduction.....	211
6.2 Methods.....	214
6.2.1 Materials and the repertory grid technique.....	214
6.2.2 Participants.....	216
6.2.3 Procedure.....	216
6.2.3.1 Pilot study.....	216
6.2.3.2 Main study.....	220
6.2.4 Analyses.....	221
6.2.4.1 Multiple groups factor analysis.....	221
6.2.4.2 Cluster analysis.....	222
6.3 Results.....	223
6.3.1 Multiple groups factor analysis.....	224
6.3.2 Cluster analysis.....	227
6.4 Discussion.....	228
6.4.1 Principal component 1: violation of baseline expectations.....	228
6.4.2 Principal component 2: logical ruthlessness.....	230
6.4.3 Principal component 3: personal likeability.....	233
6.4.4 Cluster analysis: proof of concept	235

6.4.5 Study evaluation.....	237
6.4.6 Future Directions.....	240
6.5 Chapter Summary and Conclusions.....	242
CHAPTER 7: SHARED SUBJECTIVITY IN UNDERSTANDING PROFESSIONALISM: A Q METHODOLOGICAL STUDY.....	243
ABSTRACT	244
7.1 Introduction.....	245
7.2 Methods.....	248
7.2.1 Materials.....	248
7.2.1.1 The Q sort.....	248
7.2.1.2 The current study.....	249
7.2.1.2.1 Identifying the concourse.....	250
7.2.1.2.2 Developing the Q set.....	250
7.2.2 Participants.....	253
7.2.3 Procedure.....	256
7.2.3.1 Q sort interviews.....	256
7.2.3.2 Electronic Q sort.....	258
7.2.4 Analyses.....	259
7.2.5 Intersubjectivity.....	262
7.3 Results.....	263
7.3.1 Factor descriptions.....	264
7.3.1.1 Factor 1: intellectual professionalism.....	264
7.3.1.2 Factor 2: collaborative professionalism.....	269
7.3.1.3 Factor 3: personal professionalism.....	274
7.3.1.4 Factor 4: professional professionalism.....	278
7.4 Discussion.....	283
7.4.1 Comparing results with themes from the existing literature.....	283
7.4.1.1 Being a good person.....	283
7.4.1.2 Accountability.....	284
7.4.1.3 Effectiveness.....	285
7.4.1.4 Self-management and continuing professional development.....	286
7.4.2 Points of divergence from the existing literature.....	288
7.4.2.1 Colleague relationships.....	288
7.4.2.2 The remit of professionalism.....	288
7.4.2.3 The role of sector of employment.....	289
7.4.3 Study evaluation.....	291
7.4.4 Implications and future directions.....	294
7.5 Chapter Summary and Conclusions.....	295
CHAPTER 8: A STAKEHOLDER CONSULTATION REGARDING THE MODEL OF INTERPERSONAL PROFESSIONALISM USING THE NOMINAL GROUP TECHNIQUE.....	296
ABSTRACT	296
8.1 Introduction.....	297
8.1.1 The model of interpersonal professionalism.....	297
8.1.1.1 The cognitive affective personality system theory and theories of emotional intelligence.....	298
8.1.1.2 An integrated model of interpersonal professionalism.....	300
8.1.2 Research aims.....	306
8.2 Methods.....	307
8.2.1 Materials.....	307
8.2.2 Participants.....	307
8.2.3 Procedure.....	307
8.2.4 Analyses.....	309
8.3 Results.....	310
8.3.1 Votes and scores.....	317
8.3.2 Thematic analysis.....	320
8.3.2.1 Most prominent themes.....	322
8.4 Discussion.....	322
8.4.1 Implications for the model of interpersonal professionalism.....	324

8.4.2 Study evaluation	325
8.4.3 Future research directions	326
8.5 Chapter Summary and Conclusions	327
PART IV: OVERALL DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS	329
CHAPTER 9: GENERAL DISCUSSION, CONTRIBUTIONS, AND CONCLUSIONS	330
9.1 Contributions Relevant to Psychological Theory and Measurement	330
9.2 Methodological Insights	332
9.3 Evaluation of the Approach of this Thesis	333
9.4 Future Research Directions	335
9.5 Recommendations for the Teaching and Learning of Professionalism	336
9.6 Overall Conclusions	337
REFERENCES	339
APPENDICES	387
APPENDIX A: ETHICAL APPROVALS	388
APPENDIX B: REPERTORY GRID TECHNIQUE DEFINING ELEMENTS	392
APPENDIX C: REPERTORY GRID TECHNIQUE CLUSTER ANALYSIS RESULTS	397
APPENDIX D: Q SET DEVELOPMENT	401
APPENDIX E: HTMLQ CODING	406
APPENDIX F: Q FACTOR INTERPRETATION CRIB SHEETS	410
APPENDIX G: THE MODEL OF INTERPERSONAL PROFESSIONALISM	431

List of tables and figures

Tables

TABLE 1.1	38
<i>THE FOUR-BRANCH MODEL OF EMOTIONAL INTELLIGENCE</i>	38
TABLE 5.1	177
<i>QUALITY ASSESSMENT SCORING CRITERIA</i>	177
TABLE 5.2	185
<i>DESCRIPTION OF MEASURES INCLUDED IN FINAL REVIEW</i>	185
TABLE 5.3	196
<i>NUMBER OF PAPERS RECEIVING EACH SCORE FOR EACH QUALITY CRITERION</i>	196
TABLE 6.1	217
<i>ELEMENTS PROVIDED TO PARTICIPANTS DURING PILOT AND MAIN STUDY</i>	217
TABLE 6.2	219
<i>LIKERT-TYPE RATING SCALE AND GUIDANCE FOR PILOT AND MAIN STUDY</i>	219
TABLE 6.3	224
<i>SUMMARY CHARACTERISTICS OF PARTICIPANT SAMPLE (N=36)</i>	224
TABLE 6.4	225
<i>ROTATED FACTOR LOADINGS AND SUMMARY RESULTS OF MULTIPLE GROUPS FACTOR ANALYSIS (MGFA)</i>	225
TABLE 7.1	252
<i>THEMES IDENTIFIED WITHIN THE PROFESSIONALISM CONCOURSE</i>	252
TABLE 7.2	255
<i>SUMMARY CHARACTERISTICS OF PARTICIPANT SAMPLE (N=34) AND DISTRIBUTION OF DEFINING Q SORTS</i>	255
TABLE 7.3	264
<i>ROTATED Q SORT FACTOR LOADINGS ROUNDED TO TWO DECIMAL PLACES AND SECTOR OF CURRENT OR MOST RECENT EMPLOYMENT</i>	264
TABLE 7.4	266
<i>DESCRIPTION OF FACTOR 1: INTELLECTUAL PROFESSIONALISM</i>	266
TABLE 7.5	270
<i>LONG-FORM NARRATIVE DESCRIPTION OF FACTOR 2: COLLABORATIVE PROFESSIONALISM</i>	270
TABLE 7.6	275
<i>LONG-FORM NARRATIVE DESCRIPTION OF FACTOR 3: PERSONAL PROFESSIONALISM</i>	275
TABLE 7.7	279
<i>LONG-FORM NARRATIVE DESCRIPTION OF FACTOR 4: PROFESSIONAL PROFESSIONALISM</i>	279
TABLE 8.1	304

<i>DESCRIPTION OF THE STAGES OF A TWO-PERSON PROFESSIONAL ENCOUNTER BETWEEN A PRODUCER AND A RECEIVER.</i>	304
TABLE 8.2	308
<i>CORE PHASES OF THE NOMINAL GROUP TECHNIQUE.</i>	308
TABLE 8.3	311
<i>SUMMARY OF PARTICIPANT SAMPLE CHARACTERISTICS.</i>	311
TABLE 8.4	312
<i>SUMMARY OF GROUP CHARACTERISTICS.</i>	312
TABLE 8.5	313
<i>IDEAS GENERATED BY EACH GROUP.</i>	313
TABLE 8.6	318
<i>GROUP CONSENSUS REGARDING THE PRIORITISATION OF IDEAS.</i>	318
TABLE 8.7	321
<i>THEMES AND SUBTHEMES IDENTIFIED.</i>	321

Figures

FIGURE 4.1. <i>EXAMPLE COMPLETED REPERTORY GRID USING OBJECT ELEMENTS AND RATING FOUR BIPOLAR CONSTRUCTS USING A FIVE-POINT LIKERT-TYPE RATING SCALE.</i>	115
FIGURE 4.2. <i>EXAMPLE INVERTED QUASI-NORMAL DISTRIBUTION Q SORT TEMPLATE.</i>	134
FIGURE 5.1. <i>SUMMARY OF REVIEW PROCESS.</i>	184
FIGURE 6.1. <i>BLANK REPERTORY GRID USED IN THE MAIN STUDY.</i>	215
FIGURE 6.2. <i>DENDROGRAM OF CLUSTER SEMANTIC CONTENT.</i>	227
FIGURE 6.3. <i>BIPLOT OF PC1 (VIOLATION OF BASELINE EXPECTATIONS) AND PC2 (LOGICAL RUTHLESSNESS) PARTIAL FACTOR LOADINGS.</i>	230
FIGURE 6.4. <i>BIPLOT OF PC2 (LOGICAL RUTHLESSNESS) AND PC1 (VIOLATION OF BASELINE EXPECTATIONS) PARTIAL FACTOR LOADINGS.</i>	231
FIGURE 6.5. <i>BIPLOT OF PC3 (PERSONAL LIKEABILITY) AND PC1 (VIOLATION OF BASELINE EXPECTATIONS) PARTIAL FACTOR LOADINGS.</i>	235
FIGURE 7.1. <i>QUASI-NORMALLY DISTRIBUTED Q SORT TEMPLATE.</i>	257
FIGURE 7.2. <i>EXAMPLE COMPLETED Q SORT CONFIGURATION.</i>	258
FIGURE 7.3. <i>FACTOR ARRAY FOR INTELLECTUAL PROFESSIONALISM.</i>	268
FIGURE 7.4. <i>FACTOR ARRAY FOR COLLABORATIVE PROFESSIONALISM.</i>	273
FIGURE 7.5. <i>FACTOR ARRAY FOR PERSONAL PROFESSIONALISM.</i>	277
FIGURE 7.6. <i>FACTOR ARRAY FOR PROFESSIONAL PROFESSIONALISM.</i>	282
FIGURE 8.1. <i>FACTORS AFFECTING THE PRODUCER AND RECEIVER'S INTERPERSONAL APPROACHES, IN THE CONTEXT OF INDIVIDUAL AFFECT AND THE EXTERNAL SITUATION.</i>	301

FIGURE 8.2. *THE STAGES OF A TWO-PERSON PROFESSIONAL ENCOUNTER BETWEEN A PRODUCER AND A RECEIVER.* **303**

Abstract

“All science rests upon and begins with accurate description and measurement”

Raymond Cattell (1965, p. 12)

This thesis explored professionalism from the perspective of psychometry. Being able to effectively measure professionalism is increasingly important for educators and employers among claims of a ‘crisis of professionalism’. Therefore, this thesis set out to explore how professionalism might be measured. The starting point was to review methodology associated with measuring professionalism as a psychological construct against best-practice criteria. The data suggested that current psychometric practice was poor, with no extant measures recommended for use in education or employment contexts.

These findings highlighted a need to identify priorities for increasing the quality of the psychological measurement of professionalism. A methodological review was undertaken with the results suggesting that the first priority was to create a theoretically grounded construct of professionalism. This priority was achieved via two empirical studies. Professionalism was explored as an ideographically construed concept, with elicited data suggesting that there were three underlying dimensions relevant to the shared construing among participants. These data challenged existing conceptions within the literature by suggesting that everyday understanding and experiences of professionalism may differ from the theories individuals and experts express about it.

The second psychological study explored subjectivity in perceptions of professionalism using Q methodology. Data suggested that subjectivity was shared along four dimensions by the participant group and that the different views of professionalism held by participants may, at times, actively contradicted one another. Crucially, data suggested that where the aim is to increase perceptions of professionalism, defining it as a norm-referenced concept may be of limited utility. Recommendations were instead made to further explore professionalism as a subjectively perceived construct, rather than one objectively defined.

The final aim was to contribute psychological theory to the discussion of professionalism. The previous findings were combined with existing theory to

propose an interpersonal model of professionalism, which was presented to stakeholders for consultation. Stakeholders commented that the current iteration of the model did not meet the needs of a construct definition for psychological measurement, but that it may stimulate further study and provide a useful basis for targeting education interventions seeking to improve professionalism.

Overall, it was concluded that comprehensively understanding professionalism from a psychological perspective requires further study. However, the psychological approach offers potential avenues for this to progress, both in general understanding and psychometric applications. It is recommended that educating institutions and regulatory bodies work closely with psychological research teams, with both parties making long-term commitments to collaboration. Furthermore, the data in this thesis suggests potential areas where educational innovation may be effective in improving perceived professionalism and complementary directions for future research, including gathering data sufficient to enable generalisable conclusions.

Part I: Introduction and Literature Review

Chapter 1: A Review of the Literature Relating to Professionalism

The concept of professionalism is charged with controversy in political, academic, and occupational contexts (Buck, Holden & Szauter, 2015), with stakeholders today placing increasing importance on a so-called ‘professionalism crisis’ (Anderson, Taylor & Gahimer, 2014; Baernstein, Oeschlager & Chang, 2009; Birden et al., 2014; Carter et al., 2015; Goldie, 2013; Hammer, 2006; Hicks, Lin, Robertson, Robinson & Woodrow, 2001; Monrouxe & Rees, 2012; Papadakis, Hodgson, Teherani & Kohatsu, 2004; Papadakis et al., 2005; Stern & Papadakis, 2006; Zijlstra-Shaw, Robinson & Roberts, 2012). This crisis is characterised within healthcare sectors by a lack of respect for, and compassion displayed towards, patients (Hammer, 2006; Marei, Al-Eraky, Almasoud, Donkers & Van Merriënboer, 2018; Monrouxe & Rees, 2012; Stern, 2000), and is described in the engineering workforce as a lack of ‘oven-ready’ candidates capable of working collaboratively and applying their knowledge to real-world problems (The Institute of Engineering and Technology, 2017). Healthcare trainers and educators particularly have received criticism for prioritising technical and academic learning at the expense of professionalism-related skills such as interpersonal skills and problem-solving abilities (Karnieli-Miller, Vu, Holtman, Clyman & Inui, 2010; Kenny, Mann & MacLeod, 2003; Squier, 1990; Stern & Papadakis, 2006). Consequently, professionalism has become a key focus for education, training, and development activities across all employment sectors nationally and internationally (Baernstein et al., 2009; Evetts, 2014; Fochtman, 2006).

Research suggests that professionalism assessments during training predict technical and professional competence during later training years and long-term careers (Adam et al., 2015; Gill et al., 2015; Hammer, 2006; Lievens & Coetsier, 2002; Papadakis et al., 2004; Papadakis et al., 2005). During training, professionalism-related attributes such as conscientiousness, emotional intelligence, and non-defensiveness have been found to predict final-year exam success for medical trainees to the extent of differentiating the top twenty percent of performers from the bottom twenty percent (Adam, et al., 2015). This is particularly important within medical practice and other high-stakes sectors where decisions made within

the remit of individual professionals have substantial consequences (Bertolami, 2004; Brincat, 2006).

The globalisation of marketplaces from healthcare to high growth technology, has led to increasingly complex demands for professionalism (Hassall Thomsen, 2015; Hicks et al., 2001; Sudrajat, 2017; Yang & Taylor, 2014). However, the vast majority of research has been limited to the context of medical education, due equally to its standing as one of the oldest traditional professions, and the aforementioned high-stakes nature of clinical practice (Bertolami, 2004; Birden et al., 2014; Cruess, 2006; Evetts, 2006). Despite the medical sector having a long-standing relationship with professionalism research and education, claims remain that training activities in this area are ineffective (Bertolami, 2004; Monrouxe, 2010). Even with growing pressure on educators, employers, and individuals to deliver professionalism, little progress has been made in exploring the factors affecting it and how it might be cultivated (Evetts, 2006).

The following sections explore contemporary understanding of professionalism in light of its historical context. Although professionalism will be considered in its broadest terms, the majority of evidence cited will be grounded within the field of medical education and as defined by the sociological approach to understanding professionalism. However, relevant developments in educational and psychological theory will also be discussed. The chapter concludes by discussing the implications of historical issues and relevant theory for understanding and measuring professionalism, in the context of potential further research.

1.1 Historical and Political Issues in the Understanding and Practice of Professionalism

Professionalism has been understood in different ways throughout its history (Evetts, 2006). Until the mid-twentieth century, the concept was attached to a limited number of professions, particularly those requiring specialist skills and knowledge (Birden et al., 2014; Cruess & Cruess, 1997). The training for, and regulation and monitoring of, practice within these traditional professions was managed internally by their own members. The primary reason for this was that non-members did not

have sufficient knowledge or expertise to understand the technicalities of the profession and were therefore deemed unable to appreciate the complexity of issues relating to its practice (Freidson, 2001; Inui et al., 2006; Pfadenhauer, 2006; Shirley & Padgett, 2006). This traditional approach of internal professional regulation is known as self-regulation (Birden et al., 2014; Evans, 2008; McNeill, 2001; Monrouxe, 2010). Self-regulating professions traditionally involved such occupations as medicine, architecture, law, and the clergy, which relied upon more experienced practitioners to impart to trainees information and techniques gleaned from experience (Birden et al., 2014; Evetts, 2006). Professionals were responsible for internally determining arrangements and standards for access to, and the delivery of, training, and setting standards to be met for qualification and ongoing registration to practice. The professions themselves defined the roles to be undertaken and established procedures for remediation (Cruess, 2006). However, self-regulation as the basis of professionalism is viewed as of little relevance today, with contemporary understanding instead being shaped by several major historical shifts in its conceptualisation (Castallani & Hafferty, 2006; Cruess, 2006; Trathen & Gallagher, 2009). These major shifts are discussed in the following sections as phases in the history of professionalism.

1.1.1 Phase one: the rise and partial fall of the sociological approach.

Until around 1960, professionalism was largely understood collectively as organised systems, through the lens of sociology (Cruess & Cruess, 1997; Evetts, 2006). The sociological approach to understanding professionalism is based on the concept of a social contract negotiated between profession members and the general public (Birden et al., 2014; Cruess, 2006; Hammer, 2006; Shirley & Padgett, 2006; Stern, 2000). This contract affords profession members altered status within society, carrying perceived superiority and authority based on their being granted permission to undertake activities not permitted by non-members (Birden et al., 2014; Cruess & Cruess, 1997; Pfadenhauer, 2006; Stern, 2000; Svensson, 2006; Swick, 2000; Zijlstra-Shaw et al., 2012). Such permissions might include legally undertaking safe assessment, diagnosis, and treatment of physical health issues, or safely planning the construction of a building for public use.

These permissions applied where specialised skills and knowledge were required to effectively perform the required function and where lack of expertise could result in physical or spiritual risk to the public, such as through medical mistreatment, unsafe building design, or misapplying laws or scripture (Birden et al., 2014; Cruess, 2006; Cruess & Cruess, 1997; van Mook et al., 2009). In return for these permissions, profession members were required to accept and meet standards and responsibilities above and beyond those of the common individual, such as actively self-regulating their own profession, contributing to the maintenance and enhancement of associated specialised knowledge, and embodying characteristics such as integrity, compassion, and commitment to public service (Birden et al., 2014; Cruess, 2006; Cruess & Cruess, 1997; Freidson, 2001; Inui et al., 2006; Shirley & Padgett, 2006; Stern, 2000). The sociological approach therefore distinguishes profession members from non-members via the system affording them permission to operate in autonomy (Birden et al., 2014; Cruess & Cruess, 1997; Shirley & Padgett, 2006).

The sociological approach to understanding professionalism remained unchallenged until the mid-twentieth century (Cruess & Cruess, 1997). From the 1960's, public access to information and education improved in the UK, along with an increase in publicly funded services (Evetts, 2014; Shirley & Padgett, 2006). Where professions previously operated in a payable marketplace, there was a shift towards public funding, resulting in increased scrutiny of their practices (Castallani & Hafferty, 2006). As social equality became more important in society, the professions received increasing criticism, with professionalism coming to be viewed as a device promoting self-interested upward social mobility, elitism, and the retention of social status operating to the detriment of those accessing services (Birden et al., 2014; Cruess, 2006; Cruess & Cruess, 1997; Castallani & Hafferty, 2006; Evans, 2008; Evetts, 2003, 2006; Freidson, 1994; Shirley & Padgett, 2006; van Mook et al., 2009). The professions were accused of using self-regulation to replicate professional cultures as a means of maintaining social advantage, rather than to deliver the responsibilities agreed under their social contract (Hafferty & Franks, 1994; van Mook et al., 2009).

These accusations meant that professionals accustomed to the autonomy of self-regulation began to find their workplaces increasingly driven by transparency

and accountability, particularly where services were publicly funded such as within the National Health Service (Evans, 2008; Evetts, 2003, 2014; Hafferty, 1998). The powerbase shifted from profession members to a wider dispersion amongst funders, regulators, and the general public, as stakeholders of public services (Evans, 2008; Svensson, 2006). This was perceived by many, particularly advocates of the sociological model, to threaten professionalism by reducing the autonomy of professions and, as a result, their ability to act in the best interests of those without the knowledge to do so for themselves (Birden et al., 2014; Cruess & Cruess, 1997; Evetts, 2003; Pfadenhauer, 2006). This reduction of autonomy was argued to weaken the professional identity of members, as non-member stakeholders seized control (Evetts, 2014; Finn, Garner & Sawdon, 2010).

Views regarding professionalism continued to shift during the 1980s, with the term coming to be applied to occupations other than those traditionally associated with it (Castallani & Hafferty, 2006; Evans, 2008; Evetts, 2006; Freidson, 1994). Consequently, professionalism research became more widespread, such as in the fields of teaching and nursing, and common usage of the word saw it become a descriptor for individual conduct in a range of other occupations, from senior leadership and management roles, through sales and public relations, to those in the hospitality industry (Evans, 2008; Evetts, 2006; Svensson, 2006; Yang & Taylor, 2014). This increased public awareness of professionalism prompted research relating to its development but also provoked further claims that the traditional professions had been weakened by declining self-regulation. Critics argued that the term 'professionalism' was being used errantly, due to the ongoing failure of non-profession members to appreciate what advocates of the sociological approach viewed as its correct meaning (Brown & Bhugra, 2007).

Societal change during the last century led to attempts to reframe and redefine professionalism that have been welcomed by some but also met with resistance. Although the sociological approach is generally discredited in wider arenas as providing an inadequate description of professionalism within contemporary society, it remains vehemently adhered to by many operating in today's versions of the traditional professions and, specifically, the medical sector (Cruess, 2006; Inui et al., 2006; Shirley & Padgett, 2006). Adherents to this view have even defended their position using allegations of intentional revision of

professional histories perpetrated by the social sciences (Freidson, 1994). Some such commentators have come to define professionalism as relating specifically to occupations required to defend themselves against the perceived threats of de-professionalisation (Birden et al., 2014). Whether this is due to a genuine need to preserve cultural professional traditions to safeguard knowledge and practice, or to perpetuate the social status of its members, remains unclear due in part to the cyclical nature of the argument: where non-members criticise the sociological approach on any grounds, traditional profession members may always refute such criticism as resulting from a lack of knowledge and understanding of the profession sufficient to render the argument impotent (Inui et al., 2006; Shirley & Padgett, 2006).

1.1.2 Phase two: the prioritisation of education.

Disillusionment in the traditional approach to understanding professionalism and broader societal shifts meant that new ways of conceptualising professionalism were sought, with the educational approach gaining support during the 1990's (Cruess, 2006). Aligned with more general advances in educational practices, professionalism came to be viewed as a competency to be developed through direct teaching (Birden et al., 2014; Cruess & Cruess, 1997; Hendelman & Byszewski, 2014; Hicks et al., 2001; Marei et al., 2018; Monrouxe & Rees, 2012; O'Flynn, Power, Horgan & O'Tuathaigh, 2014). Professionalism became the responsibility of trainers and educators rather than being viewed as a product of self-regulation, with the assurance of professional competence to be assessed and verified to mark the point at which training was complete (Bonke, 2006). Coinciding with increased external regulation, the competency approach to understanding professionalism related more to principles of service enhancement than status enhancement, to ensure individuals were fit to deliver the role that they were contracted to fulfil (Evans, 2008; Hoyle, 2001). Enhanced scrutiny also meant that the general public played a greater role in external regulation, becoming consumers and exercising the power this afforded them (Evans, 2008; Sockett, 1996; Svensson, 2006). Service users came to be viewed as an important source of information regarding professionalism,

with its perception by others taking a priority that contravened the traditional approach of refuting the role of non-members in regulating professions.

This shift to viewing stakeholders in more equal terms to professionals was welcomed by critics of the sociological model, but over time also began to receive criticism (Evans, 2008). The educational approach was argued to support unjustified claims of authority over profession members by senior managers in ways ineffective at assuring professionalism (Evans, 2008; Troman, 1996). Evidence in favour of this criticism came from the observation that despite a partial shift from role modelling to the formal instruction of professionalism, little impact was observed on performance in the workplace (Bertolami, 2004; Hafferty & Franks, 1994; Hendelman & Byszewski, 2014; Monrouxe, 2010). This led to arguments that professionalism may not be a ‘trainable’ characteristic that can be mastered through traditional educational methods, rendering the educational approach moot (Bertolami, 2004).

As arguments that professionalism may not be trainable increased, the focus shifted to considering professionalism as latent potential, present in some individuals and not others. It was argued that this potential should be drawn out through development activities, rather than the subject of formal instruction. Although there was some continued emphasis on the role of educators, professionalism came to be viewed as an issue of character and ongoing conduct, rather than an endeavour that concluded as training ended (Bertolami, 2004).

1.1.3 Phase three: the two new professionalisms.

‘New professionalism’ has been discussed since the early 1990’s, but the term is used in two different ways (Evans, 2008; Evetts, 2006). Although emerging around the same time as the educational approach to understanding professionalism, new professionalism initially observed lower impact due to the practical utility afforded by existing educational approaches, and therefore began to gain traction from the early 2000’s but in two different directions (e.g. Evetts, 2003; Hoyle, 2001). In one direction, new professionalism was viewed as a holistic concept that refers to the entirety of individual conduct, rather than a checklist of distinct behaviours and attitudes (Carter et al., 2015; Goldie, 2013). It represents the full

range of behaviours exhibited by an individual at different times and in different contexts, but also the personal attributes on which these behaviours are based; their core values and beliefs.

This concept of new professionalism grew from criticism of the educational approach that treating professionalism as a competency to be mastered encouraged individuals to superficially ‘act’ professional in order to deliver expected outward behaviours (Bertolami, 2004; Evetts, 2003; Finn et al., 2010; Goldie, 2013; Wear, 2006). New professionalism took a more psychological stance, speaking to a gestalt humanism encouraging individuals to truly ‘be’ professional by centring their practice on the needs of the community they serve (Birden et al., 2014; Buck et al., 2015). Professionalism came to be viewed as more about the qualities and characteristics individuals bring to a role and less about the role itself or the occupation to which it belongs (Evans, 2008). This form of new professionalism recognises that conduct is contextually bound, affected by factors such as situations encountered and who they involve, and the level and type of support received from employers, and workplace cultures and expectations (Birden et al., 2014; Carter et al., 2015; Evans, 2008; Finn et al., 2010; Fochtmann, 2006; Goldie, 2013; Marei et al., 2018; Monrouxe, Rees & Hu, 2011; O’Flynn et al., 2014; Rosenthal et al., 2011; Shapiro, Rucker & Robitshek, 2006; Troman, 1996; van Mook et al., 2009; Verkerk, de Bree & Mourits, 2007; Zijlstra-Shaw et al., 2012). In brief, new professionalism recognises the importance of context in influencing specific manifestations of professionalism but asserts that the form of these manifestations is largely determined by fundamental personal qualities that transcend contextual boundaries. These internal and external forces vie for influence, but also do not preclude the effects of stable, trait-like individual characteristics in governing professionalism (Carter et al., 2015).

The second direction through which new professionalism is viewed relates to the perceived threats facing professions today and, specifically, a rebalancing of power within professionalism (Birden et al., 2014; Evans, 2008; Evetts, 2006; Pfadenhauer, 2006). This form of new professionalism argues that increased scrutiny has threatened autonomous self-regulation, reciprocally compounding and contributing to a loss of the faithful acceptance of the authority of professionals by the general public (Pfadenhauer, 2006). There are claims, particularly from those

lamenting the loss of the traditional sociological conceptualisation of professionalism, that this loss of authority and autonomy has weakened the purpose and superiority of professionalism. However, while the sociological approach suggests that the key consequence of external regulation is a loss of professional identity, new professionalism focuses on risks to the delivery of high-quality services (Birden et al., 2014; Cruess, 2006; Evans, 2008; Evetts, 2003, 2014; Pfadenhauer, 2006; van Mook et al., 2009).

The second concept of new professionalism suggests that it is outdated to consider a possible return to self-regulation and that instead professionals must turn their attention towards the integrity of their purpose and practice, marking a shift from the prioritisation of the professional to the service user (Evans, 2008; Hafferty, 1998; Hoyle, 2001; van Mook et al., 2009). Although this second concept of new professionalism may be interpreted as ending up in the same place as the first, the motivations behind it are quite different. While the first conceptualisation of new professionalism is grounded in humanism, the second finds its drivers in the restoration of power and status.

1.1.4 The contemporary picture.

The preceding sections describe three major lenses applied to professionalism over the years. Although these lenses have arisen broadly chronologically, it may be misleading to have referred to them as phases. The sociological, educational, and new professionalism approaches are all actually still promulgated as contemporary conceptions of professionalism, leading to ongoing confusion and debate as to what it actually is (Castallani & Hafferty, 2006; Shirley & Padgett, 2006). In the face of new challenges, clarifying the concept of professionalism becomes even more central. The dawn of social media has led to questions about online conduct and its role in professionalism (e.g. Rubin, 2018; Souza, de Lorena, Ferreira, Amorim & Peter, 2017; Wang, Wang, Zhang & Jiang, 2019), and increasing educational costs have led to student-as-consumer cultures whereby attainment is replaced by purchasing power, raising questions of the conflict between consumerism and professionalism (Carey & Ness, 2001; Usman, 2016). These challenges raise questions regarding the boundaries of professionalism and when it should and should

not apply, but also invites discussion as to whether professionalism is now coming to be viewed as a commodity to be purchased for self-gain in ways that actively contravene the qualities and attributes that it seeks to promote.

Overall, the modern context of professionalism seems to require a balance of all three approaches to understanding professionalism. Its intrinsic links to vocational training draw on the sociological understanding of professionalism, while the need to formalise education in ways enabling robust assessment derives from the educational approach, and the need to ensure congruence of values and attitudes relating to public service and accepting accountability echo new professionalism (Carey & Ness, 2001). What is clear from this, however, is that professionalism is an inescapably socially constructed concept, one that appears unlikely to ever be subject to a unitary, stable, or static understanding using currently employed research methods (Evans, 2008; Evetts, 2014; Troman, 1996). Professionalism is intrinsically linked to the temporal, cultural, and societal context within which it resides and may therefore best be considered as a concept residing with any real clarity only in the eye of the beholder, and targeted using methods specifically designed to retain such idiographic and divergent viewpoints, rather than those compressing them into false consensus.

1.2 Defining Professionalism

Despite long-standing interest in professionalism, academic and professional communities have struggled to find consensus as to its definition (Anderson et al., 2014; Birden et al., 2014; Blake & Gutierrez, 2011; Buck et al., 2015; Evans, 2008; Finn et al., 2010; Freidson, 1994; Goldie, 2013; Marei et al., 2018; Mazor et al., 2007; Monrouxe et al., 2011; O'Flynn et al., 2014; Trathen & Gallagher, 2009; Wilkinson, Wade & Knock, 2009). Professionalism is perceived differently by different individuals according to the temporal, situational, cultural, and organisational contexts within which they situate their focus (Finn et al., 2010; Goldie, 2013; Monrouxe et al., 2011). Some claim that professionalism may be recognised easily and consistently by any individual, whether expert or lay, but that problems arise when they try to articulate what they have recognised (Hammer,

2006; Monrouxe et al., 2011; Swick, 2000; Trathen & Gallagher, 2009; Wilkinson et al., 2009).

Notwithstanding the issue of articulation, there does appear to be a vaguely agreed core concept of professionalism (Anderson et al., 2014; Hafferty, 2006). Professionalism is generally agreed to be multi-dimensional (Birden et al., 2014; Carter et al., 2015; Li, Ding, Zhang, Liu & Wen, 2007), relating to the non-cognitive skills, abilities, or attributes known as ‘softer’ skills, as opposed to academic or technical skills (Ben-David, Snadden & Hesketh, 2004; Bonke, 2006), but this is where consensus ends. The use of words such as ‘skills’ and ‘abilities’ is controversial in itself, as advocates of new professionalism argue that professionalism is less about outwardly displayed competencies and more about who the person is (Buck et al., 2015). Having said this, a number of broad themes are persistently identifiable within the literature.

One of the strongest themes pertaining to professionalism is being a good person, although different sources describe this using different terminology. Relevant attributes describe positive personal character, including altruism and a commitment to work in the service of and for the benefit of others; ethical conduct; being respectful; having an empathetic, caring, and compassionate nature; being trustworthy and honest; and acting with the utmost integrity (Ben-David et al., 2004; Bertolami, 2004; Bonke, 2006; Buck et al., 2015; Carey & Ness, 2001; Carter et al., 2015; Cruess, 2006; Evetts, 2003; Fochtman, 2006; Frohna, 2006; Gleeson, 2007; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Karnieli-Miller et al., 2010; Kim, Kaplowitz & Johnston, 2004; Marei et al., 2018; Mazor et al., 2007; Monrouxe & Rees, 2012; Shirley & Padgett, 2006; Stern, 2000; Swick, 2000; Thomas et al., 2007; Trathen & Gallagher, 2009; Treviño, Weaver & Reynolds, 2006; van Mook et al., 2009; Zijlstra-Shaw et al., 2012). These concepts are echoed within the guidelines of professional regulatory bodies such as the UK General Dental Council (2013) and General Medical Council (2013, 2016), the UK Government’s Committee on Standards in Public Life (1995), Financial Conduct Authority (2019), Information Commissioner’s Office (n.d.), and the Solicitor’s Regulation Authority (2019), to name a few.

Another strong theme in defining professionalism is that an individual must be willing to be held accountable for their conduct and behaviour (Carey & Ness, 2001; Cruess, 2006; Evans, 2008; Evetts, 2003, 2014; Fochtmann, 2006; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Marei et al., 2018; Swick, 2000; van Mook et al., 2009; Zijlstra-Shaw et al., 2012). Individuals must also be effective within their role, delivering excellence as standard (Carter et al., 2015; Gleeson, 2007; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Marei et al., 2018; McNeill, 2001; Swick, 2000; van Mook et al., 2009). However, the latter requirement demonstrates lower consensus than the former. Advocates of sociological professionalism generally omit this criterion (e.g. Carey & Ness, 2001; du Toit, 1995; Finn et al., 2010; Monrouxe & Rees, 2012; Shirley & Padgett, 2006; Stern, 2000), perhaps reflecting the belief that only internal peers are sufficiently well-informed to pass judgement on such competency (Evans, 2008; Freidson, 2001; Inui et al., 2006; Shirley & Padgett, 2006). This omission is marked within the guidelines of some regulatory bodies, particularly those aligning the concept of professionalism with the sociological approach, such as medicine and dentistry (e.g. General Dental Council, 2013; General Medical Council, 2013, 2016). A final recurrent theme is that professionals must self-manage, particularly in approaching ongoing learning (Ben-David et al., 2004; Carey & Ness, 2001; Doukas, 2006; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Stern, 2000; Swick, 2000). This attribute features heavily in the literature of regulatory bodies taking traditional views of professionalism, such as the UK General Dental Council (2013) and General Medical Council (2013, 2016).

Although these themes are fairly consistent, consensus regarding professionalism fragments beyond the first criterion of demonstrating integrity of character. Moreover, after those attributes listed above, the defining features of professionalism largely become an idiographic list of words from which one may pick and choose according to requirement, rather than a coherent account of the concept. For example, some authors describe professionalism as being intrinsically related to communication (Anderson et al., 2014; Carey & Ness, 2001; Carter et al., 2015; Gleeson, 2007), appropriate dress and work habits (Anderson et al., 2014;

Carter et al., 2015; Gleeson, 2007; Hammer, 2006), teamwork (Ben-David et al., 2004; Karnieli-Miller et al., 2010), time management (Ben-David et al., 2004; Gleeson, 2007) and reliability (Carter et al., 2015), although non include them all. Once one proceeds beyond the vague and nebulous concept of good character, the detail of what this actually constitutes in practice garners much lower consensus. This has led to claims that attempts to define professionalism constitute mere lists of uncontentious words unlikely to provoke controversy or disagreement, but uninformative as a result (Birden et al., 2014).

Despite professionalism-related characteristics featuring in literature pertaining to a range of occupational sectors, discussion is largely limited to a core set of professions. Despite being expanded to apply within newer professions such as teaching and nursing, the most detailed and elaborated accounts of professionalism come from the field of medicine, both practice and education (Bertolami, 2004; Birden et al., 2014; Cruess, 2006; Evetts, 2006). This medical focus means that the majority of evidence-based definitions of professionalism are applicable only within this context, using clinical contexts and medical jargon to define its scope, therefore offering limited utility today in the assessment of professionalism as a broader issue. In addition, as a heavily regulated profession, accounts of medical professionalism are heavily influenced by regulator guidelines. Such guidelines are focused on the practical applications of a definition and are therefore closer to statements of professional consensus than definitions per se, with the consensus described generally reflecting differences based on the area of specialism (Birden et al., 2014; Wilkinson et al., 2009). These differences demonstrate that even within a field where professionalism has featured in teaching and research centrally for decades, it remains a concept without agreed definition.

Overall, professionalism is largely viewed today as something unique to each profession that claims it, with its detail being sufficiently specialised as to be recognisable only to members of that profession (Freidson, 2001; Inui et al., 2006; Pfadenhauer, 2006; Shirley & Padgett, 2006). This is incongruent with the way that the word 'professionalism' is used today, with every person, regardless of their occupational expertise, acknowledged as able to recognise it in others, regardless of their occupational role (Trathen & Gallagher, 2009; Wilkinson et al., 2009). This suggests that a more generalised definition of professionalism that transcends

occupational boundaries is overdue. In this context, the core themes of goodness of character, accountability, excellence, and ongoing learning appear generalisable, applying to a wide range of occupations, albeit to a greater or lesser extent depending on the job role and remit (Bertolami, 2004; Gleeson, 2007). This has led to some postulating that professionalism is a ‘meta-skill’ used to deploy the more specific or technical skills of an occupation efficiently and effectively with a more overall sense of professional conduct or purpose, to suit the situational context (Carter et al., 2015; Zijlstra-Shaw et al., 2012).

Despite this, widespread consensus remains evasive. The occupation-specific approach to defining professionalism appears to be widening the chasm between academic and expert-led definitions of professionalism, and those used on a daily basis in general conversation. The importance of this issue, however, is often overlooked; authors generally describe a lack of consensus before moving onto reporting conclusions in the absence of a clear definition of professionalism, rendering those conclusions questionable at best (e.g. Anderson et al., 2014; Buck et al., 2015; Marei et al., 2018; Mazor et al., 2007). The implications of overlooking the definition issue when exploring professionalism are limitations to the generalisability of the construct and specifically the ability to measure it in a meaningful way. In light of this, it would be appropriate to explore relevant theoretical accounts that may provide additional insight.

1.3 Theories of Professionalism and Professional Development

Theoretical accounts of professionalism are limited. The fragmented nature of the concept means that various fields have described it different ways according to their purpose and historical context. However, as discussed in section 1.1, historical shifts have applied different ‘lenses’ to the issue of professionalism that may be used to indicate relevant theory. The following sections will consider theory from sociology and psychology to explore their potential to account for professionalism and support attempts to define and measure it.

1.3.1 Theories of professional identity and socialisation.

Theories of professional identity and socialisation are the most dominant and enduring accounts of the development of professionalism (Cruess, 2006; Kenny et al., 2003). Grounded in the sociological understanding of professionalism described in section 1.1.1, this approach accounts for professionalism as a product of the professional system; upon entering that system, individuals passively receive a process imposed externally through which they become professional (Birden et al., 2014), a process of socialisation (Anderson et al., 2014; du Toit, 1995; Carey & Ness, 2001; Cruess, 2006; Evetts, 2003, 2006; Hafferty & Franks, 1994; Hammer, 2006). Individuals are exposed to the culture of more experienced professionals, gradually adjusting their professional identities to become congruent with the values, attitudes, and behaviours of that culture (Anderson et al., 2014; Cruess, 2006; Cruess & Cruess, 1997; du Toit, 1995; Evetts, 2003; Finn et al., 2010; Goldie, 2013; Hafferty & Franks, 1994; Hammer, 2006; Kenny et al., 2003; Monrouxe, 2010). As these values and norms are internalised and become part of individual professional identity, candidates become eligible for professional status (du Toit, 1995).

Although vague regarding the mechanisms of socialisation, workplace role models are considered the primary influence upon developing professionals (Anderson et al., 2014; Baernstein et al., 2009; du Toit, 1995; Finn et al., 2010; Hafferty & Franks, 1994; Kenny et al., 2003). When seeking to improve professionalism, therefore, this is achieved by simply exposing individuals to the professional culture to prompt acculturation (Anderson et al., 2014; Baernstein et al., 2009; Coulehan, 2006). Assuming this to be true, one would expect that the amount of time spent around more experienced professionals would predict steady increases in qualities associated with professionalism. This does not appear to be the case, however, as time spent in training courses with a strong vocational or placement component, such as medical training, is actually associated with declines in characteristics such as empathy and moral reasoning (Doukas, 2006; Frohna, 2006; Hafferty & Franks, 1994; O'Flynn et al., 2014; Rosenthal et al., 2011; Shapiro et al., 2006; Stern, 2000; Thomas et al., 2007). Similarly, it might be hypothesised that professions with the strongest professional identity such as medicine, would demonstrate the highest possible levels of professionalism. Once again, evidence suggests that this is not the case, with high profile lapses leading to concerns of a

professionalism crisis within medical practice (Anderson et al., 2014; Goldie, 2013; Hicks et al., 2001; Monrouxe & Rees, 2012; Zijlstra-Shaw et al., 2012).

However, these findings do not necessarily render theories of professional identity incorrect. An alternative explanation is found in the concept known as the hidden curriculum. The hidden curriculum refers to the experiences that individuals use to inform their learning and development outside of formal teaching curricula (Finn et al., 2010; Giroux & Penna, 1979; Hafferty, 1998; Karnieli-Miller et al., 2010; Stern & Papadakis, 2006; Wear, 2006). This informal curriculum may take the form of interactions with mentors outside of the formal learning environment, such as during vocational placements or within the workplace, or communications with social networks and peers, including online interaction, and is therefore explicitly encouraged to ensure socialisation as a means of professional development (Finn et al., 2010; Hafferty, 1998; Karnieli-Miller et al., 2010; O’Flynn et al., 2014; Stern & Papadakis, 2006; Wear, 2006). These interactions are the vehicle for ‘unwritten rules’ within a profession, including cultural norms, values, beliefs, and attitudes (Coulehan, 2006; Giroux & Penna, 1979; Hafferty, 1998; Karnieli-Miller et al., 2010; Monrouxe, 2010; O’Flynn et al., 2014; Stern & Papadakis, 2006). Through role modelling, the hidden curriculum contributes to the development of professional identity, and has even been argued to contribute the majority of socialisation-based learning for some students (Fochtman, 2006; Hafferty, 1998; Hafferty & Franks, 1994; Hendelman & Byszewski, 2014; Karnieli-Miller et al., 2010; O’Flynn et al., 2014; Rosenthal et al., 2011).

By its nature the hidden curriculum is ungoverned, so although it may be effective at cultural replication, the culture being replicated may not necessarily be in line with principles of professionalism (Hafferty & Franks, 1994). The hidden curriculum acknowledges that role models may fall short of required professional standards (Baernstein et al., 2009; Kenny et al., 2003; Monrouxe & Rees, 2012; Rosenthal et al., 2011), or offer messages directly contradicting the formal curriculum (Hafferty & Franks, 1994; Inui et al., 2006; Monrouxe & Rees, 2012; O’Flynn et al., 2014). Research suggests that students in professional training are frequently exposed to unprofessional behaviour and even misconduct perpetrated by professional mentors and role models (Baldwin, Daugherty & Rowley, 1998; Bursch et al., 2013; Finn et al., 2010; Hendelman & Byszewski, 2014; Heru, Gagne &

Strong, 2009; Hicks et al., 2001; Monrouxe & Rees, 2012; Rosenthal et al., 2011). Moreover, students report that peers are a third source of role modelling alongside staff on vocational placements and teaching staff (Baernstein et al., 2009; Finn et al., 2010; Karnieli-Miller et al., 2010; Park, Woodrow, Reznick, Beales & MacRae, 2010), despite one study reporting that more than half of participants perceived their colleagues to be unprofessional (Baernstein et al., 2009). Theoretically, the hidden curriculum has equal potential to impact professionalism positively and negatively (Fochtman, 2006; Hammer, 2006). However, the vast majority of research suggests that resulting influences are negative, with some concluding that it is the main reason for observed erosion in professional qualities such as communication skills, empathy, moral development, and positivity amongst trainee healthcare professionals (Finn et al., 2010; Hafferty & Franks, 1994; Hammer, 2006; Hicks et al., 2001; O'Flynn et al., 2014; Rosenthal et al., 2011; Shapiro et al., 2006; Stern, 2000; Thomas et al., 2007; Wear, 2006).

The implication of the hidden curriculum is not that sociological theories of professionalism are incorrect, but that they may be ineffective in providing a vehicle for appropriate professional development. The potential for negative influences on professional development is argued to be largely unrecognised and unmitigated in traditional, role-modelling approaches to professional development (Giroux & Penna, 1979), meaning that where the aim is to support improvements in professionalism, the sociological approach may have limited relevance. In fact, this approach may better explain the cultural replication of unprofessional behaviour than professionalism, thus adding weight to historical criticisms that led to this viewpoint's fall from grace during the second half of the twentieth century (See section 1.1.1; Hafferty & Franks, 1994). In order to approach professionalism as a positive attribute to be developed, rather than a negative one to be avoided, alternative theory is required.

1.3.2 Educational theory.

The discreditation of the sociological approach to understanding professionalism led to increasing emphasis on the educational context. As professionalism came to be viewed as a competency that may be observed and

verified through assessment in the external world, understanding of the concept shifted similarly (Coulehan, 2006; Marei et al., 2018; O’Flynn et al., 2014; Stern, 2000). However, educational practice remained heavily influenced by sociological theory associated with professionalism, continuing to rely predominantly on role modelling for development (Baernstein et al., 2009). The field of education has been criticised for neglecting the role of theory in practice due to a pervasive preference for relying on tradition (Kolb & Kolb, 2005). Consequently, dedicated theories are lacking, with practice instead supplementing vocational sociological role modelling with additional formal education running alongside (Baernstein et al., 2009; Carey & Ness, 2001; Doukas, 2006; Hendelman & Byszewski, 2014; Hicks et al., 2001; Kim et al., 2004; Monrouxe & Rees, 2012; O’Flynn et al., 2014). Debate continues as to whether the latter can be effective while the former, and particularly the hidden curriculum associated with it, remains so influential (Hafferty & Franks, 1994; Hendelman & Byszewski, 2014; Karnieli-Miller et al., 2010; Rosenthal et al., 2011). The following sections discuss educational models commonly applied to the teaching and learning of professionalism.

1.3.2.1 Miller’s learning pyramid.

Miller’s pyramid is a model for assessing clinical competence within medical training (Cruess, Cruess & Steinhert, 2016; Goldie, 2013; Miller, 1990) extrapolated to apply to professional development, offering a theoretical framework from which to derive assessments (Al-Eraky & Marei, 2016; Cruess et al., 2016; Zijlstra-Shaw et al., 2012). Miller’s pyramid explicates a theoretical link between what an individual knows and what they actually do, with the latter being the objectively observable behavioural level of professionalism. As educational approaches to understanding professionalism focus on external behaviours, Miller’s pyramid aligns well. Miller’s pyramid acknowledges that the training an individual receives (what they know) does not equate to their real-world performance (what they do), suggesting that the further students progress from knowing to doing, the higher their attainment (Cruess et al., 2016; Goldie, 2013; Miller, 1990). Miller’s pyramid proceeds through four ascending levels of attainment, with the most novice level involving knowledge of professionalism, followed by knowing theoretically how to deliver it, before

showing how to practically deliver it, finally reaching the most sophisticated level of ‘doing’ professionalism within daily practice (Al-Eraky & Marei, 2016; Goldie, 2013; Miller, 1990; Zijlstra-Shaw et al., 2012). The four levels provide indications of the types of assessment that might be useful to appraise that stage of development (Al-Eraky & Marei, 2016; Arnold & Stern, 2006; Cruess et al., 2016; Goldie, 2013; Miller, 1990). For example, knowledge might be assessed via fact-based multiple-choice questionnaires, while ‘doing’ would be better assessed through work-place observation or peer colleague evaluations (Goldie, 2013; Miller, 1990; Zijlstra-Shaw et al., 2012).

Miller’s pyramid does not discuss the softer skills usually associated with professionalism. Miller’s original article focuses solely on assessing technical skills (Miller, 1990), specifically providing a framework for identifying what types of assessment are appropriate to the progressive levels of clinical skill (Al-Eraky & Marei, 2016; Arnold & Stern, 2006; Cruess et al., 2016; Goldie, 2013). It therefore fails to provide any account of professionalism or its development, despite being commonly associated with its assessment within the literature. As a result, Miller’s pyramid provides little support to developing a definitive understanding of professionalism.

1.3.2.2 Kolb’s experiential learning theory.

Kolb’s experiential learning theory (ELT; 1984; Kolb, Boyatzis & Mainemelis, 2001; Kolb & Kolb, 2005) is a model of learning characterised by grounding it in the role of experience, with a strong pedigree of application particularly in medical education (Kolb & Kolb, 2005). Learning is conceptualised as a dynamic process borne out of the interplay of cognitive, behavioural, social, and environmental influences (Burns & Danyluk, 2017; Kolb, 1984; Kolb et al., 2001). Kolb (1984) proposed that individuals have the capacity for four distinct but related modes of learning that are responsive to the social and physical characteristics of the external environment and situational context (Burns & Danyluk, 2017). The modes of learning are sub-categorised into two operations: grasping and transforming experience (Burns & Danyluk, 2017; Kolb & Kolb, 2005). Grasping relates to the way individuals perceive the external world and information about it (Kolb et al.,

2001; Kolb & Kolb, 2005). Transforming experience refers to the internal processing of experiences so that they may be stored to guide future behaviour (Kolb et al., 2001; Kolb & Kolb, 2005). Although focused on the receipt and processing of information, however, Kolb stressed that the ELT was not a cognitive model, because these modes are intrinsically responsive to changing conditions in the external environment, with the tension created between the modes and their context being the driving force behind learning (Kolb & Kolb, 2005). Kolb (1984) conceptualised learning as a holistic process that may not be understood by reducing it to isolated processes in cognition, emotion, or behaviour, requiring a more global perspective on conflicts between the subjective and objective worlds and how these are resolved.

When grasping information, individuals must choose one of two available modes that are related but cannot be employed simultaneously (Kolb et al., 2001; Kolb & Kolb, 2005). These modes are concrete experience (CE) and abstract conceptualisation (AC; Burns & Danyluk, 2017; Kolb, 1984; Kolb et al., 2001; Kolb & Kolb, 2005; O’Sullivan, van Mook, Fewtrell & Wass, 2012). The former relates to a preference to grasp information through sensory experience, focused on the hard data of the immediate environment without the bias of pre-conceived theories about it (Kolb, 1984; Kolb et al., 2001). The latter involves the preference to receive information in more abstract, theoretical terms, integrating events into logical theoretical frameworks (Kolb, 1984). Kolb and colleagues (2001) explained the tension between these modes using the example of driving as it is, of course, virtually impossible to physically drive a car at the same time as reading the manual. The transforming modes are similarly opposing, involving reflective observation (RO) and active experimentation (AE; Burns & Danyluk, 2017; Kolb et al., 2001; Kolb & Kolb, 2005; O’Sullivan et al., 2012). RO refers to a preference to observe events and their consequences before reflecting on them in abstract terms, while AE involves more hands-on, trial-and-error approaches (Kolb, 1984; Kolb et al., 2001).

Kolb (1984) suggested that ideally, learners occupy all four modes in an ongoing cyclical manner, in order to gain the most from learning experiences over time (Kolb et al., 2001; Kolb & Kolb, 2005). CE is the basis of RO, enabling AC to build mental representations of experiences, before AE tests these abstract concepts and guides further new experiences (Kolb et al., 2001; Kolb & Kolb, 2005). In

reality, however, most people prefer to occupy two of the four modes, one grasping and one transforming, representing their dominant learning modes (Kolb et al., 2001). In order to gain the most from learning experiences, individuals must move beyond these dominant comfort zones to grasp and transform new information most effectively (Burns & Danyluk, 2017; Kolb, 1984). This involves challenging the suppression of weaker modes that limit learning, favouring the more sophisticated and desirable approach of integrating experiences transformed by each of the four modes cyclically (Kolb, 1984). The latter ability is proposed to come later in life from mid-career onwards, with younger adults instead using learning experiences primarily to shape and maintain their mode preferences, rather than striving to this ideal place of experiential learning (Kolb et al., 2001; Kolb & Kolb, 2005). In essence, the early career individual is focused on passing their chosen internship by attaining certain learning milestones, whether in education or employment, with the broader aim of using learning to deliver excellence coming later in their career (Burns & Danyluk, 2017).

The concept of naturally shifting learning styles as a career progresses provides insight into the attraction of the ELT for professional training such as medical education as it suggests that professionalism develops with exposure to professional culture over time (Kolb & Kolb, 2005). However, it is also problematic: if the ideal learning required to deliver professionalism can only ever develop during later career stages as a result of experience, the ELT would suggest that individuals cannot be 'trained' in professionalism at all. This questions whether education-based attempts to understand, develop, and assess professionalism are even worthwhile.

Placing this issue to one side, however, the ELT also offers other advantages. For example, the ELT emphasises the role of context in professionalism (see section 1.4.2 for further discussion). Professionalism is often described as contextually-bound, as it cannot be understood in the absence of its temporal, cultural, social, and organisational context (Carter et al., 2015; Evans, 2008; Goldie, 2013; Kelley, Stanke, Rabi, Kuba & Janke, 2011; Marei et al., 2018; Troman, 1996). In addition, the ELT conceives of learning as a process rather than something defined by behavioural outcomes (Kolb, 1984). This responds to criticisms that educational approaches to understanding professionalism are overly focused on behavioural

manifestations at the expense of deeper functioning (Goldie, 2013; Rees & Knight, 2007).

Despite its strengths, however, in proposing that exposing individuals to a range of traditional educational methods including work and classroom-based learning, will be sufficient to prompt the experience required for all learning style preferences to become effective, it may be claimed that the ELT model itself adds nothing more substantial to the understanding of professionalism than advocating current methods remain in place. As previously discussed, increasing claims of unprofessionalism suggest that this strategy is ineffective (Bertolami, 2004; Hafferty & Franks, 1994; Hendelman & Byszewski, 2014; Monrouxe, 2010), and therefore undermine the utility of the ELT as a means of understanding the development of professionalism. Indeed, Kolb and Kolb (2005) describe the ELT as resting upon the assumption that learning is a process, not an outcome. This might be taken to suggest that it provides insight into ways to prompt intra-individual professional development, but that it is not be appropriate to efforts to promote inter-individual development and the assessment and benchmarking required to assess professionalism today. It is possible therefore that despite its influence in medical education, its uses therein may actually be based on *a posteriori* justification for retaining traditional methods of experiential socialisation into the profession, rather than as a dynamic approach to understanding professionalism. As a result, in the context of this thesis and its efforts to explore new avenues for understanding professionalism with potential for inter-individual measurement, Kolb's ELT appears to have limited utility.

1.3.2.3 Service-learning.

Service-learning approaches to professional development use traditional workplace socialisation tailored to meet the learning outcomes of formal education (Anderson et al., 2014; Crandell, Wiegand & Broksy, 2013; Saltmarsh, 2005). Service-learning has potential to bypass hidden curriculum issues by manipulating the environments to which trainees are exposed. Specifically, environments likely to prompt a greater depth of social and interpersonal understanding are chosen, supporting the development of humanistic qualities related to professionalism, such

as compassion and a desire to serve the underserved (Anderson et al., 2014; Wise & Yuen, 2013). For example, medical trainees may undertake clinical tasks appropriate to their level of competence with clinical supervision but do so in non-clinical placements and with non-clinical colleagues to observe and learn from, such as by providing healthcare interventions within the context of a charity working with homeless communities. Service-learning removes the historical priority placed on technical education, placing it equal to the delivery of professional services, thus requiring trainees to develop their interpersonal, professionalism-related skills to a level equal to their technical ones (Anderson et al., 2014; Bringle & Hatcher, 1996; Sandberg, 2018). Trainees are required to provide collaborative services based on equal relationships with their clients and colleagues from other professions, as a basis for developing reflective abilities, social understanding, and skills in tailoring interactions and interventions to the needs of the client (Anderson et al., 2014; Bringle & Hatcher, 1996; Crandell et al., 2013; Saltmarsh, 2005).

Studies exploring the impact of service-learning on professionalism are few, but their results suggest positive impacts on interpersonal skills, altruism, positive attitudes and compassion towards service user communities, self-management, social responsibility, commitment to ongoing learning, and critical thinking, which are all attributes considered to contribute to professionalism (Anderson et al., 2014; Bringle & Hatcher, 1996; Crandell et al., 2013; Wise & Yuen, 2013). However, a study conducted by Anderson and colleagues (2014) found that there was no significant difference between the performance of students who undertook service-learning versus those who wanted to but were not given a placement. This finding undermines claims of the effectiveness of service-learning in developing professionalism as results appear confounded by individual differences in motivation, which offers an alternative explanation of findings. In addition, it has also been argued that in attempting to bypass the hidden curriculum, service-learning may actually be inadvertently replacing it with one or more others, the implications of which remain unexplored (Swaminathan, 2007).

Service-learning aligns well with the humanistic values of new professionalism (Carter et al., 2015; Birden et al., 2014; Buck et al., 2015; Goldie, 2013). Consequently, where educational approaches are criticised for creating skilled 'actors' of professionalism (Bertolami, 2004; Evetts, 2003; Finn et al., 2010; Goldie,

2013; Wear, 2006), service-learning offers the potential to develop more congruent and holistic commitment to professional values amongst trainees. However, once again this approach relies upon role modelling to prompt development, which raises the familiar criticisms associated with sociological approaches to professionalism, rendering service-learning of limited value to discussions of a theoretically grounded psychological construct for reasons previously discussed.

1.3.2.4 Weaknesses of professionalism as an educational competency.

A prevailing criticism of educational approaches rests on the assumption that if professionalism were a competency that could be mastered, traditional educational approaches would be effective in improving it. However, explicit professionalism training has been present within vocational curricula such as those of medical education for some decades, and yet claims of eroding professionalism remain, with some claiming that attending medical school may actually be an active contributor to this erosion (Doukas, 2006; Frohna, 2006; Hafferty & Franks, 1994; Hendelman & Byszewski, 2014; Kim et al., 2004; Rosenthal et al., 2011; Shapiro et al., 2006). As a result, it is argued that theories conceiving of professionalism as an educational competency must be incorrect, as it proves resistant to tried and tested educational interventions such as those based on Miller's learning pyramid or Kolb's ELT (Bertolami, 2004). It is possible, however, that this criticism is oversimplifying the picture.

As previously discussed, educational approaches to understanding professionalism remain heavily influenced by theories of professional socialisation (Wear, 2006). This enduring dominance by sociological accounts results in traditional educational approaches remaining under prioritised in both time and importance, when compared to role-modelling placements (Bertolami, 2004). It is perhaps unrealistic therefore to expect that the addition of minimal formal education to a resilient majority of role-modelling might serve to negate the problematic influence of the hidden curriculum (Hafferty & Franks, 1994; Wear, 2006). The alleged failure of educational interventions to improve professionalism may be less reflective of ineffective approaches, and more of the failure to abandon historical

allegiance to socialisation-based methods in favour of a complete shift to formal professionalism education.

Educational approaches also conceive of professionalism as a normative concept, assuming that trainees can be validly assigned to performance percentiles for purposes of benchmarking and qualification (Stern & Papadakis, 2006). Although this would be effective for directly observable, technical skills, it is less effective for professionalism as a contextually bound, multi-dimensional, and poorly defined competency (Carter et al., 2015; Evans, 2008; Finn et al., 2010; Goldie, 2013; Troman, 1996). As a result, explicit benchmarks against which individual professionalism performance may be compared are lacking (Ben-David et al., 2004; Bonke, 2006; Marei et al., 2018; Zijlstra-Shaw et al., 2012). Attempts to remedy this issue gravitate ever more towards the more objectivity of direct behavioural observation as a means of comparing individual conduct (Goldie, 2013), but this has resulted in further criticism.

It is argued that one of the most fundamental weaknesses of educational approaches to understanding professionalism is their emphasis on observable behaviour, which oversimplifies the concept and neglects the other levels at which it operates (Goldie, 2013; Rees & Knight, 2007). When conceptualised as a competency, professionalism loses something of the construct discussed in section 1.2 relating to personal character. Where the focus is on specific behaviours exhibited within specific contexts, this encourages trainees to act professional rather than be professional (Bertolami, 2004; Evetts, 2003; Finn et al., 2010; Goldie, 2013; Wear, 2006). For example, using Miller's terminology, behaviour during classroom learning or when being observed on placement 'showing how' they conduct themselves may be exceptional, while true behaviour away from assessing eyes while 'doing' real-world professionalism may be quite different (Bonke, 2006; Miller, 1990). It is therefore argued that educational approaches to professional development encourage students to learn the correct words to use when answering related questions or to accurately predict correct answers within situational judgement tests without this in any way reflecting their real-world conduct upon qualification (Bertolami, 2004; Finn et al., 2010; Hafferty & Franks, 1994; Marchalik, 2015; Wear, 2006). This approach is claimed to have produced a generation of graduates skilled in producing the correct answer and imitating or

faking the correct behaviour upon demand in a way that rewards them with higher performance evaluations but does nothing to alter their character (Evetts, 2003; Finn et al., 2010; Goldie, 2013; Rees & Knight, 2007; van Mook et al., 2009; Wear, 2006).

The criticism that professionalism has become a role to be played as a result of educational approaches accounts for a shift in perspective towards the gestalt concept of new professionalism. Taking this perspective, the focus naturally shifts away from educational to psychological theory as a means of understanding not only the behaviour of the individual, but its psychological antecedents and the nature of links between the two. A psychological approach to understanding professionalism may therefore be alone in providing the descriptive and explanatory power required to understand professionalism at all levels from fundamental values and beliefs, to its external manifestation as objectively observable behaviours.

1.3.3 Psychological theory.

The field of psychology does not offer any dedicated theoretical models of professionalism, perhaps due to the persistent historical dominance of the field by sociological accounts that have shifted attention away from the level of the individual (Cruess & Cruess, 1997; Evetts, 2006) and the lack of a clear definition of the construct to be theorised (Freidson, 1994). However, as common usage and understanding of the concept have evolved, the concept of professionalism has progressed away from a system in which people operate and towards a characteristic or attribute connected to the individual (Rees & Knight, 2007; Zijlstra-Shaw et al., 2012). In order to keep pace with this shift, it appears reasonable for the field of psychology to turn its attention in this direction as a new and as yet unexplored subject of study.

Generalisable descriptions and understanding of the psychological construct of professionalism require quantitative exploration that, in turn, requires a theoretical account of the construct before such research may begin (Burke Johnson & Gray, 2010; Creswell & Plano Clark, 2011; Morgan, 2007; Neuman, 2000; Teddlie & Tashakkori, 2003, 2010). In the absence of dedicated psychological theory regarding

professionalism, there are a number of existing theories that might offer insight into the mechanisms governing professional development and conduct as a basis for quantitative exploration. The potential implications of these theories for the understanding of professionalism are discussed in the following sections.

1.3.3.1 Social cognitive theory.

Social learning theory (Bandura, 1977), later social cognitive theory (Bandura, 1997), is one of the most influential psychological theories of the twentieth century (Honicke & Broadbent, 2016; Watson, 2013). It offers a unified theory of human behaviour that accounts for the acquisition of behaviours or competencies through active and purposeful cognitive construction, and an account of how and when those competencies might be exercised (Bandura, 1997, 1999b). Bandura's (1977) social cognitive theory posits that behaviours are acquired by observing other people and events within the external world, based on the behaviourist principle of stimulus and response (Bandura, 1977, 1999a; Watson, 2013). It goes beyond this, however, by elaborating on the process taking place between the stimulus and response, and how this moderates outward behaviour (Bandura, 1977, 1999a). Social cognitive theory rejects the concepts of imitation and the passive imposition of behaviours through reinforcement or punishment, instead suggesting that active cognitive factors and processes intervene between the observation and subsequent reproduction of a behaviour (Bandura, 1977, 1999a; Watson, 2013); individuals observe the behaviour of others, which is then processed and used to construct a model predicting potential outcomes of reproducing that behaviour in various ways. Bandura (1997, 1999a, 1999b) stressed the importance of social context, in that individual cognitive factors and the behavioural patterns of the self and others, and events and situations surrounding those behaviours, all interact dynamically and bidirectionally in causing future behaviour.

The cognitive processing used to create predictive models introduces factors such as individual tendencies or biases, including cognitive characteristics that may be considered similar to personality traits, such as micro- versus macro-level attention, motivation or drive to learn and the nature of goals towards which that learning may be directed, and self-efficacy (Bandura, 1999a, 1999b; Watson, 2013).

Self-efficacy relates to an individual's assessment of their own ability to apply behavioural strategies in ways that successfully elicit desired outcomes (Bandura, 1991, 1997, 1999a, 1999b; Honicke & Broadbent, 2016). Self-efficacy is distinct from self-confidence in that it is based on a broader range of factors that include confidence, but also how motivated the individual is to engage in a behaviour, the level of risk they are willing to accept in doing so, and whether they feel they have adequate cognitive and motor skills to succeed (Bandura, 1997; Watson, 2013). These factors also take account of the social and environmental context during learning; individuals who will observe the consequences of attempts to master behaviours inform risk appraisals, and motivation levels are influenced by the anticipated physical and social consequences of mastery, including social appraisal and reward (Bandura, 1999a, 1999b).

Once individuals have processed an observed behaviour, models are created that allow them to choose from a number of options: whether to recreate the behaviour or not, and if so, how to adapt it to meet the idiographic requirements of maximising success and keeping risks of failure to an acceptable level (Watson, 2013). In this sense, social cognitive theory does not suggest that individuals blindly mimic the behaviour of others, but rather that they use their observations to inform and regulate their own attempts to master that behaviour (Bandura, 1991, 1999b). This process of self-regulation enables a version of the observed behaviour to become habitual, after which the usefulness of the model has been exhausted and the learning process is complete; the individual has acquired the competency of that behaviour (Bandura, 1997; Watson, 2013).

In attending to the mechanisms underlying the acquisition of behavioural competencies, social cognitive theory offers insight into the mechanisms of the hidden curriculum. When individuals observe peer colleagues and superiors perceived as successful or characterising desired qualities in the workplace, they are likely to be more motivated to attempt to master a version of their behaviours for themselves (Watson, 2013). Behaviours observed during work-based placements, such as those undertaken by medical trainees, may be viewed as exciting and highly relevant to the real world of medical practice, with the tangible rewards of positive social appraisals from colleagues and trainees, or grateful patients, making classroom-based formal curricula appear dry and of less relevance to the day-to-day

role (Bertolami, 2004; Hicks et al., 2001). This could lead to the replication of occupational cultures, even where it conflicts with formal curriculum.

Social cognitive theory acknowledges individual motivation as a key determining factor in whether learning results in behaviour change (Bandura, 1999a, 1999b). As a purposeful process, social cognitive learning is regulated by individuals to provide maximum perceived desirable outcomes and minimum perceived undesirable outcomes, in relation to their goal (Bandura, 1991). Where that goal is to integrate oneself into an existing workplace culture in order to capitalise on its perceived social benefits, the behaviours learned are likely to closely mirror those already in place within the existing professional culture. Bandura (1999b) suggested that when individuals directly observe the positive consequences of a behaviour vicariously, this increases the likelihood of attempts to master a version of that behaviour relative to, for example, receiving information about consequences without direct observation. In the context of medical education, particularly with the controversial perceived status and benefits for members of the medical profession, this explains the pervasive nature of the hidden curriculum. Where trainees observe behaviours exhibited by role models who have achieved the desired status within society, the influence of these behaviours is likely to overshadow information received during formal, classroom-based learning (Hafferty & Franks, 1994). The implication of this is that medical students observing the unprofessional conduct of qualified doctors during clinical training placements may be likely to attempt to emulate their behaviour within their own practice, even where such behaviour contradicts the instruction of classroom-based teachers.

In the context of professionalism more generally, social cognitive theory describes and explains both the acquisition and expression of professionalism-related behaviours. It provides a theoretical account of the links between context and cognition, and how they might result in observable behaviours (Bandura, 1999a, 1999b). Social cognitive theory is therefore able to provide a theoretical account of all levels at which professionalism is proposed to operate, from fundamental schema, beliefs, and attitudes, through cognitive processes, resulting in externally observable behaviours within the social and contextual environment.

1.3.3.2 Cognitive affective personality system.

The cognitive affective personality system (CAPS) theory draws together various psychological concepts central to the discussion of personality and its relationship with behaviour in the external world (Mischel & Ayduk, 2002). The CAPS theory integrates trait theories of personality with those based on socio-cognitive-affective processing (Mischel & Shoda, 1999) by combining ideas from Bandura's social learning and cognitive theories (Bandura, 1991, 1999a, 1999b; Mischel, 1979; Mischel & Shoda, 1998) with trait-based conceptualisations of stable personality, such as the big five (Mischel & Shoda, 1998; Goldberg, 1992). It also draws on constructionist theories such as personal construct theory to understand how the external is construed (for detail, see chapter 4 section 4.2.2; Kelly, 1955; Walker & Winter, 2007).

Personality refers to stable elements of an individual's character thought to play causal roles in behaviour (Mischel & Shoda, 1995). Personality is a controversial concept. Traditional personality theories propose broad, underlying traits differing between individuals but remaining relatively stable throughout adulthood intra-individually, causing similar stability in behavioural tendencies (Mischel & Shoda, 1999). Despite the intuitive attraction of an account of the enduring and consistent nature of the introspectively perceived self, personality theories have been criticised as unable to account for behavioural variation and for ignoring empirical and theoretical literature demonstrating the role of psychological processing in, and contextual influences on, behaviour (Mischel & Shoda, 1999). In other words, if individuals are subject to stable, enduring traits that govern behaviour, the question is raised as to why such wide variation is observed in cognitive and affective processing and behaviour, even in a single individual across the course of their adult life (Mischel, 1969, 1973, 1977; Mischel & Shoda, 1995, 1999).

Rival theories attempt to overcome this issue by attributing behavioural variation to idiographic socio-cognitive-affective processing. Bandura emphasised the role of social context (Bandura, 1999a, 1999b), while Kelly sought to understand the depth of idiographic construing (Butler, 2009a, 2009b; Catania & Randall, 2015; Cattell, 1965; Kelly, 1955; Walker & Winter, 2007). Mischel (1969), however,

conceptualised personality as a product of stable and change-resistant cognitive sets that include biases such as attentional preferences, and style, such as individual approaches to problem solving. These sets are known as cognitive-affective units (CAUs), which guide processing associated with understanding and producing behaviour in ways that suggest an underlying consistency of character (Mischel, 1973; Mischel & Shoda, 1995, 1998, 1999). CAUs are psychological processes that both enable and constrain behaviour by interacting with each other and the situation people find themselves in (Mischel & Shoda, 1999). CAUs include encoding, competencies and self-regulation, expectancies, values/goals/motivations, and affect (Mischel, 1973; Mischel & Shoda, 1999; Shoda, Mischel & Wright, 1994).

1.3.3.2.1 Cognitive-affective units.

Encoding refers to the idiographic system by which individuals cognitively process and organise incoming information (Mischel, 1973; Mischel & Shoda, 1995; Shoda et al., 1994). Individuals understand their world according to how they encode it, or the way they construe objects and events within it (Kelly, 1955; Mischel & Shoda, 1998, 1999). Although all CAUs including encoding are specific to the individual, groups of people may share encoding or construing tendencies, and thus share similar resulting views or behaviours (Mischel, 1973; Mischel & Shoda, 1995, 1999). During social learning, observed behaviours are cognitively processed to enable encoding, with this processing subject to individual tendencies such as those relating to attention (Mischel, 1973). For example, if an individual is predisposed to attend to positive behavioural consequences more readily than negative, they are more likely to encode that behaviour as eliciting positive consequences, ignoring negative impacts. This results in the individual being more motivated to master that particular behaviour and driven to produce it again at other times, based on its anticipated benefits.

Mastered behaviours become *competencies*, acquired via direct experience and observational learning, as described by social cognitive theory (described in detail in section 1.3.3.1; Bandura, 1999a; Mischel, 1973; Mischel & Shoda, 1995). Once the active cognitive learning process is complete, individuals have constructed and encoded a blueprint for their version of a given unit of knowledge or behaviour

(Mischel, 1973; Mischel & Shoda, 1998, 1999). These blueprints provide the potential for the behaviour to be produced again in the same or a similar form at another time (Mischel, 1973, 1979).

Individuals monitor, evaluate, and adapt their efforts to achieve goals by applying their competencies (Mischel, 1973; Mischel & Shoda, 1995; Shoda et al., 1994). *Self-regulation* involves setting internal standards for the use of competencies and assessing the likelihood of those standards being attained (self-efficacy; Bandura, 1991, 1999b), and drawing on encoded potential for acceptable expected consequences (Mischel, 1973; Mischel & Shoda, 1999). Individuals make plans to organise and deploy competencies in goal-directed ways, monitoring success in moving towards those goals and adjusting behaviour or goals themselves to maximise desirable outcomes (Mischel, 1973; Mischel & Shoda, 1998, 1999). This means that although individuals have the competency to act in line with formal curricula, for example, they may choose not to, instead favouring strategies observed through the hidden curriculum, if they appear to deliver additional benefit.

Self-regulation involves anticipating the consequences of behaviours, known as *expectancies* (Mischel, 1973; Mischel & Shoda, 1995; Shoda et al., 1994). Expectancies are based on the social context of events and the outcomes observed or experienced when competencies were acquired (Mischel & Shoda, 1998; 1999). Bandura suggested that behaviours are motivated by forethought, with related expectations crucial to behavioural acquisition (Bandura, 1991, 1999a, 1999b). Expectancies help individuals to choose whether or not to use previously acquired competencies, are acquired through both observational learning and direct experience, and are reciprocally influenced by the effects of self-efficacy (Mischel, 1973; Mischel & Shoda, 1998, 1999).

Expectancies are broadly understood as if-then statements (e.g. *if I do x, then y will happen*; Mischel, 1973; Mischel & Shoda, 1995). Where a situation allows for multiple possible behaviours, the subjective value of consequences expected from each behaviour in that situation become a discriminating factor (Mischel, 1973). Where individuals favour social praise over higher grades, for example, they may value vocational outcomes over academic. Where the absence of information prevents the generation of specific expected consequences, such as in unfamiliar

situations, behaviour is based on more generalised expectancies (e.g. *if I am nice, then people will respond positively*; Mischel, 1973). Consequently, in unfamiliar situations, individuals will likely fall back on tried and tested competencies that have previously delivered valued expectancies at other times.

As new consequences are experienced, new expectancies are encoded into individuals' pre-existing cognitive systems with connections to related competencies, meaning that the likelihood of producing certain behaviours may change over time. However, where new consequences lead to cognitive dissonance by conflicting with strongly held assertions about the self, they may have minimal or non-existent effects on the likelihood of that behaviour being produced again in future. Similarly, directly experienced consequences versus those received 'theoretically' via the communication of information are encoded more strongly and are therefore more resistant to future change (Mischel, 1973). This suggests that expectancies acquired through direct experience, such as during work-based learning rather than teaching, are more strongly encoded and therefore have a greater impact on future behaviour. However, where new expectancies fail to reinforce strongly held beliefs that the individual has about themselves or the physical and social world around them, they may have no impact at all regardless of whether they are directly or vicariously experienced.

Expectancies may be *valued* to a greater or lesser extent, depending on individual subjectivity (Mischel, 1973; Mischel & Shoda, 1995; Shoda et al., 1994). Individuals may be *motivated* by internal rewards such as positive mood, or external reward such as desirable outcomes (Mischel & Shoda, 1998, 1999). Mischel (1973) suggests that individual motivations, values, or *goals*, are relatively stable, but vary depending on situational factors, such as the possibility for social comparison or how happy the individual feels that day (Mischel & Shoda, 1995). This suggests that although individuals have relatively stable lifelong goals in general terms, how much they are valued or how motivated individuals are by them will vary across different specific situations.

The way a person feels affects the processing and encoding of incoming information and coping and self-regulatory responses (Mischel & Shoda, 1995). *Affect* or mood, and related physiological sensations and arousal, affect socio-

cognitive processing. For example, a positive/negative mood will induce greater/lesser expectations of positive behavioural consequences or enhanced/diminished estimations that goals are achievable respectively. As a result, changing mood can both affect and be affected by the encoding of incoming information and individual self-efficacy, expectancies, motivations, and self-regulatory strategies (Mischel & Shoda, 1995). In its interaction with all of the other person variables, affect is the internal context in which each must operate. This suggests that despite individual tendencies being influenced by the other CAUs, affect moderates these influences from moment to moment, thus potentially resulting in behavioural variation even within a short space of time.

CAUs as an account of behavioural variation within personality received much criticism, however, along with other cognitive accounts. Critics claimed that in listing cognitive processes such as self-efficacy, self-regulation, or expectancies, social cognitive theories and the model of CAUs offered no explanation of precisely how these processes interact to produce behaviour (Mischel & Shoda, 1999). This criticism opposed those of trait theories of personality: trait theories were accused of accounting for subjectively enduring personal identities while failing to account for behavioural variation, while cognitive theories were accused of accounting for behavioural variation while neglecting the enduring sense of self and its observation by others. In response, the CAPS theory was developed.

1.3.3.2.2 An integrated theory of personality and cognition.

The cognitive affective personality system theory integrated the cognitive processes described by the CAUs with traditional trait theories of personality (Mischel & Shoda, 1998, 1999). CAPS theory suggests that the potentially overwhelming flow of incoming information during daily life must be cognitively processed into some form of manageable, organised system (Mischel, 1969, 1979). This system is arranged in ways that feature characteristic patterns of cognitive-affective-behavioural functioning (Mischel & Shoda, 1995, 1998). Within these characteristic patterns lies the concept of enduring personality (Mischel & Shoda, 1995, 1998). Alongside this, the CAPS theory also acknowledges the hugely influential role of social and environmental context (Mischel, 1969). The complex

and bidirectional interplay of person variables, including CAUs, and situation variables, such as social and environmental contextual cues, gives rise to the potential for wide variation in observable behaviours, even where situations appear superficially similar to observers (Mischel, 1969).

As an example, consider an individual whose CAUs mean that they tend to attribute negative interpersonal reactions to their own failings or unlikeability. This individual will be primed to detect features of a situation confirming this encoding. This means that even where the objective features of a situation appear dissimilar, such as a work meeting versus a social outing, this individual may detect similarity based on their subjective sensitivities that result in predictable behavioural variations, and vice versa (Mischel & Shoda, 1999). Put another way, even where behavioural variation appears incoherent to observers, it will be consistent with the personality of the individual in the context of their idiographic CAPS as it interacts dynamically with the world around them. The CAPS theory speaks to the interaction between personality as a product of stable cognitive sets and more changeable factors, both internal and external, in ways that allow for the simultaneity of enduring selves and behavioural variation (Mischel, 1969, 1979). Mischel and Shoda's (1995, 1998) CAPS theory integrated trait and processing theories to form a single, unified theory of personality and behaviour (Mischel & Shoda, 1998, 1999). It describes personality as resulting from stable and enduring individual differences in characteristic cognitive sets that may be interpreted as traits, but with the crucial addition of individual differences in processing systems within which the former reside, which gives the potential for huge yet predictable behavioural variation (Mischel & Ayduk, 2002; Mischel & Shoda, 1998, 1999).

In the context of professionalism, the CAPS theory is the first theoretical account discussed to offer the potential to explore changing patterns of behaviour that might be perceived as professional development. Where individuals understand their own CAPS, they can undertake meta-cognitive activities to identify patterns of activation relating to problematic behaviours and intervene to reconstrue or re-encode them more adaptively (Mischel & Shoda, 1999). The CAPS theory also accounts for the multi-dimensional complexity of behaviours such as professionalism, by not only acknowledging but integrating the role of social factors as integral to behavioural consistency and variation. This responds to criticism that

cognitive accounts of professionalism are too objective or detached due to their neglecting the role of contextual factors (e.g. Zumbo, 2007). The CAPS theory offers insight into why work-based learning and the hidden curriculum overshadow formal curricula, but also why this is not the case for all individuals, and how interventions may be targeted to tackle these issues. The CAPS theory offers understanding as to how individuals approach unfamiliar situations, responding dynamically to features of that situation as it develops in ways that is both consistent with their character and responsive to small fluctuations in cognitive processing, affect, and context. This accounts for the role of context in making professionalism such an elusive concept. In combination, these strengths mean that the CAPS theory holds significant advantages over previously discussed educational, personality, or cognitive theories alone in accounting for professionalism.

1.3.3.3 Emotional intelligence.

Emotional intelligence describes abilities in processing and using emotional data, and their contribution to reasoning (Brackett, Rivers & Salovey, 2011; Mayer & Salovey, 1993, 1995; Mayer et al., 2016; Mayer, DiPaolo & Salovey, 1990; Mayer, Salovey & Caruso, 2004b, 2008; Mayer, Salovey, Caruso & Cherkasskiy, 2011; Salovey & Grewal, 2005). Emotional intelligence involves accurately appraising the emotions of both the self and others, and effectively using and regulating emotions in the self to aid cognition and in others in the pursuit of goals (Mayer & Salovey, 1993; Mayer et al., 2016; Mayer et al., 1990). High emotional intelligence is associated with emotionally skilled and well-adjusted individuals, whereas low emotional intelligence manifests as impaired social and personal functioning (Mayer et al., 1990). Although not able to account for the concept of professionalism in its entirety, emotional intelligence is included here to indicate its complementarity to the CAPS theory. It is a form of intelligence based on competencies with high relevance to the interpersonal concept of professionalism.

Emotional intelligence is argued to be a specific form of intelligence within a hierarchical model of general intelligence (Mayer et al., 2016; Mayer et al., 2008; Mayer & Salovey, 1995). The results of research exploring the factor structure of emotional intelligence suggest that it is similarly hierarchical (Mayer, Caruso &

Salovey, 1999, 2016). The four-branch model argues that emotional intelligence relates to four distinct but related areas of ability, ranging from the most basic emotion-related abilities to more complex, as described in table 1.1 (Brackett et al., 2011; Brackett et al., 2006; Grewal & Salovey, 2005; Mayer et al., 2004b, 2008; Mayer et al., 2011; Salovey & Grewal, 2005).

Table 1.1

The four-branch model of emotional intelligence.

Branch	Associated abilities/reasoning
4. Managing emotions	<ul style="list-style-type: none"> • Effectively manage own emotions to achieve a desired outcome • Effectively manage the emotions of others to achieve a desired outcome • Self-regulate and apply emotional problem solving • Manipulate interactions • Use emotional information to prompt emotional growth and development • Engage with helpful emotions and disengage from unhelpful emotions • Be open to unpleasant emotions to derive maximum emotional information from a situation
3. Understanding emotions	<ul style="list-style-type: none"> • Analyse incoming information • Identify and label complex emotions • Identify and label simultaneous emotions • Identify trends in shifting emotions and use this to predict likely future emotions and emotional shifts • Understand the relationships between emotions
2. Facilitating thought using emotion	<ul style="list-style-type: none"> • Use emotional information to prioritise thinking • Generate emotions that aid judgement, memory, or empathy
1. Perceiving emotion	<ul style="list-style-type: none"> • Perceive own emotions accurately • Perceive the emotions of others accurately • Accurately express own emotions as desired • Accurately convey the emotions of others using language • Identify deceptive or dishonest emotional expressions • Perceive emotional content within art or music • Detect and identify emotions from non-verbal cues such as facial expression or posture

Note. Adapted from Mayer et al. (2016)

At the most basic level, emotional intelligence involves the ability to perceive and express the emotions of oneself and others accurately (Brackett et al., 2011; Grewal & Salovey, 2005; Mayer et al., 1999, 2016; Mayer et al., 2004b, 2008; Mayer et al., 2011; Salovey & Grewal, 2005). This information is used to facilitate thought and cognition at the next level of emotional intelligence. At the next level of sophistication, emotional intelligence requires the ability to understand emotional information. Finally, at the most advanced level, individuals are able to effectively

manage their own emotions and the emotions of others in goal-directed ways (Brackett et al., 2011; Grewal & Salovey, 2005; Mayer et al., 1999, 2016; Mayer et al., 2004b, 2008; Mayer et al., 2011; Salovey & Grewal, 2005).

Emotional intelligence has been described using the analogy of linguistic fluency, in that some individuals can generate emotions and process emotional data more readily, while for others it requires more effort (Mayer & Salovey, 1993). This means that individuals differ in their ability to access emotion-related competencies. Highly emotionally intelligent individuals can limit or expand their awareness or experience based on emotional data, with the resulting moods acting as regulators of cognitive processing. For example, where an interpersonal interaction is not going well, highly emotionally intelligent individuals have the self-awareness and self-regulatory skills to limit the impact of their frustration or displeasure. High emotional intelligence also brings the competency of generating a wider range of emotionally relevant behaviours and using these effectively in interpersonal exchanges to achieve a desired goal (Mayer & Salovey, 1993). Taken together, these abilities indicate that even where their motivations to engage in professionalism are low, for example, highly emotionally intelligent individuals can successfully convey the opposite and achieve more positive interpersonal outcomes as a result.

At this point, it is prudent to note the significant confusion within academic and media-related literature relating to emotional intelligence. The four-branch model describes a set of specific, related but distinct abilities, within a hierarchical model of intelligence (Mayer et al., 1999; Mayer, Perkins, Caruso & Salovey, 2001; Mayer et al., 2008). However, owing primarily to the publication of a best-selling book by Daniel Goleman in 1995, this model was somewhat lost in the miasma of popular opinion (Mayer et al., 2011). Goleman's (1995) emotional intelligence draws on the four-branch model but combines it with unrelated characteristics such as political awareness, without theoretical or empirical rationale, in ways that divorce the concept from its theoretical underpinnings (Brackett et al., 2011; Mayer et al., 1999; Mayer et al., 2001; Mayer et al., 2004b, 2008; Mayer et al., 2011). Referring to Goleman's (1995) account and related works as the 'journalistic emotional intelligence', Mayer and colleagues have strived to minimise confusion and clarify their model multiple times, emphasising its heightened scientific utility for research purposes over the former (e.g. Mayer et al., 1999; Mayer et al., 2001;

Mayer et al., 2008). For the purpose of clarity, this thesis refers to the four-branch model in all subsequent references to emotional intelligence, not journalistic emotional intelligence.

Although conceptually distinct from personality theories, theories of emotional intelligence complement the CAPS theory in understanding behavioural variation. Within the CAPS theory, emotional intelligence may be considered to describe a set of competencies that have particular relevance to the interpersonal context of professionalism. Both the CAPS theory and the four-branch model also emphasise the role of socio-structural context on behaviour; one cannot have the competency to act appropriately professional without a normative, institutional or cultural understanding of what ‘appropriate’ looks like (Salovey & Grewal, 2005). Empirical evidence supports this aspect of both theories as data suggests that higher emotional intelligence is associated with higher inter-personal respect (Mayer et al., 2001), social competence (Brackett et al., 2006; Brackett et al., 2005; Lopes et al., 2004; Lopes, Salovey, Côté & Beers, 2005; Lopes, Salovey & Straus, 2003; Mayer et al., 2008), self (Lopes et al., 2003; Salovey & Grewal, 2005) and peer-reported positive social relationships (Lopes et al., 2004; Salovey & Grewal, 2005), and self-reported empathy (Mayer et al., 1999). In professional contexts, data has also been found to suggest that higher emotional intelligence at work predicts positive outcomes and attributes including higher service-user satisfaction (Virtue, Pendergast, Tellez, Waldron & Ismail, 2017); more productive interpersonal relationships, and higher integrity and leadership effectiveness (Mayer et al., 2008; Rosete & Ciarrochi, 2005); and peer and supervisor-rated interpersonal sensitivity, stress tolerance, pro-sociality, and leadership potential (Lopes et al., 2005; Salovey & Grewal, 2005). As such, it may be reasonable to expect that highly emotionally intelligent individuals would be more skilled at manipulating interactions in ways that successfully communicate and demonstrate professionalism.

Emotional intelligence also offers opportunities for measuring professionalism-related abilities. A variety of measures are available, but that developed by the authors of the theory of emotional intelligence is specifically aligned to the four-branch model and is therefore grounded in a valid theoretical account of the construct, and was the basis of the majority of the empirical studies noted above. The Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) was

refined from earlier measures developed by the same team (e.g. the Multi-factor emotional intelligence scale; Mayer et al., 2001; Mayer et al., 2004b; Mayer et al., 2011; Mayer, Salovey, Caruso & Sitarenios, 2003; Salovey & Grewal, 2005). These measures offer the advantage of not relying upon self-report data, an issue common to other measures. Self-report data is theoretically counter indicated in measuring general intelligence, as intelligence does not equate to the introspective ability to accurately assess one's own abilities (Mayer et al., 1999; Mayer et al., 2011; Salovey & Grewal, 2005). Empirically, self-report data also fails to offer more than minimal correlations with objectively measured intelligence (Paulhus, Mysy & Yik, 1998; Mayer et al., 1999; Mayer et al., 2011). Relating to emotional intelligence specifically, self-reported intelligence is not predictive of social competence, while more objective ability measures of emotional intelligence are (Brackett et al., 2006).

Although the theory of emotional intelligence, and measures based upon it, offer opportunities to measure abilities potentially involved in the broader construct of professionalism, their use is not without controversy. The theory of emotional intelligence remains relatively unchallenged, but critics level significant scepticism at the psychometric characteristics of published measures (Mayer, Salovey & Caruso, 2004a). This, however, is not the focus of this thesis, although the full implications of psychometric theory are discussed in detail in chapter 2. Setting aside the issue of measurement, the potential value of the theory itself to understanding competencies related to professionalism remains.

1.3.3.4 Theories of action and values.

Theories of action and values are conceptually different to the theories discussed in sections 1.3.3.1 to 1.3.3.3 and have a consistent history of being applied within sociological research. However, this approach is included within the section discussing psychological theory due to its groundings in cognitive and social psychology (Argyris & Schön, 1978; Friedman & Rogers, 2009). The authors themselves sought to clarify their assumption that all deliberate action is grounded in cognitive psychology, and its interaction with the social environment, self-regulatory strategies, and beliefs people hold about the world around them (Argyris & Schön, 1978). Although in later developments, theories of action were applied to

organisational systems, their original conceptualisation was in relation to individuals and their professional development (Argyris & Schön, 1978; Friedman & Rogers, 2009). As such, this approach is discussed in the context of its psychological implications and areas of complementarity with socio-cognitive and personality theories.

Theories of action developed from the empirical observation that the theories people articulate as accounting for their professional actions do not always mirror those that appear to be the true basis of their behaviour (O'Hare, 1987). It was proposed that behaviour arises from individual mental representations of action strategies (Argyris & Schön, 1978; Houchens, Hurt, Stobaugh & Keedy, 2012), and further that these mental representations offer strategies to promote professional development and growth (Argyris & Schön, 1978). Similar to the competencies of the CAPS and social cognitive theories, these representations are based on how individuals perceive or encode the external world, the people and objects within it, and the ways that they interact (Argyris & Schön, 1974; Houchens et al., 2012). Theories of action accept that outward behaviours, such as physical movement, articulated communications, or expressed emotions, result from a complex interplay of psychological and contextual factors. These factors include associations learned from direct and observed experiences between cause and effect and action and reaction, and the immediate characteristics of the context and environment (Argyris & Schön, 1974; Friedman & Rogers, 2009). With so many factors in play upon the production of any single behaviour, theories of action suggest that the motivations and values underlying actions may not be fully conscious to the individual (Argyris & Schön, 1974; Friedman & Rogers, 2009; Houchens et al., 2012).

Individuals hold complex theories about the world, theories that overlap or interrelate in ways that can obscure the reality of behavioural motivations (Argyris & Schön, 1974). For example, if a restaurant server is asked why they offered free drinks to a disgruntled customer, they may articulate that they value the reputation of the business and were seeking to provide good customer service, in accordance with company policy. This prepared articulation of their actions is known as their *espoused theory*. However, at another less accessible level, the true motivators for the server's behaviour may be that they believed that that was what the customer wanted and so offered it in order to avoid the chagrin of attempting to resolve the

complaint without recompense before finally offering a tangible gesture. This true motivation is known as the *theory-in-use*. Theories-in-use may be in line with or contradictory to the espoused theories, and individuals may be unaware of this either way (Argyris & Schön, 1974, 1978; Friedman & Rogers, 2009; Harnett, 2012; O'Hare, 1987). As such, outward allegiance expressed towards espoused theories may be entirely genuine and so should not be understood as attempts to actively deceive other parties, but rather as the natural by-product of the psychological mechanisms governing behaviour.

In relation to learning and development, behavioural change requires individuals to modify their theories-in-use to directly affect outward behaviours such as those relating to professionalism (Argyris & Schön, 1974). The tacit nature of these theories makes this difficult to achieve, but it may be achieved using double-loop learning. Single loop learning involves adjusting behavioural strategies in trial-and-error approaches to seeking different outcomes (Argyris & Schön, 1974; Houchens et al., 2012). Using the analogy of Argyris & Schön (1978), it is an action-based strategy not dissimilar to turning the heating off when a room is too hot. Double loop learning requires reflection upon the assumptions and fundamental policies underlying actions as a means of modifying both strategies and their basis, closer to modifying the heating schedule to something more appropriate to avoid the over-heating issue in the first place (Argyris & Schön, 1978; Houchens et al., 2012). Behavioural change based on double loop learning is more likely to achieve the desired outcome and bring satisfaction to the learner (Houchens et al., 2012).

In the context of professionalism, theories of action highlight the importance of self-knowledge and reflective learning to the end of growth and development, but the theory itself offers little insight into the breadth of professionalism as a holistic construct. As such, this theory might have generalised implications for professional development, and in particular the espoused theories people have about it, but cannot in itself account for the construct.

1.3.4 Conclusions regarding theories of professionalism.

There is no single, comprehensive theory that can account for professionalism as a holistic concept, but there are a number of consistent themes arising within those discussed. The sociological theory of professional identity, the educational concepts of ELT and service-learning, and socio-cognitive-affective theories all emphasise the role of context, and therefore work-based learning, in professional development. Educational theories of ELT and service-learning, and theories of action, also suggest that this is insufficient alone, as focused, explicit attempts must also be made to achieve higher professional performance and double loop learning, but without accounting for specifically how this might be achieved. Similarly, Miller's learning pyramid and theories of action both suggest a difference between knowing professionalism and 'doing' professionalism but fail to account for the psychological mechanisms underlying the latter.

Only the social cognitive and CAPS theories explicitly account for how contextual factors influence behaviour. The CAPS theory also accounts for the stable sense of self prized by theories of professional identity, but within a detailed account of exactly how this arises out of the dynamic interplay of cognitive, affective, behavioural, social, and environmental factors. On balance, it appears that the CAPS theory offers the most comprehensive account of the construct of professionalism, drawing on or complementing aspects of all the other theories discussed, but without losing sight of the contextual nuances so intrinsic to this socially constructed phenomenon.

1.4 Measuring Professionalism

In order to deliver the guarantees required of trainers and educators amongst growing interest in professionalism (Baernstein et al., 2009; Evetts, 2006, 2014; Stern & Papadakis, 2006), the ability not only to accurately measure professionalism, but to do so in a way that enables the comparison and benchmarking of skill, such that acceptable performance thresholds might be identified is required (Bonke, 2006; Stern & Papadakis, 2006). Tools to deliver this are understandably in increasingly high demand (Halpin, 2017). In response, the

already high number of published tools claiming to validly measure professionalism is continuing to increase (Mazor et al., 2007). The purpose of this measurement is to assign standards of achievement to inform qualification decisions (Ben-David et al., 2004; Bonke, 2006; Lynch, Surdyk & Eiser, 2004; van Mook et al., 2009). Such decisions carry weighty consequences for the lives of students, service users, and other stakeholders, and so the quality of evidence on which they are based is of prime importance.

1.4.1 Contemporary practice in the educational assessment of professionalism.

The assessment of professionalism is most advanced within those vocational training courses most closely linked historically to the concept, such as medical and dental education. Within these areas, there is some consistency in approaches taken to assessing professionalism, and some of the most common approaches and data sources currently used are discussed in the following sections.

1.4.1.1 Observer ratings.

Despite being time consuming and resource-expensive to arrange compared with paper-based tests, observer ratings are one of the most consistently reported forms of professionalism assessment (Zijlstra-Shaw et al., 2012). Data suggests that assessments of student professionalism made by observing tutors predict behaviour in later training years (Adam et al., 2015). Observers are asked to rate behaviour during simulated or role-played situations with elements relevant to professionalism, such as the hypothetical ethical dilemmas simulated in the objective structured/standardised clinical examination used within medical training assessment (Eva, Rosenfeld, Reitter & Norman, 2004; Goldie, 2013; Mazor et al., 2007; Zijlstra-Shaw et al., 2012).

Ratings may be provided by a range of observers, including tutors or supervisors (Fochtmann, 2006; Goldie, 2013), such as for the mini clinical evaluation exercise (American Board of Internal Medicine, n.d.) and professionalism mini evaluation exercise (Cruess, McIlroy, Cruess, Ginsburg & Steinert, 2006),

which both rely on ratings of observations of clinical interactions (Goldie, 2013). Patients or service users may be asked to evaluate performance using, for example, the patient assessment questionnaire (Hurst, Prescott-Clements & Rennie, 2004) or the physician's humanistic behaviours questionnaire (Weaver, Ow, Walker & Degenhardt, 1993), both of which ask patients to assess services received from a specific individual (Goldie, 2013). Observer ratings may also be sought from a mixture of sources for tools requiring multiple observer ratings such as 360-degree feedback mechanisms (Goldie, 2013) or the Amsterdam attitudes and communication scale (de Haas, Oort, Oosterveld & ten Cate, 2001).

Observer ratings offer the opportunity to seek multiple ratings for the same individual, some from the same observer (Arnold, 2002; Norcini, 2006). This is claimed to offer a more authentic representation of performance, based on real-world conduct (Norcini, 2006). However, despite being so widely used, reliance upon observer ratings to assess professionalism is problematic. Arranging observer ratings is time and resource intensive, meaning that simulated situations are often used in place of real-world ones, limiting the generalisability of ratings to real-world performance (Arnold, 2002). Observer ratings are also criticised as overly subjective and varying widely between raters, even where they receive specific training (Arnold, 2002; Eva et al., 2004; Fochtman, 2006; Mazor et al., 2007; Norcini, 2006; Zijlstra-Shaw et al., 2012). This suggests that observer ratings fail to offer the level of valid and objective assessment appropriate to the decisions made based on their results. Ratings are also usually completed retrospectively and therefore rely on biased individual recall, further undermining their validity (Fochtman, 2006).

Data suggests that observer ratings vary depending on the gender, race, ethnicity, or age of the individual being observed (Eva et al., 2004; Fochtman, 2006). They also suffer from so-called halo effects, whereby positive attributes such as technical skill or personal likeability, are equated with professionalism even in cases where it is lacking, thus resulting in artificially inflated ratings (Arnold, 2002). Observer ratings are susceptible to student ability to manipulate their behaviour during assessment in ways unreflective of typical practice. The ability to 'fake good' results in claims that observer ratings are incomplete in assessing professionalism, due to their sole reliance upon behavioural data (Goldie, 2013; Marchalik, 2015; Rees & Knight, 2007; van Mook et al., 2009).

Some observer ratings, particularly those from non-educational raters such as peer colleagues or placement supervisors and service users, are subject to reluctance to judge individuals critically (Arnold, 2002; Zijlstra-Shaw et al., 2012). Raters cite reasons such as not wanting to compromise working or mentoring relationships and scepticism that their reviews after such short interactions should be used for assessment purposes (Norcini, 2006; Zijlstra-Shaw et al., 2012). In addition, raters are less willing to provide negative evaluations where potential consequences are high, such as where individuals are on the brink of remediation action (Norcini, 2006). The lengthy process of providing observer ratings also increases potential for rushed or incomplete evaluations (Blake & Gutierrez, 2011). Taken together, these issues undermine observer-rated assessment by suggesting it provides only a partial image of individual behaviour.

In combination, these weaknesses suggest that observer ratings are not a valid source of professionalism assessment (Fochtman, 2006; Mazor et al., 2007). Consequently, continued reliance upon observer ratings for important decision making has been described in such strong terms as embarrassing, owing to the transparently disparate and biased nature of the results elicited (Block, 2008). Consequently, there are calls for multi-level (e.g. individual, interpersonal, and institutional), multi-method assessment enabling triangulation of behavioural and attitudinal conduct, and individual values, to comprehensively assess professionalism (Buck et al., 2004; Goldie, 2013; Wilkinson et al., 2009).

1.4.1.2 Paper-based tests.

Paper-based tests are common within educational assessment, offering more objective scoring criteria and lower resourcing implications than observer ratings (Blake & Gutierrez, 2011). Paper-based tests of professionalism include the defining issues test (Rest, 1979) and objective structured video exam (Humphris & Kaney, 2000). However, paper-based tests are often multiple-choice, assessing professionalism at the level of factual knowledge (Goldie, 2013), meaning that are considered unlikely to provide comprehensive assessment of professionalism (Zijlstra-Shaw et al., 2012). According to Miller's pyramid, this level of assessment is insufficient to indicate how an individual would actually behave in practice

(Miller, 1990; Zijlstra-Shaw et al., 2012). As such, while paper-based tests perform the function of assessing knowledge relating to professionalism, they are ineffective at assessing actual professional conduct.

1.4.1.3 Self-assessment.

Self-assessment of professionalism uses self-report, asking individuals to evaluate their own professionalism or an aspect thereof (Goldie, 2013). Examples include the Groningen reflection ability scale (Aukes, Geertsma, Cohen-Schotanus, Zwierstra & Sleats, 2007) and Penn State College of Medicine professionalism questionnaire (Blackall et al., 2007). Self-assessment surveys are time and resource efficient ways to prompt student reflection on their own conduct (Arnold, 2002). They are useful in supporting intra-individual development but as a basis for objective, inter-individual assessment, are limited.

Self-report measures are subject to biases associated with introspection and therefore correlate poorly with more objective assessments of professionalism (Arnold, 2002). Self-report data is subject to gender effects and less accurate for lower performing students than high, suggesting that they are least helpful for those students requiring the most support (Arnold, 2002). This may be due to the transparent nature of self-report questionnaires, which risks social desirability bias and the ceiling effects associated with individuals rating themselves in ways that they perceive would be most socially desirable (Arnold, 2002; Kelley et al., 2011). These issues suggest that self-report methods of assessing professionalism are not sufficiently objective or valid to enable benchmarking and qualification decisions.

1.4.1.4 Critical incident data.

Critical incident data quantitatively tallies lapses and other incidents relating to professionalism reported throughout a training programme and are common within medical education (Hodges, McLachlan & Finn, 2009). Incident reporting forms based on professional standards are commonplace within curricula with a heavy work-based placement element, enabling supervisors to report the real-world behaviour of trainees and helping to identify unprofessional behaviour early to

enable intervention (Goldie, 2013; Hodges et al., 2009). Critical incident data provides the objectivity of quantitative incident scores to support inter-individual assessment, but also enables reporters to provide qualitative comments to support intra-individual development, thus offering a more authentic, real-world indication of individual performance (Papadakis & Loeser, 2006).

However, when used as part of summative assessment, critical incident data should be used with caution. Critical incident data may be biased by negative halo effects (Kuczewski, 2006). This occurs where, once a supervisor has reported one transgression, however minor, they may be overly sensitive to other minor transgressions, as evidence confirming their initial report. Where no precipitating incident takes place, similar minor lapses by other students may go unreported, limiting the integrity of the data (Kuczewski, 2006). This sensitivity to behavioural outliers may further bias data by resulting in few reports regarding individuals performing moderately, meaning that the level of data across individuals is unequal, therefore preventing comparison (Papadakis & Loeser, 2006).

Critical incident report forms are completed retrospectively, raising memory bias and recall issues similar to observer ratings (Papadakis & Loeser, 2006). Completing critical incident reports is also time-consuming, so detail or reporting at all may suffer at busy times (Hodges et al., 2009). Critics claim that the number of report forms completed differs based on departmental culture, with some departments more likely to report minor incidents than others (Papadakis & Loeser, 2006). Departments seeking to report only major incidents will find them relatively rare, meaning that reports represent outliers rather than average performance (Hodges et al., 2009). Overall, critical incident data is not commonly used for the purposes of assessment due to its apparent limitations, which suggest it may be better used as a stimulus for intra-individual reflection exercises than inter-individual benchmarking.

1.4.1.5 Psychological measurement tools.

The forms of assessment described within sections 1.4.1.1 to 1.4.1.4 are all limited to some extent, undermining the case for their use as a basis for high-stakes

decision making. This has led educators to increasingly turn their attention towards the measurement of psychological attributes, as behavioural indicators alone are deemed insufficient to robustly assess professionalism (Rees & Knight, 2007; Zijlstra-Shaw et al., 2012). Psychological measurement uses tools validated according to psychometric theory. Psychometry or psychometrics concerns the quality of tools associated with measuring psychological attributes or characteristics, such as professionalism (Furr & Bacharach, 2014; Jones & Thissen, 2007; Miller & Lovler, 2016). A full description and discussion of psychometry is available in chapter 2, but in brief, it concerns the validity of inferences made from scores obtained on psychological measures (Furr & Bacharach, 2014; Messick, 1990; Miller & Lovler, 2016; Zumbo, 2007), as based on evidence relating to content (Furr, 2011; Furr & Bacharach, 2014; Messick, 1990), criterion (Cronbach & Meehl, 1955; Furr, 2011; Kane, 2001, 2013; Messick, 1990; Miller & Lovler, 2016; Newton & Shaw, 2014), and construct validity (American Psychological Association, 1954; Kane, 2001, 2004, 2013; Miller & Lovler, 2016), dimensionality (Furr, 2011; Furr & Bacharach, 2014; Jones & Thissen, 2007) and reliability (Borsboom, 2005; Furr, 2011; Miller & Lovler, 2016; Newton & Shaw, 2014; Raykov, 1997). Psychometry is based on detailed and highly elaborated theory used to guide best practice in validating robust measures, including those targeting professionalism as a psychological attribute.

A systematic search revealed forty-five professionalism measures published between 2007 and 2017 claiming validation according to psychometric principles (for full results, see chapter 5), but concerns have been raised repeatedly regarding the quality of evidence used to validate such published measures, along with calls for improvements (Birden et al., 2014; de Mendonça, Cotta, Lelis & Carvalho Junior, 2016; Jha, Bekker, Duffy & Roberts, 2007; Li et al., 2017; Lynch et al., 2004; Mazor et al., 2007; Veloski, Fields, Boes & Blank, 2005). However, reviews of professionalism measures undertaken to date have not approached the issue from the viewpoint of psychometric quality (e.g. Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; de Mendonça et al., 2016; Veloski et al., 2005) and so conclusions regarding the appropriateness of such assessment as a basis for qualification, registration, employment, and progression decisions are not yet possible. The time is therefore ripe for a comprehensive and systematic evaluation of practice validating measures

of professionalism, and the level of quality that may be inferred from this. A review targeting the psychometric quality of the measurement of professionalism as a by-product of the methodology used for validation would enable recommendations for robust measures that offer potential for valid decision making regarding individual professionalism.

1.4.2 The role of context.

Professionalism is described as a contextually-bound concept (Birden et al., 2014; Carter et al., 2015; Evans, 2008; Fochtmann, 2006; Kelley et al., 2011; Troman, 1996; van Mook et al., 2009; Verkerk et al., 2007), meaning that its semantic meaning and manifestation vary across individuals, historical time points, cultures, and situations (Goldie, 2013; Marei et al., 2018; Monrouxe et al., 2011; Wilkinson et al., 2009; Zijlstra-Shaw et al., 2012). Professionalism also varies within the same individual depending on their current phase of career development (Monrouxe et al., 2011; O'Flynn et al., 2014; Rosenthal et al., 2011; Shapiro et al., 2006). Its contextually bound nature complicates attempts to define professionalism in absolute terms. However, this issue cannot be ignored if educators are to be able to measure professionalism in ways that assure the quality of graduates. As a result, many educating institutions base assessments on normative definitions and consensus statements, but this brings the cyclical criticism once more that such statements neglect the role of context (Goldie, 2013).

The role of context is more than a theoretical one. Evidence suggests that measurement of attitudes as a proxy for professionalism is a poor predictor of behaviour where it is mediated by context (Rees & Knight, 2007; Goldie, 2013). Moreover, students report that context acts as a cue for them to 'switch on' professionalism only under certain circumstances (Finn et al., 2010; Treviño et al., 2006). These findings suggest that effective measurement of professionalism should take contextual factors into account in order to enable generalisable conclusions regarding global conduct (Buck et al., 2015; Goldie, 2013). No resounding resolution has as yet been offered to the issue of context in professionalism. However, it is clear that in the absence of a theoretical account of its contribution, the measurement of professionalism remains controversial. In order to more fully appreciate the impact

of this issue, it would be prudent to review the theoretical basis of extant measures of professionalism, to ascertain whether any offer a resolution to this question.

1.4.3 Conclusions regarding the measurement of professionalism.

The literature relating to measuring professionalism suggests, although not explicitly, some consensus that the future of robust and valid assessment lies in psychological measurement. However, the current state of the art in this regard remains unclear. Evidence suggests that extant measures are poorly validated (Goldie, 2013; Jha et al., 2007; Kuczewski, 2006; Lynch et al., 2004; Veloski et al., 2005; Wilkinson et al., 2009) and that data collected using them is vulnerable to social desirability bias (Buck et al., 2015) and correlates only weakly with behaviour (Buck et al., 2004; Goldie, 2013; Rees & Knight, 2007). This suggests that existing measures fall short of their intended mark. Best practice psychometric validation (for full details see chapter 2) requires time and resource-intensive ongoing processes of evidence gathering progressing over long periods (Kane, 1990, 2001, 2004, 2013; Newton & Shaw, 2014). Intensive and large-scale data collection is required to provide norm group data necessary to undertake inter-individual benchmarking for educational purposes (Evers, Sijtsma, Luassen & Meijer, 2010). As a result, there are risks that with so many measures claiming psychometric validation for measuring professionalism (Mazor et al., 2007), these measures do not offer the level of validation expected in psychometry.

It is generally accepted as preferable to improve existing psychological measures before embarking on developing new ones (Goldie, 2013; Kuczewski, 2006), but it is unclear whether this is possible in the measurement of professionalism owing to a lack of systematic evaluation of the quality of measures published to date. It is therefore recommended that a systematic review targeting the methodology of validation for measures of professionalism be undertaken before further progress might be made or conclusions drawn.

1.5 The Literature Relating to Professionalism: Concluding Remarks

The history of professionalism is long and turbulent, with ongoing issues relating to political, academic, and occupational aspects of related activities. However, its importance is unlikely to diminish and has continued to increase over recent years, resulting in a large and varied body of published research and opinion on the subject. Despite the range of content within this published literature, however, the field has struggled to make progress in finding consensus. The problem of defining professionalism remains a source of lively debate today, despite sector regulators publishing definitive guidelines against which professionals are held to account. The lack of agreed definition of professionalism is traceable to a lack of dedicated theory explaining the concept and the mechanisms at play in its behavioural manifestation. There also remains a gap in knowledge relating to the quality of measurement used to assess professionalism, which immobilises the field somewhere between educational and psychometric practice.

This thesis approaches professionalism as a psychological concept. As such, it is suggested that the key focus at present ought to be the quality of psychological measurement. Reviewing this would enable recommendations as to the future of professionalism-related research. Where possible, extant measures may be used to explore the concept further, and where required, they may be improved to enable this. Where potential is not found for the use of extant measures, the development of a new measure may be indicated using best-practice psychometric validation grounded in a coherent, dedicated theory of professionalism. Either way, a robust measure of professionalism would enable the dual outcomes of effective measurement for educational and occupational ends, and quantitative measurement required to test hypotheses enabling generalisable conclusions as to its nature as a psychological construct.

1.6 Chapter Summary

This chapter reviewed the existing literature relating to professionalism. Despite significant consensus around the growing importance of professionalism, it was found that the concept remains nebulous, with gaps in knowledge relating to its

definition and how it may be theorised and measured. This suggests that current practice in measuring professionalism may be undermined by methodological aspects of validation, but a more comprehensive review of practice would be needed in order to draw firm conclusions. In order to progress understanding of professionalism, this thesis will explore it through the psychological lens, starting with the appraisal of psychometric validation practice in this area. The next chapter will detail the theoretical underpinnings and technical content of psychometric theory, as a basis for the robust appraisal of existing measures of professionalism.

Chapter 2: A Review of the Psychometric Literature

Note to reader: the previous chapter discussed the importance of the valid measurement of professionalism in the context of increasing pressure for employers and educators to deliver a workforce equally skilled in technical competence and professionalism (Baernstein et al., 2009; Evetts, 2006, 2014; Stern & Papadakis, 2006). Extant methods of measurement grounded in the educational approach to understanding professionalism have been criticised for encouraging a generation of graduates skilled in ‘acting the professional’ and deficient in ‘being the professional’ (Bertolami, 2004; Evetts, 2003; Finn et al., 2010; Goldie, 2013; Wear, 2006). In response, attention has turned towards the validation of psychometric measures conceptualising professionalism as a personal attribute or characteristic (Rees & Knight, 2007; Zijlstra-Shaw et al., 2012). However, validation practice in developing such measures has been poor, with widespread misunderstanding of the underlying psychometric theory being a likely reason for this (Birden et al., 2014; Borsboom, Mellenburgh & van Heerden, 2004; Cizek, Rosenberg & Noons, 2008; de Mendonça et al., 2016; Duckor, 2017; Hogan & Agnello, 2004; Jha et al., 2007; Jonson & Plake, 1998; Li et al., 2017; Lynch et al., 2004; Mazor et al., 2007; Messick, 1979; Veloski et al., 2005; Wolming & Wikström, 2010).

The majority of measures published to date are in the field of medical professionalism (see the findings of chapter 5), often developed by researchers with clinical rather than psychological foci. Given that psychometric theory is a long-standing, complex, and highly technical area of psychology requiring significant dedicated study to fully understand its historical and statistical contexts and the ongoing debates still raging today, it is perhaps understandable that such researchers are potentially naïve to the full implications of this theory for the measurement of professionalism. In order to proceed with an investigation of the potential for a psychometric understanding of professionalism within this thesis, this chapter discusses psychometric theory in detail. This chapter is deemed appropriate here due to the interdisciplinary applications anticipated to arise out of the research findings and the documented propensity for confusion regarding psychometric theory among the anticipated readership.

Psychometry concerns the quality of tools associated with testing, measuring, or assessing a psychological attribute or characteristic, such as professionalism (Furr & Bacharach, 2014; Jones & Thissen, 2007; Miller & Lovler, 2016). The use of psychometry relating to professionalism is increasing, drawing attention to issues relating to its application. Before these issues may be further explored, psychometric theory and its implications for best practice in the development and use of measures should be understood.

Psychometry has a long and turbulent history, with commentators from psychology and other fields invoking arguments based on statistical, physical, historical, and practical foundations (Duckor, 2017). Controversy is far from resolved, and yet the proliferation of psychometric tests over the past hundred years has been vast, thanks in part to their use in supplementing educational assessment (Newton & Shaw, 2014). The current state of the art remains confused with many debates ongoing, but new psychometric tools are published every year citing a variety of evidence to support claims of thorough and rigorous validation. Such tools are increasingly used to make decisions regarding professional registration, recruitment, and promotions (Halpin, 2017; Miller & Lovler, 2016; Newton & Shaw, 2014). Given the weight of such decisions, it is prudent to consider critically arguments associated with psychometry and the related validation of tools. The following sections discuss major contributors to contemporary understanding of and practice in validation, in order to consider the evidence required to support inferences made from a measure of professionalism.

2.1 Operationism

The temporal context of psychometry informed its development, namely the rise and fall (or lack thereof) of operationism. Psychometric theory and practice have been heavily influenced by operationism, which states that the definition of a phenomenon equates to nothing more or less than the method by which it is measured, and that switching from one measurement method to another therefore changes the attribute itself (Maul, 2017; Maul, Irribarra & Wilson, 2016; Newton & Shaw, 2014; Sijtsma, 2012a). In its purest form, operationism suggests that

phenomena to be measured, such as psychological attributes, cannot be theorised independently of the measure designed to capture them (Maul, 2017).

Operationism was grounded in the positivist movement of the physical sciences but was criticised and rejected almost immediately as providing no meaningful avenues for scientific progress (Green, 1992; Maul, 2017; Maul et al., 2016). For example, if a distance is measured on the same day at the same time using two different rulers, operationism suggests that their generating the same result has no meaning whatsoever and cannot therefore be used to draw any inferences (Green, 1992). As a result, operationism has long since been considered outmoded by the scientific community in general. However, it has always remained particularly resilient as a rationale for psychological research, still taught in general psychological education today (Green, 1992; Maul, 2017; Maul et al., 2016; Michell, 1997, 2000; McGrane, 2015; Newton & Shaw, 2014).

Positivism was the dominant paradigm during major phases of development in psychological research (for full details, see chapter 4; Maul et al., 2016). However, psychological theory was much less advanced in its scientific journey than the physical sciences at that time and so received criticism for lacking theory on which to base its enquiries (Duckor, 2017; Humphry, 2017; Sijtsma, 2012b). It is perhaps therefore unsurprising that operationism held an attraction for psychology, in that it offered the diplomatic rationale for rejecting criticism relating to its methods that enabled the field to continue progressing (Maul et al., 2016; Michell, 1997; Newton & Shaw, 2014). According to operationism, there need be no theoretical account of a psychological attribute and how it causes variation in scores on a given measure, because these attribute and measure are defined by each other (Maul, 2017).

Critics claim that operationism survives in psychology as received cultural wisdom, accepted without challenge or critical appraisal internally and without defence to ongoing criticism externally (Green, 1992). This lack of external perspective on operationism means that its more recent conceptualisation within psychology is somewhat distorted from its original philosophical underpinnings (Green, 1992; Newton & Shaw, 2014). This opens the field to the dual criticisms of

not only adhering to outdated theory, but applying it inexpertly too (Newton & Shaw, 2014).

The existence of psychometry within psychology suggests that inferences may be made on the basis of test scores relating to psychological attributes. However, such claims are incongruent with operationism, which states that the only valid inference from a measure is that it indicates only that which it measures (Newton & Shaw, 2014). The concept of operationism within psychology today appears divorced from this philosophy, instead being used in other ways without explicit rationale: some claim that operationism states that a measure only measures that with which it correlates (Wolming & Wikström, 2010), while others use it to refer to proxy measures for non-observable psychological attributes (Green, 1992). Either way, the enduring presence of operationism within psychology appears to have effectively perpetuated a myth that psychological attributes can be measured without being defined in a theoretical framework prior to quantification. This myth enables researchers developing measures of professionalism to rely upon the rationale of operationism to operationalise the concept in whichever way serves their particular purpose. There is therefore a justification for systematically exploring the psychometric properties of extant professionalism measures before decisions as to their use may be made, particularly their theoretical grounding.

2.2 Classical Test Theory and its Core Principles

Classical test theory (CTT) is the oldest and most dominant theoretical account of measurement within psychology (Borsboom, 2005; Hambleton & Jones, 1993). CTT interprets psychometric data as observable signals indicating the existence of latent variables or attributes that cannot be directly observed (Furr, 2011). Fundamentally, CTT states that a score observed using a psychometric tool is comprised of the true score, which is an accurate signal of the construct being measured, and error (Brennan, 2011; Furr, 2011; Furr & Bacharach, 2014; Hambleton & Jones, 1993; Jones & Thissen, 2007; Miller & Lovler, 2016; Raykov, 1997; Starkweather, 2012; Suen & Lei, 2007; Yang & Green, 2011). CTT specifies that error is random and assumes that the true and error scores are independent and therefore uncorrelated (Barchard & Hakstian, 1997; Borsboom, 2005; Brennan,

2011; Hambleton & Jones, 1993; Jones & Thissen, 2007; Miller & Lovler, 2016; Raykov, 1997; Yang & Green, 2011).

CTT has dominated psychometry throughout its history, offering straightforward procedures with statistical assumptions that are easily met during data collection, including easily accessible participant sample sizes (Borsboom, 2005; Fan, 1998; Hambleton & Jones, 1993). Consequently, CTT is widely subscribed to and has a strong track record of application and utility extending beyond the remit of psychology (Hambleton & Jones, 1993). Such widespread use of CTT-based psychometric validation is taken by some to indicate methodological rigour. However, critics argue that this assumption is unfounded and suggest that the ease of its application has led to CTT being commonly used as a tokenistic offering of validation that, when misapplied, cannot deliver on its promises (Maul, 2017). In order to explore this criticism further, the basic principles and contemporary applications of CTT-based validation will be considered.

2.2.1 Validity.

Validity is commonly understood as the ability of a psychometric tool to measure the variable that it claims to target (Borsboom, 2005; Borsboom et al., 2004; Bringmann & Eronen, 2016; Furr, 2011; Furr & Bacharach, 2014; Kline, 2000; Newton & Shaw, 2014). In terms of CTT, this refers to the ability of the test to deliver a true score that accurately reflects the target variable (Furr, 2011; Hambleton & Jones, 1993; Raykov, 1997; Starkweather, 2012; Yang & Green, 2011). However, more accurately, validity is not a property of a test itself (Hogan & Agnello, 2004; Kane, 2001; Messick, 1990), but rather the inferences that may be made from data gathered using it; assessing validity involves questioning interpretations of data gathered, rather than the test used to gather them (Furr & Bacharach, 2014; Messick, 1990; Miller & Lovler, 2016; Zumbo, 2007).

Validity is intrinsically linked to the purpose of a measure; a tool may be assessed as valid for one purpose when used with one population, but not for another purpose/population (Furr, 2011; Newton & Shaw, 2014). Assessing validity is known as validation, which explores the quality of a measure or, more specifically,

the appropriateness of how its results are interpreted (Messick, 1990; Newton & Shaw, 2014). Although validation draws on a range of evidence, validity is the crucial, overriding question to be addressed (Cizek et al., 2008; Hogan & Agnello, 2004; Kane, 1990; Newton & Shaw, 2014). Where a measure is invalid, this undermines conclusions drawn from its data, potentially rendering them meaningless (Cizek et al., 2008; Furr, 2011; Kane, 1990; Messick, 1990; Wolming & Wikström, 2010; Zumbo, 2007). Validity is therefore the single most important concept to the process of psychometric validation.

Despite almost universal consensus regarding the cruciality of validity to psychometric validation (Cizek et al., 2008; Furr & Bacharach, 2014; Hogan & Agnello, 2004; Kane, 1990; Newton & Shaw, 2014), best practice in assessing it is disputed (Bringmann & Eronen, 2016; Newton & Shaw, 2014). This is perhaps unsurprising when considered from a historical perspective. In its original form, validity theory was complex and technical, and so during the twentieth century the American Psychological Association (APA) published professional validation standards and guidelines in an attempt to streamline the field and enhance the quality of psychometric tools in use (Newton & Shaw, 2014). The earliest APA guidelines defined validity in an applied manner by breaking it down into separate and distinct types, each being relevant to a specific aim of testing (APA, 1954; Cronbach & Meehl, 1955; Furr, 2011; Furr & Bacharach, 2014; Kane, 2001; Messick, 1990; Miller & Lovler, 2016; Newton & Shaw, 2014; Travers, 1951; Wolming & Wikström, 2010). Three of these types are still discussed today: content, criterion, and construct validity (APA, 1954; Cronbach & Meehl, 1955; Furr, 2011; Furr & Bacharach, 2014; Guion, 1980; Messick, 1990; Miller & Lovler, 2016; Newton & Shaw, 2014). In the latter half of the twentieth century, validity theory evolved in ways rendering these validity typologies obsolete, as validity came to be viewed as a single, unidimensional concept: construct validity. Content and criterion validity came to be viewed as types of evidence used to estimate the overall construct validity of a measure (Bringmann & Eronen, 2016; Cizek et al., 2008; Furr, 2011; Furr & Bacharach, 2014; Jonson & Plake, 1998; Kane, 2004; Messick, 1979; Miller & Lovler, 2016; Newton & Shaw, 2014; Wolming & Wikström, 2010; Zumbo, 2007).

Today, validation is viewed as an ongoing process of compiling evidence from a range of sources to provide an argument towards the overall (construct) validity of a measure for a specific and defined purpose (Bringmann & Eronen, 2016; Kane, 2001; 2013; Miller & Lovler, 2016; Zumbo, 2007). This arguments-based approach to validation suggests that validity is non-binary and that validation therefore speaks to the strength of inferences that may be made from a measure, rather than an absolute presence or absence of validity (Bringmann & Eronen, 2016; Furr, 2011; Furr & Bacharach, 2014; Zumbo, 2007). Contemporary validation best practice therefore involves an ongoing process of gathering evidence regarding the ways that data collected using a given measure may be interpreted when administered in a given context with a given population (Furr, 2011; Messick, 1990; Newton & Shaw, 2014; Wolming & Wikström, 2010; Zumbo, 2007). This means that validation is intensive and onerous, and unlikely to be achieved within a single validation study.

Validity theory has evolved rapidly, with research practice struggling to keep pace. As a result, validation practice adhering to historical iterations of validity theory is still evident in research today, despite the theory having moved onto the arguments-based approach some years ago (Hogan & Agnello, 2004; Jonson & Plake, 1998; Kane, 1990; Sijtsma, 2012a). The sections that follow consider the types of evidence that may be combined to provide a best practice assessment of validity from a contemporary perspective, as recommended by the arguments-based approach.

2.2.1.1 Content validity.

Also known as logical validity (Newton & Shaw, 2014), content validity relates to the extent to which items or questions within a scale reflect content theoretically relevant to the target attribute (Furr, 2011; Furr & Bacharach, 2014; Messick, 1990). Content validation involves reviewing items to remove irrelevant content and ensure that retained items adequately reflect the full breadth of the attribute (Furr, 2011; Furr & Bacharach, 2014; Guion, 1980; Kane, 2013; Messick, 1990; Miller & Lovler, 2016). Historically, content validation was deemed relevant to measuring ability or expertise and typically relied upon sampling theory to ensure

that items were randomly and comprehensively sampled from the full breadth of the skill domain (APA, 1954; Cronbach & Meehl, 1955; Kane, 2013; Newton & Shaw, 2014). In contemporary terms, content validation utilises expert panels to assess each item according to their own knowledge and expertise of the target attribute (Furr, 2011; Kane, 2001, 2004).

Content validation assumes that expert panels have broad and detailed knowledge of the target attribute, but also expertise in psychometric theory, to enable the transformation of attribute knowledge into effective psychometric items (Furr, 2011). It also assumes panels are sufficiently sized to represent the entirety of views regarding the attribute and its theoretical basis and to offset risks of individual bias (Furr, 2011; Kane, 2001, 2013). Robust content validation requires panels to systematically assess the content of items against the content of the target construct, but also to consider whether the structure of items (i.e. their wording, ordering, and response options) is appropriate to its assessment (Secolsky, 1987).

Content validation provides evidence that measures have been assessed against a framework of existing knowledge and theory, the importance of which is widely cited (Borsboom, 2005; Bringmann & Eronen, 2016; Cronbach & Meehl, 1955; Embretson & Gorin, 2001; Kane, 2013; Markus & Borsboom, 2013; Newton & Shaw, 2013). The importance of theoretically grounded measures stems from our understanding of how the measure accesses signals of the target attribute. Specifically, a theoretical understanding of an attribute enables a theoretical account of how it causes variation in scores on a measure (Borsboom, 2005; Borsboom et al., 2004; Bringmann & Eronen, 2016). Without a theoretical account of how an attribute affects measurable behaviours, it cannot be assumed that a measure targets its items appropriately.

There are number of threats to robust content validation arising out of the subjective nature of expert judgements. Firstly, where logical validation is undertaken by the research team who drafted the items of a measure, there is reduced likelihood of detecting bias or omissions within the content and increased likelihood of confirmation bias (Furr, 2011; Kane, 2001, 2013). Secondly, researchers can confuse content validity for face validity. Face validity speaks to the intuitive relevance or attractiveness of a measure to a potential user or the general public, and

therefore does not rely upon the critical contributions of experts required for content validation (Furr, 2011; Furr & Bacharach, 2014; Miller & Lovler, 2016; Secolsky, 1987; Sireci, 1998). Face validity evidence carries less-demanding assumptions regarding the expertise of the panel and their focus, and so may be gathered easily and quickly using any prospective test users, in exercises suggested to be closer to market research than validation (Furr & Bacharach, 2014; Miller & Lovler, 2016; Mosier, 1947; Secolsky, 1987). However, face validity evidence is often reported alongside published measures as evidence of content validity, despite it being conceptually different and unrelated to the validation argument. For example, Connell and colleagues (2018) place content validity before face validity in their publication's title, although the study did not actually involve expert consultation and instead analysed data from interviews with prospective test users regarding item acceptability. This is a clear case of face validity misrepresented as content validation, meaning that the authors' claims of content validity are unfounded, undermining their case for validity overall.

Having said this, there is debate as to whether content validation evidence contributes meaningfully to judgements regarding validity as the appropriateness of inferences. Some argue that content validation concerns sampling rather than interpretation issues (Messick, 1990). In other words, although content validation tells us how an expert interprets the content of a measure, it cannot provide insight into the meaning of scores it returns. As a result, content validation is unable to support decisions regarding inferences and may therefore be errantly considered validity at all (Messick, 1990; Newton & Shaw, 2014). Despite this, when undertaken in a systematic way, content validation still ensures that items are logically related to the target construct and grounded in existing theory, and so it remains an accepted contributor to robust validation (Furr, 2011; Kane, 2013; Lennon, 1956; Newton & Shaw, 2014; Sireci, 1998).

2.2.1.2 Dimensionality and factor structure.

Dimensionality, or factor structure, refers to the internal structure of data elicited by a measure, and specifically the type and number of underlying constructs or variables it represents (Furr, 2011; Furr & Bacharach, 2014; Jones & Thissen,

2007). Statistical evidence regarding dimensionality is used to support or refute validation by checking for discrepancies between the theorised or intended factor structure of a measure and the actual structure of the data it elicits (Furr, 2011; Furr & Bacharach, 2014). Dimensionality evidence also has implications for scoring tests, ensuring that only scores representing the same psychological attribute are aggregated and interpreted accordingly (Furr & Bacharach, 2014).

Items in a single tool might measure a single variable (unidimensional) or multiple variables (multidimensional) that may or may not be related to each other in ways that suggest summing their scores provides information about higher level or more general latent variables (Furr, 2011). To ascertain the most appropriate scoring approach, dimensionality must be established during validation. If a tool is found to measure two unrelated variables, a single score for the tool is inappropriate as it would reflect no meaningful variable. In such cases, separate scores for each variable or factor should be produced. In addition, the two scales should be subject to separate validation judgements to ensure transparency regarding the quality of each (Furr, 2011).

Factor analysis explores correlations between items as a means of grouping them into factors. If all items correlate to a similar level, this is interpreted as indicating unidimensionality, but where sets of items correlate with each other and not with others, a multidimensional factor structure is suggested (Furr, 2011). Establishing the factor structure of data elicited by a psychometric tool involves statistical procedures falling into two categories: exploratory and confirmatory. Exploratory factor analysis is used where there is no *a priori* expectation as to the factor structure of data (Miller & Lovler, 2016; Yanai & Ichikawa, 2007), such as where topics are relatively unexplored and there are no theoretical accounts from which to hypothesise a factor structure (Miller & Lovler, 2016). Exploratory factor analysis requires researchers to determine the correct factor structure using their best judgement on a case by case basis (Furr, 2011). There are, however, some guidelines available to support this process.

One of the most common approaches to deciding factor structure is to retain any factor returning an eigenvalue larger than one, known as the Kaiser-Guttman criterion (Auerswald & Moshagen, 2019; Yanai & Ichikawa, 2007). Eigenvalues

speak to the amount of variance accounted for by a factor and so retaining only those with an eigenvalue exceeding one theoretically ensures that each factor accounts for a minimum of one item or variable's worth of variance (Auerswald & Moshagen, 2019; Girden & Kabacoff, 2011). Despite being widely used, however, this approach is purely statistical and is therefore highly influenced by statistical artefacts (Auerswald & Moshagen, 2019). As a result, it is generally viewed as best avoided in validation practice (Costello & Osborne, 2005; Furr, 2011; Furr & Bacharach, 2014). An alternative approach is to examine a scree plot, which is a graphical representation of factor analysis results used to explore the evidence for different factor structures on an idiographic basis (Costello & Osborne, 2005; Furr, 2011; Furr & Bacharach, 2014; Yanai & Ichikawa, 2007). The specifics of this approach are beyond the remit of this thesis, but it is worth noting that it offers an advantage over the first in that it is led by the data rather than statistical procedures and is therefore more accurate than relying upon eigenvalues alone (Furr, 2011). The inspection of scree plots is subjective, however, and consequently may still lack clarity (Auerswald & Moshagen, 2019). Overall, it is deemed most appropriate to use a range of evidence to inform factor structure decisions, including eigenvalues, scree plots, and factor loadings. This is in alignment with practice recommended by the arguments-based approach to validation to balance a range of evidence to inform inference decisions (Furr, 2011; Kane, 2004, 2013).

Confirmatory factor analysis is used where there is one or more *a priori* hypothesis regarding factor structure (Auerswald & Moshagen, 2019; Furr, 2011; Miller & Lovler, 2016; Yanai & Ichikawa, 2007), such as where there is robust theory of a construct or where previous data, including that from exploratory factor analysis, suggest a specific structure (Furr, 2011). It is common to observe a combination of exploratory and confirmatory factor analyses undertaken on a single data set, firstly to explore the factor structure and then to confirm researcher decisions relating to it (Furr, 2011).

Undertaking factor analysis for validation purposes requires significant sample sizes to ensure the stability and clarity of factor structures, and therefore the appropriateness and generalisability of inferences made from it (Auerswald & Moshagen, 2019; Costello & Osborne, 2005; Furr, 2011; Igundunasse, 2016). Although exact sample size requirements are debateable, they are generally

sufficiently burdensome that many validation studies simply do not meet them (Costello & Osborne, 2005; Igundunasse, 2016). One of the major threats to validation using factor analysis is therefore the misapplication of the technique using too small a participant sample, which undermines all inferences that may be made from data, and therefore its overall validity.

As with all analyses, even when undertaken correctly, factor analysis is subject to limitations that should be noted during validation. Critics claim that although factor analysis can elaborate the nature of the data from a measure, decisions as to its meaning remain subjective. On the basis of factor analysis alone therefore, an overall judgement of validity is not possible (Bringmann & Eronen, 2016; Humphry, 2011). Factor analysis is therefore best used as part of a broader validation programme alongside other sources of evidence.

2.2.1.3 Criterion validity.

Criterion validity refers to the extent to which the results of a psychometric tool relate to the results of another measure hypothesised to target the same or a closely related construct (Cronbach & Meehl, 1955; Furr, 2011; Kane, 2001, 2013; Messick, 1990; Miller & Lovler, 2016; Newton & Shaw, 2014). Also known as empirical validity, criterion validation evidence takes the form of correlations with an independent criterion providing a gold standard measure of the target variable (Guion, 1980; Messick, 1990; Newton & Shaw, 2014). In previous iterations of validity theory, criterion validation was divided into categories. Convergent validity referred to correlations between scores and a theoretically related criterion (Furr, 2011). Discriminant validity referred to correlations (or more specifically lack thereof) between scores and a theoretically unrelated criterion (Furr, 2011). Predictive validity referred to correlations between scores and a future criterion, most commonly future behaviour in the workplace (APA, 1954; Cronbach & Meehl, 1955; Messick, 1990; Miller & Lovler, 2016; Newton & Shaw, 2014). Similar to predictive, concurrent validity referred to correlations between scores and present-day rather than future behaviour (APA, 1954; Cronbach & Meehl, 1955; Miller & Lovler, 2016; Newton & Shaw, 2014). Today, these categories are viewed as forms

of criterion validity evidence supporting an overall validation argument but cannot be relied upon alone to provide conclusive evidence of validity.

The main assumption underlying criterion validity is that the criterion represents a gold standard approach to measuring the target attribute (Guion, 1980; Messick, 1990; Newton & Shaw, 2014). This means that the criterion measure must have a strong validation argument already in place drawing on a range of evidence (Gullisken, 1950; Messick, 1990). If such an argument is not in place, inferences relating to the statistical relationships between scores and the criterion are undermined. For example, a criterion measure with low reliability may inaccurately measure the true ‘signal’ of the target attribute, therefore obscuring relationships and leading to errant results being invoked as validation evidence (Furr, 2011). This undermines the validation argument, calling into question all inferences made from each measure.

Criterion validation also assumes that the criterion measure be based on different theoretical assumptions and/or mechanisms to the measure being validated, to minimise threats of common method variance (Bringmann & Eronen, 2016; Chang, van Witteloostuijn & Eden, 2010; Podsakoff, MacKenzie, Lee & Podsakoff, 2003). For example, where self-report measures are validated against data from another self-report measure targeting the same construct, both data sets are subject to the same confounds, due to their being based on the same theory of measurement (Bringmann & Eronen, 2016; Chang et al., 2010). Method variance introduces systematic error into validation evidence by providing alternative or confounding explanations for criterion correlations, and method factors can account for more than a third of the common variance observed between two data sets (Podsakoff et al., 2003). Using different validation and criterion data sources, procedural controls such as collecting data using each measure at different times or in different contexts, and statistical methods controlling for or attempting to detect method variance on a post-hoc basis all serve to mitigate common method variance (Chang et al., 2010; Podsakoff et al., 2003). However, where its impact is not accounted for, falsely inflated or deflated criterion correlations are likely to be observed (Chang et al., 2010; Podsakoff et al., 2003). In such cases, criterion correlations cannot be interpreted as reflecting relationships between target constructs, as they may in reality reflect relationships between common method factors instead, an issue

particularly problematic where shared factors constitute a limitation of each method of measurement independently (Bringmann & Eronen, 2016). Once again, failing to account for method variance criterion validity evidence undermines the overall validity argument.

Due to its purely statistical nature, empirical validation may also be threatened by features of datasets used for correlations. Where a dataset's range is artificially limited in some way, such as only including scores from poor performers in order to support their development, validity coefficients are similarly limited (Furr, 2011). Correlation coefficients also assume both datasets are equally distributed, meaning that unexplored differences in distribution patterns, perhaps resulting from limited sample sizes, would violate this assumption and lead to falsely limited correlations (Furr, 2011). Both of these issues would artificially suppress criterion correlations, limiting the conclusions that may be drawn from them regarding validity.

Criterion validation also receives theoretical criticism. As empirical validation evidence became more popular during the twentieth century, so too did the tendency to rely upon it as a sole source of validation evidence. Theories as to how constructs related to measure scores were neglected in favour of an atheoretical approach guided exclusively by correlation coefficients (Newton & Shaw, 2014). A hypothetical yet pertinent case example of the potential consequences of relying upon empirical validation alone was described by Travers (1951), who postulated that if a purely empirical approach was taken to create a measure intended to predict which employees would achieve the highest seniority in their professional lives, unintended consequences might ensue (Newton & Shaw, 2014). Travers (1951) proposed that despite being highly successful at predicting success within the organisation, the measure happened to be developed within an organisation with a historical culture of anti-Semitism. Despite its apparent success in identifying individuals well-suited to the leadership culture of that organisation, Travers (1951) suggested that under further scrutiny, the measure may be found to actually identify Jewish versus non-Jewish employees, reflecting the anti-Semitic bias present within that workforce. Although empirically sound, inferences drawn from this measure would clearly not be valid, and such validation approaches ought therefore to be used with caution considering their potential limitations (Newton & Shaw, 2014; Travers,

1951). Despite this issue being much discussed in the literature, appropriate application of criterion validation procedures is rare. The psychometric properties of criterion measures are given little attention, being omitted entirely in many cases, therefore remaining unknown (Kane, 2004; Sijtsma, 2012a). Where criterion validation is misused in this way, it becomes meaningless as a form of validation evidence (Newton & Shaw, 2014).

The issues described above relate more to the misapplication of criterion validation than the principle itself. Criterion validation is valuable when contextualised by a range of validation evidence. For example, where no single gold standard criterion is available, using a range of criteria data would offer additional validation support. Each additional measure that relates as expected, both theoretically and empirically, reduces the likelihood that all measures are equally invalid (Bringmann & Eronen, 2016). Contemporary understanding suggests that rigorous validation involves both a range of criterion and logical validation evidence (Cronbach, 1990; Gulliksen, 1950; Newton & Shaw, 2014). Where the properties of the criterion measure, dataset, and hypothesised causal relationships are known, criterion evidence can still only ever form part of the validation picture.

2.2.1.4 Construct validity.

Although considered validity in its entirety for some decades (Cronbach & Meehl, 1955; Embretson & Gorin, 2001; Jones & Thissen, 2007; Kane, 2004; Messick, 1990; Newton & Shaw, 2014; Wolming & Wikström, 2010; Zumbo, 2007), construct validity was originally conceptualised by the APA as a type of validity equal to those described above (APA, 1954; Cronbach & Meehl, 1955; Jonson & Plake, 1998). It referred to how observed scores related to the target psychological attribute and so construct validity introduced the idea that validation related to the theory of an attribute rather than a test itself (APA, 1954; Kane, 2001, 2004, 2013; Miller & Lovler, 2016). In other words, construct validity provides support for theory explaining how the target attribute causes variance in observed test scores. Contemporary understanding is not too far removed from this, but the importance of construct validity has increased.

Today, construct validity pertains to the meaning of signals captured during psychological measurement and acknowledges that these signals are merely indications of underlying latent attributes (Borsboom, 2005; Borsboom et al., 2004; Bringmann & Eronen, 2016; Miller & Lovler, 2016; Newton & Shaw, 2014). Construct validity, and therefore validity overall, relates to the way that test scores may be interpreted in light of the theory on which they are based, necessitating a shift from the historical focus on test attributes and towards the more modern focus on the inferences they permit (Bringmann & Eronen, 2016; Jonson & Plake, 1998; Kane, 1990, 2013; Messick, 1990; Wolming & Wikström, 2010). Construct validity states that neither logical nor empirical evidence alone is sufficient to support a robust validation argument (Newton & Shaw, 2014).

Construct validity concedes that latent psychological attributes are not directly observable due to the number and complexity of factors interacting to generate observable human behaviour. Construct validation therefore requires an overall estimated judgement of the best available evidence to support the validity of indirect inferences regarding the target attribute (Newton & Shaw, 2014). Validity is not a tangible absolute that is present or absent, but the strength of the argument supporting the specific inferences made from scores on a specific test when used for a specific purpose (Bringmann & Eronen, 2016; Kane, 2001, 2013; Messick, 1990; Wolming & Wikström, 2010).

The validation argument should draw upon a range of evidence supporting the theory explaining how the target variable causes variance in observed test scores (Borsboom, 2005; Borsboom et al., 2004; Bringmann & Eronen, 2016; Messick, 1990; Sijtsma, 2012a, 2012b). This can only be achieved by comprehensively considering all of the types of evidence described in sections 2.2.1 to 2.2.1.3 (Kane, 1990, 2004, 2013; Newton & Shaw, 2014; Messick, 1990; Wolming & Wikström, 2010). This evidence ought to be viewed in relation to the construct theory, rather than the methods used to observe it (APA, 1954; Messick, 1990). Divorcing validation from relevant theory risks creating measures that meet empirical standards but fail to provide any meaningful contribution to the scientific process (Bringmann & Eronen, 2016). In short, relying upon empirical or logical validation in isolation is failing to validate at all, instead presenting an incomplete picture insufficient to assess construct validity in any meaningful way.

In practice, construct validation supports both the construction and appraisal of psychometric tools. Using the construct theory, test constructors can design measures based on the hypothesised ways that attributes cause variations in responses (Bringmann & Eronen, 2016). In requiring multiple lines of enquiry to support a validation argument, construct validation also provides evidence based on different assumptions that may be used to triangulate the argument (Bringmann & Eronen, 2016; Maul, 2013). This means that evidence can be used in combination to capitalise on instances where one source overcomes the shortcomings of another. For example, if one source of support comes from correlations with a self-report measure of the attribute or an attribute theorised to relate to it, using a second source of evidence from observational data minimises risks of common method variance confounding inferences (Maul, 2013, 2017). Construct validation therefore provides more robust arguments to support inferences. A final advantage of the arguments-based approach to construct validation is that the argument need only be proportionate to its intended use (Kane, 2013; Messick, 1979). Measures considered limited in validity could still be used to advance knowledge or theory in ways enabling the construction of more sound measures, as long as their data are used in a manner appropriate to their validation argument (Bringmann & Eronen, 2016; Kane, 2013).

Despite the advantages of arguments-based construct validation, misuse is long-standing, widespread, and pervasive (Embretson & Gorin, 2001; Hogan & Agnello, 2004; Jonson & Plake, 1998; Maul, 2013, 2017; Messick, 1979; Wolming & Wikström, 2010). There is a persistent gap between theoretical understanding of validation and ways it is used in practice (Jonson & Plake, 1998; Kane, 1990, 2004; Maul, 2013; Wolming & Wikström, 2010). Failing to consider validity theory in test construction results in failure to define target attributes theoretically before construction begins (Cizek et al., 2008; Duckor, 2017; Embretson, 1998; Embretson & Gorin, 2001; Miller & Lovler, 2016) and single forms of evidence, such as logical or empirical, being used in isolation (Bringmann & Eronen, 2016; Kane, 1990, 2001; Messick, 1979; Travers, 1951). Validation practice continues to be atheoretical and therefore fails to acknowledge the role of causality in understanding how attributes theoretically cause variance in test scores, thus perpetuating the widespread failure to validate theories rather than tests.

Construct validity is the single most important concept underpinning psychometric quality. It requires that prior to test construction, detailed theory of a target attribute is developed (Sijtsma, 2012b) and that this theory is then used in line with the best practice arguments-based approach to guide test construction (Bringmann & Eronen, 2016). Following this, a range of evidence-gathering activities should be undertaken to justify the appropriateness of inferences to be made from test data (Bringmann & Eronen, 2016; Jonson & Plake, 1998; Kane, 1990, 2013; Messick, 1990; Wolming & Wikström, 2010), with the strength of the evidence required being determined by the measure's intended use (Bringmann & Eronen, 2016; Kane, 2013; Messick, 1979). Robust validation therefore requires that all evidence-gathering activities should contribute to the ultimate aim of strengthening the overall validation argument.

2.2.2 Reliability.

Reliability relates to the extent to which a psychometric tool accurately reflects the variable it intends to measure (Furr, 2011). Reliability concerns how well true scores are detected by measures, compared to error (Furr, 2011; Yang & Green, 2011). Reliability estimates assess how accurate, precise, or consistent measures are in detecting the signal of the true score (Borsboom, 2005; Furr, 2011; Miller & Lovler, 2016; Newton & Shaw, 2014; Raykov, 1997).

Although often misinterpreted as a discrete aspect of psychometric quality, the importance of reliability actually lies in its relation to validity. The reliability of a measure influences the statistical results of empirical validation and the ability to interpret data gathered using the measure post-validation (Furr, 2011). In both cases, reliability evidence provides information regarding the meaning of test scores both during validation and ongoing use (Newton & Shaw, 2014). Where reliability is low, the signal of the target construct is obscured by random error. Consequently, observed scores may underestimate or reflect an incomplete version of target variables. This questions inferences made from the data and therefore their validity (Furr, 2011). Moreover, reliability alone cannot be used to validate a measure. Measures can demonstrate high reliability estimates without accurately targeting the correct construct (Newton & Shaw, 2014). For example, a clock may return the same

time at the same moment each day, but it does not mean that that time is correct. This means that reliability and validity are inextricably linked in their relevance to validation, with both being necessary and neither being sufficient alone to guarantee psychometric quality.

Various approaches are available to statistically estimate reliability, including alternate forms and test-retest reliability (Furr, 2011; Furr & Bacharach, 2014; Miller & Lovler, 2016; Webb, Shavelson & Haertel, 2007). However, one approach is overwhelmingly dominant in psychometric validation literature: internal consistency, specifically estimated using Cronbach's alpha coefficient (Barchard & Hakstian, 1997; Brennan, 2011; Furr & Bacharach, 2014; Hattie, 1985; Maul, 2017; Raykov, 1997; Starkweather, 2012; Webb et al., 2007; Yang & Green, 2011). Cronbach's alpha repeatedly splits the items of a test in two and correlates the two resultant sets. It calculates the correlations between all possible halves of a test to estimate the level of relatedness amongst all items (Miller & Lovler, 2016). Cronbach's alpha assumes that the variance of true scores and error scores are normally distributed and uncorrelated, and when these assumptions are satisfied, provides a maximum estimate of the precision of a measure (Barchard & Hakstian, 1997). The appeal of Cronbach's alpha stems from it being relatively easy to apply and interpret, with consensus around the thresholds demonstrating low, moderate, and high internal consistency, offering an objective interpretation framework (Furr & Bacharach, 2014; Webb et al., 2007; Yang & Green, 2011). However, despite such widespread use, coefficient approaches to estimating reliability, and particularly their application, receive criticism.

Coefficient results are often errantly interpreted as assessing unidimensionality (Borsboom, 2005; Fisher, 2017; Furr, 2011; Furr & Bacharach, 2014; Hattie, 1985). A unidimensional measure targets a single construct, therefore providing validity evidence that relates to how scores may be interpreted. Interpreting internal consistency coefficients as indicating unidimensionality is theoretically incorrect, with claims regarding dimensionality requiring separate exploration using factor analysis (Fisher, 2017; Furr & Bacharach, 2014; Miller & Lovler, 2016). Cronbach's alpha actually rests upon the assumption that measures are unidimensional, so using this metric to evidence its own assumption is fundamentally illogical (Raykov, 1997; Yang & Green, 2011). The implication of

this errant practice is that even when used correctly, failure to appreciate the theoretical and empirical basis of coefficient alphas can lead to them being interpreted incorrectly and therefore undermining the validation argument they intend to support.

A further limitation of coefficient alphas is their high susceptibility to statistical artefacts. Including additional response options to item formats arbitrarily increases reliability estimates (Furr, 2011; Hambleton & Jones, 1993; Hattie, 1985), as do additional items (Borsboom, 2005). Even in cases of known multidimensional measures, high numbers of items always result in high internal consistency estimates, regardless of the number of constructs being measured (Barchard & Hakstian, 1997; Furr, 2011; Hattie, 1985; Raykov, 1997; Starkweather, 2012). Reliability coefficients should be interpreted in this light, with judgements made regarding their value in the context of the structural characteristics of the measure being investigated.

Despite the requirements of reliability coefficient procedures being relatively undemanding when compared with more modern techniques (Furr & Bacharach, 2014; Hambleton & Jones, 1993), the majority of validation practices fail to meet the required statistical assumptions or to acknowledge the implications of this (Furr & Bacharach, 2014; Starkweather, 2012; Yang & Green, 2011). For example, coefficient alphas assume that error is random, but it is highly unlikely that any psychometric tool would meet this assumption. Owing to the complex nature of human behaviour, systematic error is almost universally accepted to be an omnipresent issue in psychological measurement, a fact that reliability coefficients based on CTT are unable to account for (Furr, 2011; Hambleton & Jones, 1993). In assuming uncorrelated error and true scores, CTT sets an impossible standard (Borsboom, 2005). There is therefore support for an argument that coefficient alphas are unhelpful in assessing reliability, and therefore validity, during validation.

Reliability coefficients also assume that all items within a measure are equal in terms of variance, difficulty, and accuracy. However, modern developments in psychometric theory, including theoretical and empirical arguments grounded in item response theory, suggest that this assumption can also never be met or justified (Starkweather, 2012). Where the statistical assumptions of reliability coefficients are

violated, results may overestimate, accurately reflect, or underestimate reliability, dependent upon the nature and extent of those violations, in ways unknown to test users (Brennan, 2011; Yang & Green, 2011).

Finally, perhaps the most pervasive misuse of Cronbach's alpha stems from misunderstanding of the fundamentals of reliability and its relationship to validity. Despite reliability being a form of evidence that supports a validation argument, it is often applied and interpreted as a detached concept (Borsboom, 2005; Bringmann & Eronen, 2016). Reliability estimates are often used in isolation to support claims of high quality and robust measurement. However, reliability estimates are dependent upon the sample from which they are derived and as a result, cannot be claimed as a static property of a test, but rather its precision on a specific occasion when used with specific participants (Borsboom, 2005). Reliability evidence should therefore be combined with a broader assessment of validation evidence from a range sources for inference claims to be justified (Borsboom, 2005; Furr & Bacharach, 2014; Zumbo, 2007). Although reliability is a key component of validation, it should be carefully applied and interpreted as part of an integrated body of validation evidence rather than in isolation.

Robust reliability evidence is necessary for best practice validation. However, such evidence alone is insufficient for validation. Reliability estimates are flawed in their propensity to be influenced by statistical artefacts, including where statistical assumptions are not met. As a result, although an important part of validation, reliability estimates should be used with caution, with full consideration given to their limitations in the context of broader validation evidence. Despite this, reliability metrics are frequently misused, applied as either sufficient to validate a measure in isolation or without due regard to the statistical assumptions to be met, constituting a commonly encountered threat to the overall validation of measures.

2.2.3 Critique of classical test theory.

Despite being the overwhelmingly popular choice of psychometricians throughout the history of psychological measurement, CTT is not without criticism. In empirical terms, a major limitation of CTT concerns the assumption that true and

error scores are uncorrelated (Barchard & Hakstian, 1997; Borsboom, 2005; Furr, 2011; Hambleton & Jones, 1993; Jones & Thissen, 2007; Starkweather, 2012). This assumption requires that meaningless noise within test scores is random. However, consensus is that systematic error is ubiquitous in psychometry, due to relying on indirect indications of underlying variables. This means that the method used to measure these indirect indicators creates noise in itself (Maul, 2013). Variance resulting from method of measurement is systematic, in that it relates in some predictable way to the data it elicits (Miller & Lovler, 2016). As all psychometry rests upon measures, systematic method variance must always be present and the fundamental assumption of CTT therefore violated. Similarly, response biases such as demand characteristics, where participants exhibit tendencies to provide what they consider to be the 'correct' answer to questions based on subtle cues picked up as part of the research process (Orne, 1962; Rosnow, 2002; Sharpe & Whelton, 2016; Shaughnessy & Zechmeister, 2015), and practice effects, where repeatedly taking the same test changes performance based on previous experience (Miller & Lovler, 2016), also introduce systematic variance (Maul, 2013). Where systematic error is present, error scores cannot be assumed to be random, so this fundamental assumption of CTT cannot be upheld. Consensus among critics is therefore that CTT falls at the first hurdle and that all derived validation practices are meaningless as a result (Borsboom, 2005; Furr, 2011; Hambleton & Jones, 1993).

Theoretically speaking, another criticism of CTT-based validation approaches is that they rely upon non-falsifiable theoretical concepts (Borsboom et al., 2004). It is argued that focusing on the properties of measures themselves leads to validation practice being divorced from psychological theory regarding the causality of responses. As a result, measures are operationally defined and therefore may only be considered to measure that which it measures and nothing more. This makes inferences regarding psychological attributes or latent variables impossible (Duckor, 2017). It is argued that the theory of psychological measurement is circular and so cannot be falsified according to the principles of scientific enquiry (Borsboom et al., 2004; Brennan, 2011; Fan, 1998). Furthermore, even if such philosophical issues were hypothetically resolved, empirical evidence suggests the same issue.

Maul (2017) conducted a series of studies specifically intended to develop poor psychometric tools, with resulting inferences ranging from flawed to

meaningless. Maul (2017) manipulated test items by inserting nonsense words and entirely blank items and subjected their data to CTT-based validation. Theoretically, poor items should have resulted in poor validity evidence, as they reflected no psychological attribute whatsoever. However, nonsense items were consistently found to return similarly high estimates of reliability to standard test items using Cronbach's alpha and were included within factors proposed to reflect meaningful latent variables under factor analysis. The only hint of falsification was that criterion correlations for nonsense items were slightly lower than standard items, although both were objectively poor performers in this domain. Maul (2017) concluded that under CTT validation, almost any results may be accounted for as valid. This inability to falsify hypotheses under the scientific method undermines the rigour of CTT-based psychometric validation. This suggests that robust validation may require an even broader range of evidence than heretofore considered.

2.2.4 Summary of classical test theory.

Contemporary understanding is that validity reflects a unitarian construct conceptually derived from construct validity. Validity is the single most important aspect of psychological measures, as it ensures that inferences from test scores are appropriate. Validity speaks to the potential uses of a measure and the meaning of the results it elicits in this context. Concepts previously understood to be types of validity, such as content and criterion validity, are today considered types of evidence supporting an overall validation judgement. Factor structure and reliability evidence also support this judgement. Robust validity judgements draw on all of these concepts and related evidence, with sound understanding of the meaning and contribution of each, in line with the construct theory and relevant empirical evidence.

2.3 Item Response Theory

The theoretical underpinnings of target constructs are important to validation, but item response theory (IRT; Embretson, 1983; Furr, 2011; Hambleton & Jones, 1993; Jones & Thissen, 2007; Lord, 1952; Lord, 1953) prompted much more

detailed discussion of their relevance. Specifically, IRT considers the role of causality in psychological measurement and whether the causes of participant responses to individual items are in line with those suggested by the attribute theory (Borsboom et al., 2004; Maul, 2017; Furr, 2011; Newton & Shaw, 2014). Advocates of IRT claim that it surpasses all other forms of validation and may therefore be used as a single form of validation evidence in isolation (Borsboom et al., 2004; Furr, 2011). Advocates of IRT claim that the tick box application of CTT-based validation procedures leads researchers to neglect the role of theory in measurement, further contributing to a lack of understanding regarding causal relationships between attributes and participant responses (Borsboom, 2005; Borsboom et al., 2004; Bringmann & Eronen, 2016).

IRT recognises that observed scores reflect the true score and error, in addition to multiple other characteristics, and considers these issues to have a cognitively causal role (Bringmann & Eronen, 2016; Furr, 2011). Employing a statistical modelling technique, IRT estimates the likelihood that a specific participant will respond in a specific way to a specific test item, without these estimates being falsely inflated or deflated according to the validation sample (Furr, 2011; MacDonald & Paunonan, 2002). In practice, IRT estimates the difficulty of each item (or how difficult they are to endorse if a tool does not concern ‘correct’ answers), the ability of each item to discriminate individuals exhibiting low levels of the target attribute from those exhibiting high levels, and the likelihood that each item could be responded to correctly or endorsed as a result of random error or guessing. IRT models these estimates to provide an overall estimate of the psychometric quality of each item for a given participant (Furr, 2011). IRT provides evidence that can be used on a case by case basis to support (or contraindicate) the use of that particular item with a specific participant (Hambleton & Jones, 1993; Sijtsma, 2012a).

A second advantage of IRT is that it offers a route to validation for measures targeting an attribute for which there is no available theory. In such cases, IRT becomes the theory in practical terms by building a model of causal mechanisms in participant responses that contributes evidence to the validation argument in the absence of dedicated construct theory (Sijtsma, 2012a). IRT is also extremely flexible and can readily adapt to the complexity of target attributes and measures,

although more complex models naturally require higher levels of expertise to undertake the required modelling (Hambleton & Jones, 1993).

A final advantage of IRT is the strict and somewhat onerous set of assumptions to which it is held, offering a robust method of validation (Hambleton & Jones, 1993). However, this advantage also has its downside as this robustness is accompanied by additional demands that result in many researchers avoiding IRT in favour of the simpler statistics and more easily met assumptions of CTT (Hambleton & Jones, 1993). IRT demands significant expertise and specifies the thresholds for data to meet its assumptions, including requiring significantly larger sample sizes than CTT (Hambleton & Jones, 1993). Consequently, the impact of IRT has been limited (Furr & Bacharach, 2014) and when it is used, this is often inexpertly. Despite IRT models providing explicit requirements for data to be modelled, data gathered by researchers often fails to meet them (Sijtsma & Emons, 2013). The fact that data is often observed as failing to meet the required assumptions when either CTT or IRT are used as a basis for validation invites comparison and contrast between the two theories, as they both suffer from common misuse. As a result, IRT and CTT have drawn much comparison, both in theoretical and empirical terms.

2.3.1 A comparison of classical test theory and item response theory.

Although CTT-based validation procedures are most common in the psychometric literature, there is growing debate comparing them directly with IRT and its growing popularity and evidence-based support (MacDonald & Paunonan, 2002). In theoretical terms, the field generally favours IRT as its estimates are not contingent upon sample characteristics (Cook, Eignor & Taft, 1988; Rudner, 1983). CTT validation statistics are dependent upon the sample used to calculate them (Barchard & Hakstian, 1997; Borsboom, 2005; Rudner, 1983). As a result, CTT reliability estimates may reflect the reliability of a test when used with a sample identical to the one with which it was validated, with no conclusions possible regarding its precision when used with other samples (Borsboom, 2005). Sample independent validation data, such as that provided by IRT modelling, would result in more generalisable inferences (Hambleton & Jones, 1993; Hambleton, Swaminathan & Rogers, 1991; MacDonald & Paunonan, 2002). IRT approaches are further valued

theoretically for acknowledging the role of cognitive processes in psychometrics. IRT considers cognitive factors to be causal, therefore delivering measures that are more sensitive to true variance owing to their ability to disentangle systematic sources of variance such as cognitive biases (Bringmann & Eronen, 2016; Embretson, 1983, 1998, 2004).

Empirical evidence comparing IRT and CTT is mixed, but the weight of evidence appears to favour IRT. Empirical evidence suggests that IRT validation statistics are indeed independent of the influence of sample characteristics, with the same being untrue for CTT metrics (Fan, 1998; Hambleton & Jones 1993; Hambleton et al., 1991; MacDonald & Paunonan, 2002; Rudner, 1983; Tinsley & Dawis, 1977), but the picture is far from clear cut. Both Cook and colleagues (1988; MacDonald & Paunonan, 2002) and Fan (1998; MacDonald & Paunonan, 2002) found that estimates resulting from IRT and CTT were both unstable and sample dependent. The former study found CTT results to be more problematic, while the latter found the opposite. Additionally, other authors have found little difference between the two in terms of results (MacDonald & Paunonan, 2002). Many conclude that on balance, with little conclusive evidence of statistical superiority either way, CTT-based validation approaches are preferable simply due to their ease of application and accessibility for researchers (Cook et al., 1988; MacDonald & Paunonan, 2002).

It ought to be noted, however, that the evidence described above all rests on secondary analysis of existing validation data. Where sample dependency is the key issue, it is not possible to take these findings as generalisable, as the extent to which the findings are sample specific is unclear (Fan, 1998; MacDonald & Paunonan, 2002). In response, more recent studies have used simulated data to run the same comparative analyses, with this data having the added advantage of enabling the user to manipulate sample characteristics that are otherwise unknown. As a result, the ‘correctness’ of validation statistics can actually be compared with the true characteristics of the sample (MacDonald & Paunonan, 2002). Simulated data suggests that while CTT and IRT approaches return similar results relating to the ability of measures to estimate the target attribute and difficulty of individual test items, IRT is superior in estimating the ability of test items to discriminate high and low levels of the target attribute (MacDonald & Paunonan, 2002).

Overall, the weight of evidence suggests that although both are dependent on sample characteristics to an extent, IRT-based validation statistics are less so. Under certain circumstances (as dictated by the validation sample used), the accuracy of CTT-based validation statistics may therefore be unreliable. Having said this, the requirements of IRT, both in terms of assumptions to be met and expertise required within the research team, are much higher than for CTT, leading to limited uptake (Embretson & Gorin, 2001; Furr & Bacharach, 2014). As a result, CTT remains the dominant validation approach, although the impact of IRT has been increasing slowly since its inception (Fan, 1998; Embretson, 2004; Furr, 2011). Overall, the lack of clarity in empirical findings relating to IRT versus CTT validation practices suggests that best practice would make use of both approaches during an ongoing process of validation, in order to provide the strongest possible validation argument.

2.4 Generalisability Theory

Generalisability theory (Cronbach, Gleser, Nanda & Rajaratnam, 1972; Furr, 2011; Furr & Bacharach, 2014; Miller & Lovler, 2016) derives from that of CTT (Brennan, 2011; Furr, 2011). CTT assumes that observed test scores consist of the true score accurately reflecting the attribute under study and random error, thus failing to acknowledge or account for systematic error (Furr, 2011; Hambleton & Jones, 1993; Webb et al., 2007). Generalisability theory accepts that random error is but one source of variance within an observed test score and that reliability estimates must also account for other sources, including systematic error (Brennan, 2011; Furr, 2011; Furr & Bacharach, 2014; Suen & Lei, 2007; Webb et al., 2007). Sources of variance are termed facets and might include characteristics such as the influence of common methods, bias introduced by using different scorers, and the effects of different administration contexts and participants (Furr, 2011). Generalisability theory seeks to extricate the contribution of each facet and interactions amongst them, and enable decisions as to which facets are relevant to deciding how the true and observed scores are best defined. This enables researchers to ascertain the extent to which the true score is detected versus other facets, and therefore the appropriateness of inferences made from it (Brennan, 2011; Furr, 2011; Furr & Bacharach, 2014; Miller & Lovler, 2016).

Generalisability theory has developed in parallel with IRT, although the latter is more well-known, due in part to generalisability theory's CTT roots and its relationship to the issue of precision or reliability (Brennan, 2011). Superficially, generalisability theory offers a detailed and complex elaboration on CTT (Brennan, 2011). Being prized for their accessibility and ease of application and interpretation, CTT-based statistics remain more popular, as generalisability theory offers researchers a demanding journey into statistics potentially deemed prohibitively onerous by comparison (Brennan, 2011). Consequently, generalisability theory is often overlooked in modern psychometry in favour of both IRT and CTT.

Beyond this, there is theoretical precedent for comparing IRT and generalisability theory directly. IRT explores measures at the item level, being concerned with predicting the true score relative to the causal mechanisms associated with responding to test items (Brennan, 2011; Suen & Lei, 2007). Generalisability theory explores measures at test level, seeking to predict observed scores without reference to latent variable models (Brennan, 2011; Suen & Lei, 2007). Brennan (2011) summarises this issue succinctly as IRT being to generalisability theory as individual trees are to a forest; focusing on either renders the other out of scope. As a result, tentative steps have been taken towards combining IRT and generalisability theory, although progress in this regard is limited to date and so IRT and generalisability theory are generally still applied in parallel within contemporary practice (Brennan, 2011).

Generalisability theory improves upon CTT-based validation in estimating precision or reliability (Brennan, 2011; Suen & Lei, 2007). However, much like IRT, it has yet to come close to the popularity of CTT, with the main reason for this being that the procedure is less accessible and requires appreciation of the detailed, complex, and technical statistical framework within which it sits (Brennan, 2011). Despite this, as what many believe to be an improvement on basic CTT-based validation, generalisability would still be expected to feature in a best-practice, ongoing approach to arguments-based psychometric validation.

2.5 The Ethics of Validation

The role of ethics in validation is a source of significant controversy (Cizek et al., 2008; Furr & Bacharach, 2014; Zumbo, 2007). Messick (1979, 1990, 1995) argued that as the validity of a measure pertains to the ways it is intended to be used, responsible validation requires consideration of related ethical implications (Jonson & Plake, 1998; Miller & Lovler, 2016; Newton & Shaw, 2014). For example, the threshold for an acceptable strength of validation argument may only be understood relating to its intended purpose and the likelihood and consequences of potential misuse or misunderstanding; the consequences of a poor measure of colour preference may be inconsequential when compared with a poor measure claiming to predict future work performance.

The controversy associated with this position relates not to the soundness of reasoning, but to the implications for science and research more broadly. Considering ethical implications in psychometrics has been described as risking the politicisation of science, moving beyond the creation of knowledge and understanding towards the consideration of whether certain types of knowledge ought to be created at all (Furr & Bacharach, 2014; Lees-Haley, 1996). Lees-Haley (1996) argued that this inappropriately introduced personal values and individual preference into science. This counter argument presupposes that science must be value-free and objective in line with the values of positivism. As will be discussed in detail in chapter 4, positivism is not the paradigm within which this thesis sits and so the ethical implications of measures will be given further consideration in coming chapters.

The ongoing controversy relating to the ethics of validation has resulted in it being given little consideration in published tests (Cizek et al., 2008; Newton & Shaw, 2014). This is perhaps surprising as professional psychometric standards recognise the importance of this issue and recommend its consideration in validation (APA, 2017; Miller & Lovler, 2016). As a minimum, it is therefore recommended that potential ethical issues associated with test use be acknowledged transparently when reporting validation so that users may consider them in their own inferences.

2.6 Conclusions Regarding Psychometric Best Practice

Although debate continues in psychometry, significant progress and some level of consensus has been achieved over the last century. Generally speaking, validity is viewed as a single, unified concept central to psychometric quality, and is agreed as best investigated by an ongoing and comprehensive validation approach using multiple forms of evidence pertaining to validity, reliability, and dimensionality (Newton & Shaw, 2014). Such an approach is distilled well by Kane's arguments-based approach to validation (1990, 2001, 2004, 2013; Newton & Shaw, 2014), which suggests that validation ought to be viewed as a spectrum of evidence and that this evidence may vary in nature and strength for different measures and purposes (Kane, 2013; Messick, 1990; Newton & Shaw, 2014). For example, evidence might suggest that a measure is sufficiently robust to be used for purposes of personal development, but not as a basis of making life-changing decisions regarding recruitment or promotion. Similarly, measures demonstrating weak validation evidence may still have value in generating theories for further testing and elaboration (Bringmann & Eronen, 2016; Kane, 2013).

The arguments-based approach to validation also responds to claims of circular and non-falsifiable validation in psychometry by conceptualising validation as an ongoing process of ruling out confounding explanations, rather than as a single, absolute event (Newton & Shaw, 2014). In this sense, arguments-based validation returns to the fundamental scientific method of generating hypotheses to be tested and potentially falsified regarding a measure's ability to detect and accurately quantify an attribute (Bringmann & Eronen, 2016). Having said this, there is also some consensus in describing the shortcomings of psychometrics, which most often lie in their application rather than theory. Regarding statistical validation, the overreliance on CTT and the belief that certain CTT-based statistics alone are sufficient to validate measures regardless of their outcome, are agreed to be prevalent and problematic issues (Cronbach & Meehl, 1955; Duckor, 2017; Maul, 2017). Regarding theoretical aspects of validation, the absence of clearly defined target attributes and theoretical frameworks through which to understand empirical validation results also constitutes a consistent violation of psychometric assumptions (Borsboom, 2005; Bringmann & Eronen, 2016; Cronbach & Meehl, 1955;

Embretson, 1983; Embretson & Gorin, 2001; Kane, 2013; Maul, 2017; Newton & Shaw, 2014).

On balance, it appears that while psychometry offers clear guidance as to best practice, there remains a persistent gap between psychometric theory and its application for research and other purposes, owing mainly to a determination to adhere to concepts generally considered outdated and inferior to their modern counterparts (Cizek et al., 2008; Hogan & Agnello, 2004; Jonson & Plake, 1998; Newton & Shaw, 2014; Wolming & Wikström, 2010). As such, it would be advantageous for research relying heavily upon psychometry to consider the strength of validation evidence before embarking on activities to push the boundaries of current understanding, boundaries that may yet be undermined should the psychometric evidence to date prove unsatisfactory.

Regarding implications for the measurement of professionalism, discussions within chapter 1 suggest that this is as subject to the psychometric theory-practice gap as any other application. Although not specifically targeting the methodological technicalities of psychometric theory, reviews evaluating the quality of existing measures of professionalism cite major shortcomings and call for urgent improvements (e.g. Lynch et al., 2004; Veloski et al., 2005; Jha et al., 2007). The more general shortcomings in validation practice discussed within this chapter mirror these findings and further support the need for a robust and systematic evaluation of psychometric practice in the measurement of professionalism before further research based on psychometric data may be undertaken.

2.7 Chapter Summary

This chapter discusses the history and contemporary understanding of validity theory and its implications for best practice in the validation of psychometrics. It concludes that although validity theory is well-developed and comprehensive, providing theoretical and empirical evidence for robust and ongoing arguments-based validation, this is rarely observed in practice. Evidence discussed in chapter 1 suggests that the picture is similar for professionalism-specific psychometric tools. In the context of this thesis, it is important to evaluate the

validation arguments for extant psychometric tools targeting professionalism before they may be used for further research, against the key technical criteria cited within this chapter. The next chapter will discuss the research questions to be explored to this end.

Chapter 3: Research Aims and Objectives

The literature reviewed in chapter 1 highlights the importance of professionalism in today's society and its relevance for trainers and educators, employers and regulators, and the general public. The historical context of professionalism provides insight into why it attracts significant controversy, with different parties vehemently defending their different viewpoints as to what it is, and connotations of social (in)justice colouring debate. These issues were discussed specifically in relation to their implications for the measurement of professionalism, as a means of benchmarking performance. Such measurement would form the basis of decisions regarding qualification, employment, promotion, and fitness to practise throughout the career of an individual and across the myriad sectors with an interest in this issue, and so would necessitate measurement of the highest quality. Traditional educational assessments used to measure professionalism to date have been heavily criticised, resulting in increased interest in measuring the construct as a psychological attribute or characteristic. Psychological measurement, or psychometrics, is a challenging area of measurement science with a rich, detailed, and technical theory underpinning it. A rigorously validated measure of professionalism developed in line with theory could potentially provide the robust assessment reported as lacking in the teaching and learning of professionalism to date.

Measuring professionalism psychometrically introduces a whole new arena of debate; namely that regarding best-practice validation of psychological measures. The field of psychological measurement speaks to the argument made to support the drawing of inferences from the data collected using a specific measure. The strength of this argument indicates the level of validity of conclusions, such as those required within the context of professionalism relating to individual employment. Best practice in forming this argument requires an intensive, ongoing approach to collecting a range of validity evidence, from sources relating to classical test theory, such as content and criterion validity, factor structure and dimensionality, and reliability; and those relating to more contemporary developments in psychometry, such as item response theory and generalisability theory. However, there is evidence

of a pervasive, almost habitual tendency for general practice in psychometry to fall considerably short of these standards.

3.1 Aims and Objectives

In order to address the potential for the psychological measurement of professionalism, the aims of this thesis are:

1. To review the current state of the art in measuring professionalism as a psychological construct;
2. to identify actions required in order to increase rigour in the measurement of professionalism;
3. to contribute to and stimulate the field of psychology in its study of professionalism; and
4. to lay the foundations for a psychological theory of professionalism that may generate future directions for further study.

These outcomes will be achieved via the following objectives:

- I. To undertake a systematic review of the methodology of the psychological measurement of professionalism, with reference to contemporary psychometric best practice.
- II. To respond to the identified need to develop a theoretically grounded definition of professionalism by providing empirical evidence regarding the construct from a phenomenological viewpoint.
- III. To use empirical data and existing psychological theory to develop a new theoretical model of professionalism as a dynamic, interpersonal phenomenon.
- IV. To conduct a pilot stakeholder consultation regarding the new model of professionalism to explore its potential acceptance by, and utility

for, interested professional communities, as a potential basis for a new psychological measurement tool.

These aims and objectives will be achieved through a series of empirical research studies, addressing the research questions listed below:

- A. What is the evidence that validation undertaken of existing measures of professionalism published in all occupational sectors between 2007 and 2017 meets the standards required by best practice psychometric theory? (Chapter 5)
- B. What aspects of construing are shared by a group of adults from a range of occupational sectors when considering the professionalism of others? (Chapter 6)
- C. What are the characteristics of shared subjectivity in the perceptions of professionalism amongst a group of adults from a range of occupational sectors? (Chapter 7)
- D. What are the priority feedback points identified through stakeholder consultation regarding the proposed model of interpersonal professionalism? (Chapter 8)

3.2 Thesis Structure

The following chapters will describe and discuss research undertaken in the delivery of these aims and objectives. Chapter 4 outlines the methodological approach of the thesis, discussing the paradigmatic, methodological, and methods-related issues considered and decisions made in resolving them. Chapter 5 describes a methodological review of the measurement of professionalism, using a systematic search strategy and evaluating validation practices used in the development of published measures of professionalism against rigorous, best-practice psychometric criteria (objective I).

Chapters 6 and 7 describe empirical research seeking to understand the concept of professionalism from the perspective of individuals experiencing it (objective II). Chapter 6 describes a study exploring the way that professionalism is construed ideographically by individuals using the repertory grid technique grounded in the personal construct theory of personality. It explores the way that professionalism is actually experienced by recipients, rather than the way that they have learned to describe it within their experienced history. Chapter 7 describes a study exploring professionalism as a subjectively viewed construct, using the Q sort method derived from Q methodology. It discusses the way that individuals perceive the construct of professionalism by bringing objective form to their tacit, subjective perspectives. Both chapters 6 and 7 use quantitative analytical techniques to identify areas of shared construing and subjectivity across multiple individuals that may form the basis of a general model of professionalism.

Chapter 8 describes a new model of professionalism as a dynamic and interpersonal psychological process based on the empirical findings of chapters 6 and 7, and relevant existing psychological theory (objective III). It also describes a study providing a conceptual example of how this model might be elaborated using the expertise of professional communities with a key stake in the understanding and measurement of professionalism (objective IV). The findings of this stakeholder consultation are discussed with reference to how they might be applied within a future programme of research aiming to undertake the best-practice validation of a new measure of professionalism. A general discussion of the outcomes of the research described in chapters 5 to 8 is provided in chapter 9, along with overall conclusions and recommendations.

3.3 Ethical Issues

The research undertaken in thesis did not raise any significant ethical issues and all studies were approved by the University of Liverpool research ethics sub-committee for the Institute of Psychology, Health, and Society. The study described in chapter 5 did not require ethical approval. The pilot and main studies described in chapter 6 were approved under reference 1922, the pilot and main studies described in chapter 7 were approved under reference 1922 (amendment), and the study

described in chapter 8 was approved under reference 4987. Copies of the applications, related documentation, and approval confirmation letters are provided in appendix A. In line with the approvals provided, the student researcher and thesis supervisors had received dedicated ethics training provided by the University of Liverpool prior to undertaking any research activities.

3.3.1 Interviews and focus groups.

Where face-to-face interviews were undertaken, the preferred interview venue was a University of Liverpool facility. However, where requested by the participant, public places such as coffee shops or the facilities of another university were also used. Where the latter were used, participants were informed of the inability of the researcher to guarantee that conversations would not be overheard, and the offer of a private University of Liverpool room reiterated. Schedules were produced to guide interactions, which were submitted in full to the research ethics sub-committee for approval.

Interviews could not be undertaken anonymously. However, where recordings or extracts of recordings were transcribed, transcripts were redacted to remove personal information in case of future requests to re-use the data. During focus groups, the identities of participants were also known by other participants. In these cases, the participant information sheet, focus group ground rules, and interview schedule all requested that the confidentiality of others in the session be respected and that no contributions were discussed outside of the session.

3.3.2 Online data collection.

Where data was collected using online questionnaires, participants were invited to complete studies using their own electronic device or computer. Questionnaires were presented to participants using specialist software as a webpage with a submission confirmation at the end. The study described in chapter 6 was displayed using the University of Liverpool's Qualtrics software subscription (<https://www.qualtrics.com/uk/>), in compliance with the ethical approval provided for that study. The study described in chapter 7 was displayed using version 1.0.4 of

the HTMLQ specialist software package associated with Q methodology (Aproxima Gesellschaft für Markt- und Sozialforschung Weimar, 2015), in compliance with the ethical approval provided for that study. The coding used to programme the HTMLQ software is available in appendix E. Online data collection was undertaken anonymously. No identifying information was gathered from participants and following submission, their data was stored under an assigned participant number.

3.3.3 Participant information and consent.

Participants were provided with an information sheet describing the study prior to taking part, which was submitted in full to the research ethics sub-committee for approval. Participants contacting the research team directly were provided with a copy of the information by email before deciding whether to take part. Participants attending interviews or focus groups were also provided with a hard copy of the information before the interview commenced. Participants completing online data collection without prior contact with the research team were presented with the information electronically before commencing the questionnaire.

Participants provided fully informed consent to take part in the study prior to commencing data collection. For interviews or focus groups, participants read and completed a hard-copy consent form. For online participants taking part in the study described in chapter 6, the consent information was displayed electronically with a checkbox system preventing their progressing any further in the questionnaire before completing it. For the study described in chapter 7, in the absence of checkbox functionality, a statement that clicking the 'next' button indicated consent to take part in the study was included. The information provided informed potential participants of the limitations of their consent, specifically the time after which they would no longer be able to withdraw their data.

3.3.4 Data management.

Data collected took the form of audio recordings and hard copy field notes from interviews or focus groups, and electronic data for online participation. All electronic data, including audio recordings and questionnaires completed online were

stored on the University of Liverpool secure server. Hard copies were stored within a lockable office in a University of Liverpool campus building. In line with University of Liverpool policy, data was stored for a period of five years after conclusion of the programme of study. Upon expiry of the retention period, all hard copies of data were destroyed and all electronic copies permanently deleted.

Part II: The Methodological Approach

Chapter 4: Methodology

The literature review reported in chapters 1 and 2 identifies potential shortcomings of previous professionalism-related research, including poor psychometric quality of quantitative measures resulting from a lack of theory in which to ground construct definitions (Goldie, 2013; Jha et al., 2007; Kuczewski, 2006; Lynch et al., 2004; Veloski et al., 2005; Wilkinson et al., 2009). In response, a methodological review was undertaken to establish the psychometric quality of extant measures of professionalism (see chapter 5). This review concluded that no measure could be recommended for use due to a lack of sound, theory-led definitions of professionalism. In order to address this issue for future measurement, a theory-building approach was deployed to define the construct of professionalism using research rooted in the mixed methods research paradigm to establish a phenomenologically grounded understanding of professionalism. This chapter explains and discusses the methodological approaches taken to achieve this, including an exploration of the philosophical and theoretical underpinnings of each approach and measures taken to maximise integrity in their application.

4.1 Mixed Methods Research

Mixed methods research (MMR) has experienced huge growth in popularity over recent decades (Bergman, 2008; Cameron, 2011; Tashakkori & Teddlie, 2010a; Teddlie & Tashakkori, 2010). MMR rejects the dichotomy of using purely qualitative or quantitative research methods alone in favour of employing the blend of methods most suited to achieving rounded understanding of complex phenomena (Burke Johnson & Onwuegbuzie, 2004; Cameron, 2011; Feilzer, 2010; Tashakkori & Teddlie, 2010a, 2010b). However, MMR remains comparatively new within psychology and so controversy remains around its constitution and best practice (Cameron, 2011; Teddlie & Tashakkori, 2003).

4.1.1 Philosophy and epistemology.

Philosophy relates to the way an individual views the world and the assumptions upon which that view is based (Cameron, 2011; Creswell & Plano Clark, 2011; Mertens, 2007; Morgan, 2007; Neuman, 2000; Teddlie & Tashakkori, 2010). In research terms, this is the interrelated set of assumptions underpinning the research paradigm, including ontology and epistemology (Burke Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2011; Mertens, 2007). Ontology concerns beliefs about the nature of reality: a single, objective reality subject to a single truth versus multiple realities constructed by the multiple individuals experiencing them (Burke Johnson & Onwuegbuzie, 2004; Guba, 1990; Mertens, 2007; Teddlie & Tashakkori, 2010). Epistemology speaks to the understanding of knowledge and how it is generated. The former ontological stance is associated with an epistemology of objective observation to derive general rules that transcend contextual issues. The latter ontological stance is associated with an epistemology that knowledge is a product of the observer's interpretation of reality, from which general laws may not be extrapolated (Burke Johnson & Onwuegbuzie, 2004; Guba, 1990; Mertens, 2007; Morgan, 2007; Teddlie & Tashakkori, 2009, 2010).

These two philosophical standpoints underpin different scientific approaches. Understanding the philosophical viewpoint of a researcher provides context to their work, allowing readers to challenge their interpretation and application of methodologies in this context (Dougherty, Slevc & Gran, 2019). Historically, psychological research has developed along two opposing pathways resulting from the opposing philosophical viewpoints described above: the quantitative and qualitative paradigms (Burke Johnson & Onwuegbuzie, 2004). Advocates of either approach adhere to a single paradigm, claiming that one philosophy cannot be reconciled with the other, a viewpoint known as the incompatibility thesis (Burke Johnson & Onwuegbuzie, 2004; Creswell, 2010; Teddlie & Tashakkori, 2003, 2009, 2010). Purists adhering to the incompatibility thesis and arguing for the superiority of their chosen paradigm do so with a singular vehemence that has led to this debate being described as the paradigm wars (Cameron, 2011; Feilzer, 2010; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003, 2009).

The quantitative paradigm, associated with the positivist epistemology, is based on the assumption of a single, objectively verifiable reality (Burke Johnson & Gray, 2010; Burke Johnson, Onwuegbuzie & Turner, 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Guba, 1990; Neuman, 2000; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003, 2009). Quantitative methodologies seek absolute truths using predominantly numeric data (Burke Johnson & Onwuegbuzie, 2004; Creswell & Creswell, 2018; Neuman, 2000; Teddlie & Tashakkori, 2003, 2009). They use deductive logic to form general theories about the world, derive specific hypotheses from them, and then collect data and use inferential statistical methods to test those hypotheses (Burke Johnson & Gray, 2010; Creswell & Plano Clark, 2011; Morgan, 2007; Neuman, 2000; Teddlie & Tashakkori, 2003, 2010). As the results of quantitative research relate to a single reality, findings can be extrapolated from to draw generalised conclusions about the same phenomenon in different social, environmental, or temporal contexts (Burke Johnson & Onwuegbuzie, 2004; Denzin & Lincoln, 2008; Neuman, 2000; Teddlie & Tashakkori, 2009).

The qualitative paradigm, derived from the constructivist epistemology, asserts that knowledge can only capture a representation of a situation or object through the lens of the individual observing it. Qualitative research findings are embedded within the socially constructed environment, representing one of multiple subjective realities (Burke Johnson & Onwuegbuzie, 2004; Burke Johnson et al., 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Guba, 1990; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003). There are close links between constructivism and subjectivity meaning that qualitative data is idiographic, pertaining to phenomenological human experience described in textual or narrative form (Burke Johnson & Gray, 2010; Burke Johnson & Onwuegbuzie, 2004; Creswell & Creswell, 2018; Denzin & Lincoln, 2008; Teddlie & Tashakkori, 2003, 2009). The idiographic nature of qualitative research means that findings apply only to the specific individuals in the specific context in which they were studied, and so do not provide generalisable insights (Denzin & Lincoln, 2008). Qualitative analytical approaches are underpinned by inductive reasoning, whereby specific observations are used to build more generalised theories to account for them

(Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2009).

More recently the so-called third research paradigm of MMR has entered the war. It moves beyond the previously tribal approaches, claiming that qualitative and quantitative research are two sides of the same coin that may be used in tandem to offer advantages inaccessible to either employed in isolation (Burke Johnson & Onwuegbuzie, 2004; Burke Johnson et al., 2007; Cameron, 2011; Creswell & Plano Clark, 2011; Feilzer, 2010; Neuman, 2000; Teddlie & Tashakkori, 2003, 2009). Grounded in the pragmatist epistemology, MMR asserts that while an objective reality does exist, fully understanding it in the context of human subjectivity requires a blend of both objective/quantitative and subjective/qualitative methodologies (Biesta, 2010; Teddlie & Tashakkori, 2009, 2010).

Pragmatism rejects the incompatibility thesis (Teddlie & Tashakkori, 2003, 2009), providing an alternative to the objective-subjective dichotomy known as inter-subjectivity. Inter-subjectivity acknowledges the role of social context in research (Burke Johnson & Gray, 2010; Morgan, 2007; Teddlie & Tashakkori, 2010), stating that observations are recorded via the dual lenses of the researcher and participant viewpoints, thus requiring a reflexive research approach that acknowledges and explores this (Morgan, 2007). Using intersubjectivity to explore areas of consensus and contention, researchers can establish shared interpretations of results with participants before they are more widely communicated. This creates a bridge between the subjectivity of the participant and the objective world that they share with other individuals (Morgan, 2007). Supporting this bridge is the target of MMR, achieved by applying the most effective and appropriate methods to the scenario being studied (Biesta, 2010; Creswell & Plano Clark, 2011; Feilzer, 2010; Miller, 2006; Morgan, 2007; Teddlie & Tashakkori, 2009).

Pragmatist MMR rejects the 'either/or' approaches of positivism and constructivism, instead acknowledging that the underlying paradigm of research may actually be multiple paradigms (Biesta, 2010; Burke Johnson et al., 2007; Tashakkori & Teddlie, 2010b; Teddlie & Tashakkori, 2010, 2012). This pluralist approach is associated with an iterative approach to knowledge-creation, whereby observations about a phenomenon lead to the creation of theories that are then tested,

reviewed, refined, and re-tested using the methods most suited to answering the research question at a given point in this cycle (Burke Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010a, 2010b; Teddlie & Tashakkori, 2012). MMR therefore provides opportunities for both inductive and deductive logic, supporting both theory building and testing. A key feature of MMR is that researchers can choose to employ one, the other, or multiple paradigms depending on their utility to the question being explored and the phase of the cycle currently being undertaken (Burke Johnson & Onwuegbuzie, 2004; Burke Johnson et al., 2007; Teddlie & Tashakkori, 2003, 2009, 2010). Researchers can pick and choose from methods associated with each paradigm as practicality dictates, an opportunity known as methodological eclecticism (Teddlie & Tashakkori, 2010, 2012). Consequently, MMR offers greater flexibility than purist quantitative or qualitative approaches and is therefore particularly well-suited to tackling complex issues occurring in social contexts that involve both subjective and objective aspects (Tashakkori & Teddlie, 2010b).

4.1.2 Advantages of MMR.

The pluralism of MMR affords researchers the flexibility to choose from the breadth of existing methodologies within both quantitative and qualitative research, enabling them to choose the method most suited to the research question, rather than adapting the question to suit their paradigm (Tashakkori & Teddlie, 2010b). This is particularly advantageous where complex issues, such as social and behavioural phenomena, would otherwise require breaking down into smaller constituent parts that may be less useful to end users, such as policy makers (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010b). MMR offers opportunities to answer questions that neither quantitative nor qualitative approaches alone could (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2003), eliciting results providing more than the sum of their parts (Creswell & Plano Clark, 2011; Miller, 2006). Qualitative research elements provide greater depth while quantitative elements provide breadth of findings, and MMR thrives in acknowledging these findings may be divergent or even appear contradictory (Burke Johnson & Gray, 2010;). Accepting such

divergence acknowledges the complexity of understanding subjective beings in an objective world (Greene & Caracelli, 2003).

MMR may also deliver more valid inferences from research data, due to its capitalising on the strengths of both qualitative and quantitative methods of enquiry and offsetting their respective weaknesses (Burke Johnson & Onwuegbuzie, 2004; Creswell & Plano Clark, 2011; Greene & Caracelli, 2003; Teddlie & Tashakkori, 2003, 2009). Divergent results from MMR may therefore be interpreted as providing triangulation to support validity, as both viewpoints are used to create a more complete picture of the phenomenon under study (Burke Johnson & Gray, 2010; Burke Johnson et al., 2007; Teddlie & Tashakkori, 2010). Pragmatism views divergent evidence as a stimulus for refining understanding through both induction and deduction (Burke Johnson & Gray, 2010; Burke Johnson et al., 2007; Teddlie & Tashakkori, 2010). Rather than adhering to either of these logics alone, pragmatism aligns with abduction, which involves undertaking induction or deduction according to the practical needs of a study or both in a continuous flow of one then the other (Burke Johnson & Gray, 2010; Morgan, 2007; Teddlie & Tashakkori, 2012; Reichertz, 2004). Abduction enables observations to inform theories and theories to inform observations within research studies or across a research program, depending on what is most useful at that time (Morgan, 2007; Teddlie & Tashakkori, 2010, 2012). Abduction therefore allows a researcher to pursue a research question in the most practically effective manner, rather than in a way dictated by the direction of purist qualitative or quantitative reasoning.

MMR is also cited as advantageous in communicating research findings to a broad audience. MMR is inherently socially grounded, focusing on intersubjective consensus and contention. As a result, researchers ought to take steps to create shared understanding of research findings between themselves and participants before they are more widely communicated. This supports greater research impact due to enhanced credibility of interpretations (Morgan, 2007), but also by providing research findings in a format closely aligned with everyday human experience. Humans understand the world around them both in numeric and narrative terms, so research using both of these may make more intuitive sense to non-specialist readers (Creswell & Plano Clark, 2011).

4.1.3 Controversies.

Although methodological eclecticism is often described as a strength of MMR, offering the flexibility to target complex questions, it has also been cited as a flaw. The guiding principle of utility governing methodological decisions in MMR recommends that researchers opt for the best methods for the job. Some argue that this principle could be used to justify any methods, regardless of their scientific rigour or rationale (Miller, 2006). Without a sound philosophical framework, and the ontological and epistemological assumptions it articulates, research is argued to be non-falsifiable as researchers may call upon any philosophy depending on the methods they would like to use that day (Cameron, 2011; Greene & Caracelli, 2003; Denzin & Lincoln, 2008). Falsifiability is a key issue in research with a widely held view being that unless something may be falsified, it cannot be considered scientific fact (Popper, 1972). Overall, eclecticism is the major point of contention for critics of MMR, with their argument generally being grounded in the assumption that eclecticism negates the scientific rigour required for knowledge creation.

In response, the MMR community have paid close attention to issues of research integrity and rigour (Cameron, 2011). Guidelines have been developed to ensure mixed methods are applied appropriately, with the foremost guiding principle being synergy (Biesta, 2010; Teddlie & Tashakkori, 2010). The synergistic integration of quantitative and qualitative research methods enables the discovery of information that would not be accessible through either method alone (Feilzer, 2010; Miller, 2006). A second crucial issue in the integrity of MMR is referred to as *methodological bilingualism* (Teddlie & Tashakkori, 2003), meaning that to ensure that methods derived from both paradigms are used appropriately, the researcher should demonstrate sufficient expertise in each to be considered akin to a fluent speaker in language terms (Cameron, 2011). These standards have served to enhance the standing of MMR within psychology and enabled a framework for critique that drives further quality enhancements in the field, and issues relating to their delivery are discussed in section 4.1.4.

4.1.4 Practical considerations.

Compared to purist paradigms, MMR is relatively new, so there remain unresolved points of contention relating to the pursuit of rigour (Creswell, 2010; Miller, 2006). MMR must include a minimum of one qualitative and one quantitative component (Bergman, 2008; Creswell & Plano Clark, 2011; Miller, 2006). These components must be employed to enable synergy, but the means by-which such synergy be delivered are debated (Creswell & Plano Clark, 2011; Miller, 2006). The majority of recommendations require integration at all levels, from philosophy through theory to data collection, and analytic and interpretative methods, in order to deliver the full potential of MMR (Cameron, 2011; Feilzer, 2010; Teddlie and Tashakkori, 2010). This goes beyond undertaking multiple methods in parallel (Denzin & Lincoln, 2008; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2012). Mixed methods findings uncover different pieces of the same jigsaw, leading to more comprehensive understanding of complete phenomena surpassing the one-dimensional understanding delivered by qualitative or quantitative methods alone (Denzin & Lincoln, 2008). However, delivering synergy through integration at all levels of research planning, design, and execution is challenging, with claims that little research has yet delivered it (Bryman, 2008; Cameron, 2011; Feilzer, 2010; Kelle & Erzberger, 2004). This challenge may therefore be best met using approaches in which theory, methodology, and methods are developed together in ways that make them inextricable.

However, in the absence of an approach developed in this way, alternative routes to delivering synergy are available (Burke Johnson et al., 2007). The seminal MMR writings of Teddlie and Tashakkori (2010, 2012) state that integration must be sufficient within a single research study to enable a minimum of one complete cycle of inductive reasoning. However, others suggest that pragmatism allows for one form of research to inform the other in a more sequential fashion across multiple studies (Morgan, 2007). As a minimum, it appears that a study or programme must take and justify as many steps as possible towards integration to negate the use of multiple methods in parallel (Denzin & Lincoln, 2008; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2012). A researcher should ensure that they are providing sufficient integration to deliver synergistic outcomes rather than multiple ones, as best meets the needs of their research question.

Overall, guidelines recommend that the primary focus when determining the integrity of MMR is the research question at hand, and that methodological decisions must be made based on pragmatic consideration of what would deliver the most useful and meaningful findings (Biesta, 2010; Burke Johnson & Onwuegbuzie, 2004; Niglas, 2010; Teddlie & Tashakkori, 2010). Ultimately, MMR relies upon a perpetual return to the question of what methods would be most effective in answering the research question.

4.1.5 The approach of this thesis.

The approach of this thesis was grounded in pragmatism using a bottom-up orientation. The issue of professionalism is inherently applied being a socially constructed issue, and so the usefulness and meaningfulness of findings were a major factor in research planning (Biesta, 2010; Burke Johnson & Onwuegbuzie, 2004; Niglas, 2010; Teddlie & Tashakkori, 2010). A pragmatic approach was taken to identifying the methodologies and methods best suited to answering each of the research questions under study, resulting in the completion of a qualitative systematic methodological review, two mixed methodology studies, and one use of qualitative focus group data, to contribute to an overall pragmatic and abductive theory-building approach.

4.2 Theories, Methodologies, and Methods

This section will discuss the theories informing methodological decisions within this thesis and discuss the theoretical and practical implications of the methods employed.

4.2.1 The lexical hypothesis.

The lexical hypothesis is a research approach commonly associated with the study of individual differences in personality, most notably contributing to the empirical basis of the so-called Big Five personality factors (Goldberg, 1992; John & Srivastava, 1999; Poropat & Corr, 2015). Although professionalism is a construct

broader than personality alone, the lexical hypothesis has relevant implications for its study. The foundations of the lexical hypothesis were described by Galton (1884; Poropat & Corr, 2015), further developed by Allport and Odbert (1936; Poropat & Corr, 2015). The lexical hypothesis assumes that there are differences between individuals which cannot be explained by causal inferences and are observed and articulated within everyday life (Mollaret, 2009). It is hypothesised that languages develop in such a way to evolve descriptors of the most important or recognisable individual differences, with the most salient taking the form of a single word, such as ‘extraversion’ or ‘introversion’ (John & Srivastava, 1999; Poropat & Corr, 2015). The lexical hypothesis is that there will be single word descriptors within all languages that capture key differences between individuals, and that these descriptors may be interpreted as corresponding to personality traits (Ashton & Lee, 2005; John, Angleitner & Ostendorf, 1988; Livianiere & De Raad, 2016; Mollaret, 2009).

Allport & Odbert (1936; John & Srivastava, 1999; Mollaret, 2009) progressed the lexical hypothesis to form the lexical research approach. They explored the role of meaning in single word descriptors to provide a framework by which researchers could systematically analyse lexicons to identify individual differences in personality. The influence of Allport & Odbert’s (1936; Mollaret, 2009) approach is evident throughout the history of personality and individual differences research, most influentially in the development of the Big Five personality traits and sixteen personality factors described by Cattell (1943, 1945a, 1945b), both of which remain hugely influential in psychology today (Boyle et al., 2016; John & Srivastava, 1999; Mollaret, 2009; Poropat & Corr, 2015).

4.2.1.1 Contribution and critique.

The lexical hypothesis is evident in the groundings of seminal authors whose work underpins much psychological practice and theory today (Boyle et al., 2016; Mollaret, 2009; Poropat & Corr, 2015). The indirect nature of psychological research necessitates that researchers find a way to narrow the descriptive words used in society to those that most accurately encompass meaningful individual differences worthy of further study. The lexical hypothesis enabled this by recognising the

importance of everyday language in encoding salient differences between individuals (Ashton & Lee, 2005). In essence, the lexical hypothesis allowed the individuals who laid the foundations for modern psychology, including Gordon Allport, Raymond Cattell, Francis Galton, and Lewis Goldberg, to begin their paradigm-shifting and disruptive research (Boyle et al., 2016; Mollaret, 2009; Poropat & Corr, 2015).

Despite the influence of the lexical hypothesis and theories of personality based upon research using it, critics have questioned its philosophical underpinnings. The positivist paradigm requires deductive reasoning undertaken by testing hypotheses derived from pre-existing theory. The purpose of this reasoning is that the data generated during the latter phase confirms or challenges the hypothesis and by extension, the theory underpinning it. This potential for theories to be proved false is a key aspect of the positive paradigm; for a concept to be deemed meaningful and useful in positivist science, it must be possible to demonstrate that it is incorrect (Mollaret, 2009; Popper, 1972). Critics argue that the lexical hypothesis is non-falsifiable, as it cannot be demonstrated that individual differences are not encoded in single word descriptors (Mollaret, 2009). As such, the lexical hypothesis is neither meaningful nor useful to science, with the same being true for all evidence based upon it. This criticism only holds true however where one subscribes to a purist positivist worldview (Burke Johnson & Gray, 2010; Burke Johnson et al., 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Neuman, 2000; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003), and so in rejecting the incompatibility thesis (Burke Johnson & Onwuegbuzie, 2004; Creswell, 2010; Teddlie & Tashakkori, 2003, 2010), MMR also rejects this criticism of the lexical approach. Pragmatism acknowledges that complex socio-behavioural phenomena, such as the interpretation and description of individual differences, may not be fully understood through purist research. Instead using methods that allow for socially constructed phenomena, rather than those viewing individual behaviour as absolute 'truths' of reality, may be better suited to answering research questions relating to individual differences.

The lexical approach also receives criticism on methodological grounds. Critics argue that it identifies lay descriptors used to define that which people find important, rather than the meaningful underlying psychological phenomena causing

individual differences (Poropat & Corr, 2015). However, once again this argument is based on a purist positivist approach where making generalisable, objective, causal inferences is of prime importance (Burke Johnson & Onwuegbuzie, 2004; Denzin & Lincoln, 2008; Neuman, 2000). However, while causal explanations are important where research targets them, pragmatism allows for research aiming to describe or articulate socially constructed phenomena, rather than always endeavouring to discover the generalisable laws causing them.

A third criticism of the lexical approach concerns semantic content, arguing that the complex and subjective nature of language means that all words have an element of ambiguity and so cannot be interpreted as conveying objective meaning (Ashton & Lee, 2005; John & Srivastava, 1999; Mollaret, 2009). For example, an individual in a temporary state of anxiety might be described as anxious, but this term may also be used to describe a trait-like stable characteristic of another person. While such double meaning does not preclude the lexical hypothesis, it may question its utility in understanding individual differences. Once again, this argument assumes the positivist requirement to minimise subjectivity (Burke Johnson & Gray, 2010; Burke Johnson et al., 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Neuman, 2000; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003). However, pragmatism allows for the subjectivity of language as providing additional insight through divergent findings (Greene & Caracelli, 2003). Another potential resolution to this debate is provided by the application of prototype theory (John & Srivastava, 1999; Mollaret, 2009; Rosch, 1978). This theory allows for words to be representative of prototypical categories, rather than each corresponding to a single absolute meaning. Prototype theory considers single-word descriptors as potentially belonging to more than one semantic group but acknowledges that these groups will naturally coalesce and be revealed through everyday language, as captured by the lexical approach (Mollaret, 2009).

Finally, critics have cited reproducibility issues as problematic in lexical research. Studies attempting to replicate findings across different cultures and languages have obtained results suggesting poor generalisability (Poropat & Corr, 2012). However, where research aims to provide non-generalisable findings, this issue is of little importance and therefore does not serve to undermine research aiming to describe individual differences in a contextually bound manner.

The major criticisms of the lexical approach are grounded in a purist positivist paradigm. Where a researcher subscribes to the incompatibility thesis, the lexical approach may indeed be contraindicated. However, where research has more pragmatic aims, the research question rather than epistemology guides design. In cases where complex social phenomena are targeted and exploring subjectivity would bring the benefits associated with divergent evidence, using the lexical approach is justified.

4.2.1.2 Implications for this thesis.

As a socially constructed phenomenon, professionalism speaks to individual differences. Although broader than personality alone, professionalism pertains to differences encoded within natural language. Attempts to describe professionalism in objective terms have not proved fruitful in generating consensus (Anderson et al., 2014; Birden et al., 2014; Blake & Gutierrez, 2011; Buck et al., 2015; Evans, 2008; Finn et al., 2010; Goldie, 2013; Marei et al., 2018; Mazor et al., 2007; Monrouxe et al., 2011; O’Flynn et al., 2014; Trathen & Gallagher, 2009; Wilkinson et al., 2009), and so exploring it within its social context may provide new insight. In accordance with the lexical hypothesis, this thesis assumed that meaningful differences pertaining to professionalism are encoded within natural language and sought to explore these differences by considering subjective, experiential aspects of the phenomenon through the ‘lens’ of the participant. The lexical hypothesis was used to inform a pragmatic approach to exploring professionalism as a social, contextually bound construct, thus rendering the positivist objections to the approach irrelevant in this case.

4.2.2 Personal construct theory and associated methods.

The Personal Construct Theory (PCT) of personality was proposed by George Kelly (1955; Brewerton & Millward, 2001; Cantania & Randall, 2015; Easterby-Smith, Thorpe & Holman, 1996; Hill, Wittkowski, Hodgkinson, Bell & Hare, 2016; Walker & Winter, 2007). PCT describes the way that individuals make subjective sense of the objective world and events within it (Butler, 2009a, 2009b;

Catania & Randall, 2015; Cattell, 1965; Kelly, 1955; Walker & Winer, 2007). The individual is described as a scientist systematically exploring, testing, and understanding their environment and their role within it (Hill et al., 2016; Kelly, 1955; Pipere, 2007; Walker & Winter, 2007). Through these endeavours, individuals develop a complex system of internal constructs that describe the world from their viewpoint and guide their understanding of, and behaviour undertaken within, it (Kelly, 1955; Walker & Winter, 2007).

Kelly (1955) postulated that individuals actively make sense of and navigate their environment based on their expectations of the outcomes of events (Butler, 2009a; Cooper, 2010). The scientist-individual forms theories about their environment and the things within it, derives hypotheses based on these theories regarding the outcomes of actions or events, and then tests their hypotheses in the real world, before evaluating their theories further (Butler, 2009a; Cooper, 2010). Kelly (1955) proposed that individuals actively construct their own understanding of the world by seeking themes and commonalities in their environment, construing them in an idiographic manner (Kelly, 1955). This construction results in the development of bipolar constructs through which the dichotomy of the world is understood (Butler, 2009a; Catania & Randall, 2015; Forster, 1992; Kelly, 1955; Pipere, 2007).

The construct is central to PCT as it determines how an individual relates to their environment and people, objects, and events within it, which are collectively known as elements (Brewerton & Millward, 2001; Cooper, 2010; Fransella, Bell & Bannister, 2004). An individual's system of constructs is their mental representation of reality, developed as a result of experiences as to how elements appear similar or dissimilar, an idea that resonates with the lexical hypothesis. For this reason, constructs are bipolar in nature, with the poles of a construct representing what the individual perceives as opposing characteristics. These poles are not based on objective reality, but rather are a product of the individual's internal construction of reality (Butler, 2009a; Easterby-Smith et al., 1996). Poles do not necessarily take the form of objectively logical or lexical opposites, but they form two sides of the same subjective coin for that individual (Cooper, 2010; Kelly, 1955). For example, an individual with a pole of 'sharp' relating to objects, might have the opposite pole of 'smooth'; although logic might suggest 'blunt' as a second pole, constructs represent

the individual's interpretation of the world alone, and so may not always appear overtly logical to others. Individuals exhibit a preference for a particular construct pole; this pole usually becomes apparent first and is known as the emergent pole, followed by the contrast or implicit pole (Butler, 2009a; Cooper, 2010). Where an individual is not fully aware of a construct, or has not had the opportunity to test and refine it thoroughly, the poles may not yet be encompassed in a single word, but elaborated poles will come to be associated with verbal markers or labels taken from commonly used everyday language, as suggested by the lexical hypothesis (Butler, 2009a; Mollaret, 2009; Poropat & Corr, 2015; Procter, 2009). As the use of language is grounded in experience, PCT recognises that different individuals might share labels where their associated constructs are actually different, and vice versa (Butler, 2009a; Procter, 2009).

The construct system reflects an individual's unique perspective on and interpretations of events (Butler, 2009a; Kelly, 1955; Walker & Winter, 2007). However, although idiographic, constructs are formed in a social context of shared experience (Walker & Winter, 2007). As a result, individuals may still share similarity between their construals (Kelly, 1955; Walker & Winter, 2007). Similarity forms a basis for individuals to understand one another's constructs based on their commonality, with higher levels of commonality or shared construing leading to more successful interpersonal interactions (Butler, 2009a; Kelly, 1955; Procter, 2009; Walker & Winter, 2007).

In acknowledging the social context of construing, Kelly provided a framework to understand the full range of human cognition, behaviour, and emotion, including individual differences (Walker & Winter, 2007). PCT is often categorised as a theory of personality, but it actually challenges trait approaches by recognising the impact of situational context on construing (Procter, 2009). PCT understands human behaviour as resulting from the systematic exploration of a world in which different circumstances may result in different construals and different behaviours as a result (Procter, 2009). PCT was therefore appropriately summarised by Kelly (1955) as a 'meta-theory': a theory about the theories that individuals form about their environment and relationships with it (Butler, 2009b).

4.2.2.1 Contribution.

Although not a mainstream psychological theory, evidence suggests that PCT generates a level of consensus that is rare within psychology. PCT features in almost half of articles published within the Annual Review of Psychology journals from 1955 to 2005 and is cited as a milestone whereby psychological theory and methodologies came to be employed within a wide range of research fields, from clinical research to urban planning (Walker & Winter, 2007). The theoretical value of PCT therefore lies not only in its own discrete contributions, but in its stimulating the work of other theorists who have built upon Kelly's work (Walker & Winter, 2007). The methodological legacy of PCT, however, is arguably its most significant contribution to psychology, via the method by which construing is studied. The repertory grid technique (RGT) developed by Kelly specifically to explore personal constructs remains popular in the study of individual differences today (Catania & Randall, 2015; Kačániová & Szabová, 2014; Walker & Winter, 2007). The specifics of the RGT process are discussed in more detail in section 4.2.2.3, but it is closely linked to pragmatism and MMR (Walker & Winter, 2007). Within more general texts, PCT is often categorised as rooted in constructivism (Butler, 2009a; Cooper, 2010). However, those authors considered more central to the elaboration and communication of PCT dispute this (Scheer, 2013; Walker & Winter, 2007), citing links between the work of George Kelly and key thinkers in pragmatism such as John Dewey and William James (McWilliams, 2009; Warren, 2010). The focus of Kelly's PCT was the practical utility of personal constructs to both the individual and others seeking to enhance wellbeing (McWilliams, 2009; Walker & Winter, 2007). It therefore represents a bespoke methodology that is fundamentally grounded in pragmatism, therefore meeting one of the most challenging requirements of MMR and being often cited as one of few truly mixed methods approaches to psychological research (Rocco et al., 2003).

4.2.2.2 Controversy, criticism, and critique.

PCT generates a high level of consensus amongst those conversant with its principles (Walker & Winter, 2007), leading to relatively little discussion as to its weaknesses. However, this may not reflect a lack of grounds on which to base

criticism of the theory, but rather a lack of propensity to undertake such critique. PCT is frequently cited as sitting outside mainstream psychology, described by some as a footnote of the cognitive movement (Scheer, 2013). This lack of mainstream following has resulted in a dearth of contemporary critique that could challenge the aforementioned consensus. However, the limitations of the approach might be inferred by exploring the historical controversies that led to PCT being side-lined.

Kelly's approaches to developing PCT proved contentious at the time. Kelly's PCT challenged the traditional psychoanalytic approach that had already been losing favour, but also rejected the new behaviourism, an approach rapidly gaining momentum and popularity at the time (Butler, 2009a; Cooper, 2010). PCT rejected the behaviourist principle that individuals were passive responders to environmental stimuli, instead asserting that they were active agents constructing, testing, interpreting, and evaluating the external world (Butler, 2009a; Easterby-Smith et al., 1996). Behaviourism was prized for introducing a new level of experimental rigour to psychology and so Kelly's rejection of its fundamental principles, and his move towards less directly observable phenomena, limited the impact of his work amongst peers at the time (Procter, 2009). However, when viewed from a modern perspective, the idea of individuals actively exploring their world is entirely acceptable. A major antecedent of the cognitive psychology movement that followed rapidly on the heels of behaviourism was a desire to recognise the role of cognition as a mediator of stimulus-response processes. Resulting from this later shift, individuals came to be viewed as active thinkers exercising choice in their outward behaviours (Butler, 2009a). In this sense, Kelly's theory is viewed as ahead of its time (Walker & Winter, 2007) in that it is much more at home in modern psychology than in the behaviourist context of the time.

The favour of behaviourism at this time reflected a more fundamental paradigmatic debate. Specifically, the weight of support was shifting away from talking therapies and the qualitative, narrative data on which these were based, and gravitating towards the scientific methods and rigour offered by behaviourism, and later cognitive psychology (Butler, 2009a; Walker & Winter, 2007). Kelly grounded PCT in observations that individuals understood their constructs in narrative form and so cited similarly narrative evidence in support of the theory (Procter, 2009). At a time when positivism was overwhelmingly dominant in the contemporary

paradigm wars, Kelly's use of qualitative data was less convincing to his peers than the quantitative data of behaviourism that was considered to represent the gold standard of research. Once again, however, this criticism of PCT is contextually bound and so when viewed from a more modern perspective, becomes less problematic. Although still raging, the paradigm wars are more balanced today than in Kelly's time, with the value of qualitative research for theory-building more widely recognised. As a result, criticism of PCT based solely on attacking its qualitative foundations appears irrelevant today.

During development, Kelly's PCT also received criticism for being overly focused on social context. Kelly paid more attention to the role of society in understanding personality than was common at the time. He sought to explain construing by acknowledging that it was influenced by and even shared with other individuals within the social environment (Butler, 2009a; Procter, 2009; Walker & Winter, 2007). PCT states that individuals view themselves through bipolar constructs, with the self occupying one pole and others occupying the other. People therefore do not understand themselves in isolation but through comparisons with others in the social environment, and even the potentially endless meta-construing that considers the perception of the self by others using their own self-referent constructs (Butler, 2009a; Walker & Winter, 2007). Once again, the move towards experimental psychology during the latter half of the twentieth century meant that the social emphasis of PCT left it outside of the mainstream (Walker & Winter, 2007). Since this time, psychology has come to reject the idea of understanding individuals as clinically divorced from the social environment, with modern emphasis focusing more on understanding the contributions of and tensions between the individual and society. Consequently, empirical evidence regarding social aspects of construing was more forthcoming during later decades. For example, Neimeyer and Neimeyer (1981; Walker & Winter, 2007) described evidence suggesting that individuals with higher similarity of construing showed greater positive regard for one another, and Adams-Webber (2001) reported that intimate partners were likely to demonstrate similarities in the complexity of their construct systems (Walker & Winter, 2007). Claims that PCT takes account of social context is therefore not a cause for concern in modern psychology, instead being more commonly understood to be a strength today (Butler, 2009b).

Criticisms of PCT are generally viewed as outdated today, growing mainly out of the overwhelming dominance of positivism at the time. Consistent with the more recently developed pragmatist paradigm, PCT is viewed as a useful theory to inform research aiming to answer questions that require understanding of both individual construing and its societal context, whether that research is qualitative, quantitative, or mixed methods.

4.2.2.3 The repertory grid technique.

Although there is significant heterogeneity in application, the Repertory Grid Technique (RGT) remains extremely popular in understanding individual differences in construing (Walker & Winter, 2007). Prior to the RGT, Kelly observed that methods for the study of individual differences involved statistically aggregating groups of individuals. He felt that this approach was contrary to the aims of individual differences research and so set about developing a method that avoided the loss of the individual in statistical averaging. Kelly developed the RGT to meet this end by combining phenomenology and the potential for quantitative analyses (Smith, 1995). Based on PCT, the RGT (Kelly, 1955) explores individual constructs and maps relationships between them (Butler, 2009a; Cantania & Randall, 2015; Pipere, 2007; Rocco et al., 2003; Smith, 1995). The RGT was originally developed as a clinical tool to aid diagnosis and treatment of mental health issues (Cantania & Randall, 2015; Rogers & Ryals, 2007), but has since been applied in a wide variety of settings (see section 4.2.2.3.3).

4.2.2.3.1 Terminology & process.

There are three core phases to the construction of a repertory grid: selecting elements, eliciting constructs, and completing the grid, a fully worked research example of which is described in chapter 6 (Brewerton & Millward, 2001; Cantania & Randall, 2015). The RGT uses elements from the real world (Butler, 2009a; Easterby-Smith et al., 1996) that may be individuals, objects, or events grounded in actual experience (Brewerton & Millward, 2001; Butler, 2009a; Cantania & Randall, 2015; Easterby-Smith et al., 1996). Elements should represent the full range of the

topic being explored, with around eight to ten elements being optimal (Cantania & Randall, 2015; Easterby-Smith, 1980; Easterby-Smith et al., 1996).

The researcher and participant together may select elements where ensuring maximum self-reference is important, but where the research question requires the comparison of multiple individuals' construing regarding a specific topic, it is more helpful to use standardised elements selected beforehand (Easterby-Smith et al., 1996; Jankowicz, 2004). This ensures that all participants use the same elements to maximise the usefulness of inter-individual comparison and ensures that data elicited align with the research question, as required by the pragmatic research paradigm (Jankowicz, 2004).

Critics claim that standardising elements or constructs generates data subject to the influence of demand characteristics (Orne, 1962; Rosnow, 2002; Sharpe & Whelton, 2016; Shaughnessy et al., 2015). In the context of psychological research, demand characteristics are subtle contextual cues provided by the research materials and process, and communications with the research team or other participants, that may lead participants to provide what they expect to be the most helpful responses to questions or tasks (Orne, 1962; Rosnow, 2002; Sharpe & Whelton, 2016; Shaughnessy et al., 2015). For example, where an individual is asked to complete a standardised grid in an organisational context, they may be more likely to supply what they feel to be the 'correct' answer based on the received wisdom of organisational policy rather than personal experience (Brewerton & Millward, 2001). Standardising grids also risks researchers imposing their own construing onto the process, which would limit the flexibility of the technique for exploring unrecognised or unarticulated constructs (Easterby-Smith et al., 1996). The counterargument in favour of standardised elements is that they provide standardised data, meaning that results from different individuals or departments, for example, might be compared in a more meaningful way (Brewerton & Millward, 2001; Cantania & Randall, 2015). One suggested compromise to resolve this issue where the research question requires the standardisation of data is to elicit elements as part of a pre-RGT study using a representative group of individuals, in an attempt to access issues that are likely to be meaningful and relevant to the target participants but without compromising comparability (Brewerton & Millward, 2001; Cantania & Randall, 2015).

Once elements have been selected, construct elicitation begins. The most common approach to eliciting constructs is the *triadic* approach, which is described in chapter 6, section 6.2.1 (Brewerton & Millward, 2001; Butler, 2009a; Catania & Randall, 2015; Cooper, 2010; Easterby-Smith et al., 1996). *Dyadic* approaches using pairs of elements are less common in research as dyads tend to elicit more simplistic construct systems than triads (Butler, 2009a; Catania & Randall, 2015; Easterby-Smith et al., 1996). In the triadic approach, participants are asked to compare and contrast the elements presented along a directed line of enquiry/topic such as professionalism (Brewerton & Millward, 2001; Cooper, 2010; Easterby-Smith et al., 1996). Perceived similarities are recorded in the repertory grid, a completed example of which is available in figure 4.1 (Easterby-Smith et al., 1996). This is the emergent pole forming one end of a bipolar construct (Butler, 2009a; Cooper, 2010). Perceived differences are recorded opposite the emergent pole in the repertory grid; this is the implicit or contrast pole (Butler, 2009a; Cooper, 2010). The number of constructs elicited is highly idiographic, with estimates ranging from seven to thirty constructs being expected (Brewerton & Millward, 2001; Catania & Randall, 2015; Cooper, 2010). The decision to stop eliciting constructs is taken following the lead of the participant, with their inability to generate new constructs being indicative of capturing a comprehensive representation of their construing (Brewerton & Millward, 2001; Cooper, 2010; Easterby-Smith et al. 1996).

Figure 4.1. Example completed repertory grid using object elements and rating four bipolar constructs using a five-point Likert-type rating scale.

	Sandpaper	Hammer	Screwdriver	
<i>Rough</i>	1	4	3	<i>Smooth</i>
<i>Heavy</i>	5	1	2	<i>Light</i>
<i>Cold</i>	4	1	2	<i>Warm</i>
<i>Violent</i>	3	1	2	<i>Kind</i>

Finally, participants are presented with the grid listing their emergent poles in the left-most column, their implicit poles in the right-most column, and the elements

across the top. Participants are asked to provide ratings for each element against each construct in terms of how closely they are perceived to relate to each pole using a Likert-type rating scale (Brewerton & Millward, 2001; Cantania & Randall, 2015; Cooper, 2010; Easterby-Smith et al., 1996). This concludes grid completion, and where grids are used for research purposes, direct participant involvement, with the next stage being the analysis of completed grids.

4.2.2.3.2 Analytical techniques.

Repertory grids were originally intended as therapeutic intervention tools to guide discussion and exploration (Easterby-Smith et al., 1996; Fransella et al., 2004; Walker & Winter, 2007). Analysis would therefore be informal and qualitative, with the main purpose being therapeutic progress rather than deriving information (Cooper, 2010). However, more modern research applications have necessitated methods of analysing repertory grids to evolve. The RGT is now largely considered a mixed methods approach, with both qualitative and quantitative elements being appropriate during analysis, particularly when comparing multiple grids (Brewerton & Millward, 2001; Cooper, 2010; Hill et al., 2016). The RGT has been cited as a method enabling the systematic quantification of psychological phenomena which do not easily lend themselves to quantitative study, such as attitudes and perceptions (Easterby-Smith et al., 1996).

The most commonly used quantitative analytical approaches are factor and cluster analysis, both used to reduce grid data to the underlying dimensions of grids (see chapter 2). In both factor and cluster analyses, patterns of similarity and difference between constructs or elements are of interest (Cooper, 2010). As an example, consider multiple grids completed using the elements of human and animal. These grids are analysed and groupings found that constructs relating to warmth and tails are major dimensions shared amongst the grids of different participants. Individual grids may show that humans and animals are not particularly different when it comes to warmth, suggesting that the participants do not use warmth to differentiate between the two. Conversely, if humans and animals are deemed very different on the dimension of tails, one might interpret this as indicating that the presence of a tail is a key distinguishing feature between an animal and a

human, for this group of individuals. In this example, analysis has used the detailed construing of individuals to identify the broader underlying dimensions that are shared by a group of construing individuals, but with factor and cluster analyses achieving this in different ways.

Factor analysis of RGT data poses some challenges. Traditional factor analysis is undertaken with a single data set gathered using a single set of variables. However, depending on the method used, RGT data can involve multiple datasets (participant-generated elements) or multiple variables (participant-generated constructs; Abdi, Williams & Valentin, 2013). RGT data that involves multiple datasets or multiple variables (i.e. standardised elements or constructs only) may be subjected to Multiple Groups Factor Analysis (MGFA) using specialist software specifically developed to process such data, as described by Abdi and colleagues (2013). For the purposes of clarity, discussions will assume data using standardised elements and participant-generated constructs hereafter within this section.

MGFA is based on the principles of principal components analysis, which explores areas of similarity across different grids interpreted as indicating shared construing (Abdi et al., 2013). The procedure involves first standardising data to ensure no single grid has a disproportionate impact on the final factor solution due to its size or composition. The standardised grids are then concatenated to produce a single grid known as the compromise. The compromise is subjected to principal components analysis in the same way as a single dataset would be during traditional factor analysis. The resulting principal components indicate similarity in the data suggesting underlying variables accounting for variance therein (Abdi et al., 2013). That is to say that the resultant principal components each suggest an area of shared construing amongst the participant sample. At this point, the results are rotated and interpreted as would be expected within traditional factor analysis (see section 4.2.3.5.1 for a full discussion of factor rotation and interpretation).

While factor analysis uses the similarity between ratings within a grid to reveal shared construing, cluster analysis groups constructs according to their semantic similarity (Everitt, Landau & Leese, 2001; Han, Kamber & Pei, 2011; Yim & Ramdeen, 2015). Cluster analysis summarises constructs as a smaller number of semantic dimensions, achieved in one of two ways: agglomerative or divisive. The

former begins with all constructs viewed each as an individual cluster, before merging the two most similar clusters into one (Everitt et al., 2001; Han et al., 2011). This is repeated until all constructs form a single cluster. The latter takes the opposite route, starting with all constructs in a single cluster and splitting it successively until each cluster contains only one construct (Everitt et al., 2001; Han et al., 2011). The researcher must decide at which point to stop agglomeration or division by stipulating the number of clusters to be included in the final solution. This decision is made by studying a dendrogram or tree diagram, which visually represents the agglomeration or division process by displaying clusters proceeding from one to many (Doreian, 2004; Han et al., 2011; Roux, 2018; Yim & Ramdeen, 2015).

The uses of cluster and factor analyses in RGT research are widely accepted, particularly where data from multiple grids/participants are to be summarised as shared construing (Brewerton & Millward, 2001; Cooper, 2010; Hill et al., 201). However, strict adherents to Kelly's approach generally consider quantitative analyses to be inappropriate and unjustified (Brewerton & Millward, 2001). The RGT was designed to provide idiographic data of use for clinical purposes and as such, it is a tool particularly sensitive to individual differences. This means that aggregating data into a single quantitative set may result in distortion of participants' intended meaning (Easterby-Smith et al., 1996). Although this argument is relevant where individual construing is of interest, however, exploring shared construing across multiple grids is best achieved using quantitative analyses, therefore justifying them under the pragmatic paradigm.

4.2.2.3.3 Strengths and applications.

A fundamental strength of the RGT is its *post hoc* consideration as a mixed method. Kelly is now considered to have been heavily influenced by early developments in pragmatism (Scheer, 2013; McWilliams, 2009; Warren, 2010), but the elaboration and articulation of MMR over recent decades has enabled closer links with the RGT. The RGT is philosophically rooted in both constructivism and positivism, evident in its seeking to study subjective construing using objective methods. The RGT therefore combines the strengths of both qualitative and

quantitative approaches while mitigating the weaknesses of each, thus constituting a truly mixed methodology (Rocco et al., 2003).

Another advantage of the RGT is its flexibility. Although there are some consistent features, a wide range of adapted forms of the RGT have been developed to meet different aims and objectives. This flexibility of application means that the RGT has become extremely popular in both research and clinical settings to meet different ends (Easterby-Smith et al., 1996; Rocco et al., 2003). The RGT demonstrates high utility as it is easily administered and, where it is used qualitatively, offers greater efficiency than unstructured interviews due its highly structured nature (Cantania & Randall, 2015). The RGT has been used in the diagnosis and treatment of schizophrenia, post-traumatic stress disorder, anxiety disorders, and in the prevention of suicide (Walker & Winter, 2007). The RGT is also used in educational and organisational settings (Easterby-Smith et al., 1996; Walker & Winter, 2007). On an individual level, the RGT provides a supportive tool to prompt personal development and career planning (Brophy, Fransella & Reed, 2003; Easterby-Smith et al., 1996; Walker & Winter, 2007), and on an organisational level, it enables market research, job role analysis to inform recruitment, performance evaluation, and the exploration of organisational culture and knowledge (Brewerton & Millward, 2001; Brophy et al., 2003; Stephens & Gammack, 1994; Walker & Winter, 2007). In these contexts, the RGT also has the advantage of indirectly increasing employee engagement in projects or change programmes, by engendering a sense that decision-makers are listening to and utilising employee views (Easterby-Smith et al., 1996).

A further strength of the RGT is that it is particularly useful in targeting concepts or questions where the answer may be difficult to articulate. Targeting such questions using other methods risks eliciting received views espoused within the individual's community, network, or organisation, rather than those relating to their actual personal construing. The RGT negates this issue by providing a highly structured and visual approach considered to support efforts to make the ineffable effable, therefore reducing the likelihood that participants will fall back on received knowledge in the absence of readily articulated personal constructs (Easterby-Smith et al., 1996). The RGT is particularly adept at accessing tacit information, which

may not be readily acknowledged, understood, or articulated by individuals (Cantania & Randall, 2015; Grant, 2007).

4.2.2.3.4 Criticism.

In addition to those levelled at PCT discussed in section 4.2.2.2, the RGT itself has also received criticism. One of the most common criticisms of the RGT relates not to the method itself, however, but to its applications, specifically those that are inappropriate or unjustified (Walker & Winter, 2007). Advances in research data collection and analysis software over recent decades have enabled the automated collection of large RGT datasets from which some researchers have sought to derive generalisable findings based on the positivist rationale associated with the statistical power of a large sample size (Easterby-Smith et al., 1996). However, this marks a distinct and problematic departure from the originally intended use and underlying theory of the RGT. The RGT was purpose-designed to offer a conversational prompt during therapeutic intervention and as a result, was not intended to provide a reflection of an objective and generalisable reality, thus preventing any positivist claims of generalisability of findings (Easterby-Smith et al., 1996; Fransella et al., 2004; Walker & Winter, 2007). Described by Walker & Winter (2007) as a “divorce of techniques from theoretical basis” (p. 460), this separation of theory and technique leads to misapplication of the RGT, a fact extrapolated to bring undeserved criticism to the method itself (Easterby-Smith et al., 1996).

Another controversy associated with the RGT concerns researcher misinterpretations of analytical techniques and the information they provide. As noted earlier, repertory grids are considered today to represent a mixed methods approach, as data may be gathered qualitatively but analysed quantitatively (Fransella et al., 2004; Hill et al., 2016; Jankowicz, 2004). Some cite this as effectively offsetting the potential for the response bias or interviewer influence inherent to qualitative interviews (Easterby-Smith et al., 1996; Hill et al., 2016). However, one should consider the purpose of minimising such bias; within the positivist paradigm, controlling extraneous influences serves to provide validity and reliability to maximise generalisability (Burke Johnson & Onwuegbuzie, 2004;

Denzin & Lincoln, 2008; Neuman, 2000). As the RGT is not designed to provide generalisable findings, making such claims regarding the method is inappropriate and unnecessary in light of its sitting within the pragmatic rather than positivist paradigm.

Criticisms of the RGT primarily relate to its execution rather than the theoretical or empirical basis. When applied with integrity and due regard for underlying theory, the RGT may be considered on balance to offer a rigorous MMR tool enabling the systematic quantification of psychological phenomena that are otherwise inaccessible to purely quantitative methods (Easterby-Smith et al., 1996).

4.2.2.4 Implications of PCT and the RGT for the study of professionalism.

Evidence suggests that the RGT as a derivative of PCT is a valid research method particularly suited to applications in MMR made under the pragmatic paradigm. There is a strong precedent for its use in research relating to organisational and occupational issues, and it is particularly well-suited to exploring implicit knowledge. This may be particularly helpful where individuals have not had the opportunity to consider and articulate their own views due to the dominance of received wisdom or occupational dogma. The RGT may therefore be particularly useful in the study of professionalism where the research aim is to explore shared aspects of the construing of professionalism amongst a group of individuals. It will not, however, provide a basis for generalisable findings. As a result, where findings are conducive, they may reasonably be used to build a theory of professionalism and/or act as a prompt for further research within a pragmatic research programme based on the logic of abduction.

4.2.3 Q methodology.

Q methodology is an MMR theory and methodology developed to provide a systematic and robust approach for the study of subjectivity (Baker, van Exel, Mason & Stricklin, 2010; Brown, 1980, 1993; Mason et al., 2018; McKeown & Thomas,

2013; Stenner & Stainton Rogers, 2004; Stephenson, 1935a, 1953b; van Exel & de Graaf, 2005; Watts & Stenner, 2012). Subjectivity refers to aspects of the human condition that are subjectively experienced, such as viewpoints (Brown, 1980; Watts & Stenner, 2012) and personal opinions (Brown, 1993), judgements, perceptions, attitudes, appraisals, or perspectives (Brown, 1996; Gallagher & Porock, 2010; McKeown & Thomas, 2013; van Exel & de Graaf, 2005), including impressions and evaluations of other people (Block, 2008). Subjectivity was the lifetime interest of William Stephenson, who outlined his related Q methodology in a letter published in *Nature* journal in 1935 (Brown, 1993; Stephenson, 1935b; Watts & Stenner, 2012). Throughout the remainder of his career, Stephenson worked to develop and elaborate Q methodology to enhance the study of individual subjectivity (Watts & Stenner, 2012).

Stephenson conceived of subjectivity as a complex and unstable phenomenon, meaning that objective research methods would be unable to gather consistent or reliable data regarding it as a result. Stephenson cited this as the main reason for subjectivity being largely dismissed by the positivist community in favour of phenomena more easily described using numerical data (Stephenson, 1953b). However, Stephenson (1953b) argued that separating the subjective and objective in this way was invalid and that the complexities of studying subjectivity objectively were insufficient grounds to overlook it entirely (Brown, 1972). Stephenson argued that the range of potential subjective viewpoints was of particular interest in studying individual differences and therefore sought to provide the philosophical, psychological, methodological, and statistical tools to explore it systematically and robustly (Stenner, 2009; Stephenson, 1953b). Stephenson spent the majority of his career refining and elaborating Q methodology, which is reflected in its breadth and depth. Consequently, a full discussion of the detail of Q methodology is impossible within this thesis, but the following sections outline the theoretical and empirical issues of greatest relevance to the following chapters.

4.2.3.1 Historical issues and the statistical underpinnings of Q methodology.

Q methodology is most comprehensively understood from a historical perspective. Stephenson was a student of Charles Spearman when the study of individual differences was a major focus and the dominance of quantitative research methods was arguably at one of its highest points in the history of psychology (Watts & Stenner, 2007, 2012). Stephenson termed the dominant research methodologies associated with positivism *R methodology* (Brown, 1980; Stainton Rogers, 1995; Stenner, 2009; Watts & Stenner, 2007). R methodology is characterised by external and objective observers seeking to measure the causes of the internal states of participants (Brown, 1980; Stenner, 2009; Stephenson, 1986a). The complexity of human existence requires that R methods involve breaking the issue of subjectivity down into its composite parts, measuring each separately under tightly controlled conditions (Brown, 1980, Stephenson, 1986b). A defining feature of R methodology is that the anticipated responses of participants are defined by the researcher prior to data collection via a hypothesis to be tested experimentally (Brown, 1980; McKeown & Thomas, 2013; Stephenson, 1986a). The R methodological statistical approach associated with the study of individual differences at this time was Spearman's factor analysis (FA), with Stephenson's internship under Spearman meaning that he was closely involved in its development (Brown, 1980; Watts & Stenner, 2007, 2012)

In the context of Q methodology, Spearman's R methodological FA is known as *by-variable FA* (for further detail, see chapter 2). By-variable FA is used in studies assessing multiple participants across multiple variables. It reduces data gathered to a summary in the form of fewer underlying variables or factors based on how closely associated or correlated the data are, before exploring the relationships between these factors (Block, 2008; Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2012). By-variable FA explores the relationships between variables across a population of individuals, resulting in it becoming instrumental in the study of individual differences during the 1930's (Watts & Stenner, 2012).

However, taking a statistical viewpoint, Stephenson held fundamental concerns about by-variable FA and its reliance upon correlations between variables assessed using indirect methods, such as psychometric tests. Stephenson felt that

factor analysing the results of indirect methods was more likely to reveal the factor structure of the tests used than the variables extant among participants (Brown, 1980; Stainton Rogers, 1995; Stephenson, 1953b). Stephenson felt that the indirect nature of psychological assessment meant that the results of by-variable FA tell us more about the assessments used than individuals completing them. Brown (1980) illustrated this point elegantly with the analogy of timepieces, suggesting that if a clock and watch correlate, this does not mean that they are both accurately keeping time as they may both be constructed in such a way as to be equally incorrect in their measurements. Regarding psychological testing, the common construction underlying measures means that they all impose *a priori* meanings onto the responses of individuals, rather than allowing them to express their own meaning, a factor that Stephenson argued made them all potentially equally incorrect. Stephenson argued that imposing such *a priori* objective meaning onto subjectivity was inappropriate (Brown, 1980).

Stephenson also questioned the process undertaken during by-variable FA, specifically the standardisation of data based on statistical aggregates of the participant population prior to factor extraction (Fischer & Milfont, 2010). Standardisation enables meaningful comparison of data gathered using different scales of measurement. For example, in response to the question of whether a person is heavier than they are tall, one must first transform the data into a consistent format that derives some universal meaning from the measurements of height and weight. To achieve this, both measurements are transformed into a common, or standardised, format based on their normally distributed variance. This means that the researcher can establish where both a person's height and weight sit in relation to the population mean; the person may be tall in comparison to other people, but their weight may be comparatively low. Standardisation involves aggregating large datasets to produce a normal distribution against which individual datapoints may be compared to discover their meaning within a given population. Standardised scores are population-dependent in that they are useful only in comparing individual data to the mean of that population (Fischer & Milfont, 2010).

In standardising data around the means of variables, Stephenson argued that the factors extracted are similarly derived from the mean of a given population and therefore detached from any meaning at the individual level (Block, 2008; Watts &

Stenner, 2007, 2012). Although fit for purpose when comparing individuals with a group norm (Block, 2008), such mean-referenced analyses are inappropriate to research questions targeting individual differences, where individual-referenced factors would be more appropriate and meaningful (Watts & Stenner, 2012). As an alternative, Stephenson proposed that FA be undertaken on a *by-person* basis, exploring the subjective position from which an individual views and interacts with their environment in a way retaining individual differences rather than averaging them out (Block, 2008; Stephenson, 1935b; van Exel & de Graaf, 2005; Watts & Stenner, 2007). By-person FA is considered an inverted version of by-variable FA, because it approaches individuals as variables and summarises their viewpoints as a smaller number of underlying shared viewpoints (Baker et al., 2010; Brown, 1980, 1993; Stephenson, 1935b, 1977; van Exel & de Graaf, 2005; Watts & Stenner, 2012). Similar to the inference of underlying variables in by-variable FA, the underlying viewpoints described in by-person FA reflect the main viewpoints extant amongst participants and how closely each individual viewpoint resembles the more general underlying ones. For example, the political viewpoints of a group of six individuals may be reduced to an underlying factor structure of two: right versus left-wing politics. The former viewpoint may account for more variance with three individuals closely correlated with it, while the second has only two closely correlating individuals, with the remaining individual having views that are significantly correlated with both factors or neither. In this case, the viewpoint of the final participant would not be reflected within the shared viewpoints of the group, but by-person FA would enable the researcher to identify and examine this in a way by-variable FA would not.

Although the theoretical rationale of Stephenson's approach was generally agreed to be sound by his peers, it was met with a scepticism from the outset that remains prevalent today. The main reason for this is a serious and enduring misunderstanding of by-person FA grounded in the historic resilience and dominance of R methodology. In being described as inverted by-variable FA, by-person FA was taken by some to suggest that data elicited using R methodology could simply be flipped in its configuration to list participants as variables, before subjecting it to traditional by-variable FA, known as Burt's transposed matrix model (Brown, 1980; Kline, 1994; McKeown & Thomas, 2013; Watts & Stenner, 2007).

This misunderstanding meant that Q methodology was evaluated using R methodology criteria, resulting in a conclusion that data still required standardisation prior to factoring rendering by-person FA surplus to requirement as this was already achieved by by-variable FA (Block, 2008; Brown, 1980).

Stephenson himself agreed with this conclusion but based on the argument that using R methodology data and analysis were inappropriate to the intentions of Q methodology, regardless of whether data was formatted according to R methodology or the transposed matrix model (Brown, 1980; Watts & Stenner, 2007). Stephenson noted that the transposed matrix model still required data to be standardised prior to factoring, transforming it into mean-referenced data and fundamentally disconnecting it from the theory underpinning Q methodology (Brown, 1980; Watts & Stenner, 2012). The only way to overcome this issue using the transposed matrix model would be to assume the same unit of measurement across both rows and columns within the dataset (Brown, 1980; Watts & Stenner, 2007). As most psychological data is highly unlikely to meet this assumption, Stephenson suggested that an entirely new unit of measurement was required. Stephenson therefore developed a new unit of measurement that could deliver pre-standardised data across different individuals not requiring post-hoc standardisation: the Q sort (Block, 2008; Brown, 1980; Watts & Stenner, 2012). Despite almost universal consensus regarding the errant reasoning of the transposed matrix model, however, it remains a frequently cited approach to Q methodology in European research and instructive texts (Brown, 1980; Burt, 1972; Watts & Stenner, 2007, 2012). The pervasive nature of this misunderstanding of Q methodology has limited its application in psychological research in the years since its development, which is understandable considering that most texts appropriately question the integrity of the transposed matrix model (Brown, 1980; McKeown & Thomas, 2013; Stenner & Stainton Rogers, 2004). It is perhaps disappointing, however, that those texts do not then go on to explain Stephenson's Q methodology accurately, and how the transposed matrix model issues were overcome by the development of the Q sort (for further detail, see section 4.2.3.5.1; Brown, 1980).

Q methodology provides the theory and principles of studying subjectivity, while the Q sort provides a practical method to achieve this. However, it is crucial that the application of the Q sort not be divorced from its encompassing

methodology, which can lead to misapplication and misunderstanding regarding the aims, results, and integrity of the Q approach (Brown, 1980). For example, a key difference between by-variable and by-person FA, and R and Q methodologies respectively, is that while the former uses representative samples of participants to derive generalisable findings, the latter does not strive for generalisability, recognising instead that individual differences by their very nature are not generalisable to other individuals or populations (Watts & Stenner, 2012). By-person FA explores subjectivity shared by participants in a given study that may or may not also be shared by individuals not included within the study but enables no inference as to this either way (Brown, 1980; Watts & Stenner, 2012). Failing to appreciate the non-generalisable nature of Q findings leads researchers to errantly seek representative participant samples in order to ensure generalisability. However, this application of the Q sort is inappropriate and unjustified within Q methodology and so it is important that Q methods are used with appropriate reference to the ‘-ology’ from which they are derived.

4.2.3.2 The theory of operant subjectivity.

Stephenson explicated Q methodology as situated within the theory of operant subjectivity. Despite echoing the dominant R methodology terminology of behaviourism at the time, Stephenson’s use of the word ‘operant’ represented a very different approach to understanding psychological concepts (Watts & Stenner, 2012). Stephenson rejected the false dichotomy of separating internal/subjective and external/objective aspects of reality, instead suggesting that subjectivity was a form of behaviour to be captured and examined objectively in the same way as items extant only in the external world (McKeown & Thomas, 2013; Stenner, 2009; Stephenson, 1953b; Watts & Stenner, 2012). Stephenson theorised that by operating on their environment, individuals might express their subjectivity objectively, hence the term ‘operant subjectivity’ (Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2012).

Operant subjectivity explains how subjectivity might be captured objectively. Stephenson suggested that rather than applying tests to individuals in ways attempting to separate the subjective from the objective, individuals should be

applied to stimuli operantly (Stainton Rogers, 1995). Operant subjectivity states that a person's viewpoint may be captured objectively via its representation in a Q sort through the operation of the individual on the stimuli to be sorted (Brown, 1972; Stephenson, 1953a).

4.2.3.3 Concourse theory.

Q methodology also relies upon concourse theory (Brown, 1993; Stephenson, 1986a), which relates to the stimuli to be arranged within a Q sort. Concourse theory applies to all Q sort stimuli but may be most easily understood as relating to linguistic stimuli. The concourse is the entire domain of possible stimuli relating to a topic or in this case, the entire range of possible statements that might be made about it (Brouwer, 1999; Brown, 1993). If the statements used within a Q sort are considered to be a sample, the concourse would equate to the population or universe from which that sample is drawn (Brown, 1980; Stainton Rogers, 1991; Stephenson, 1986a). For non-linguistic stimuli, the concourse might take the form of all possible objects, sounds, or images associated with a topic (Brown, 1993; Stephenson, 1986a). Stephenson suggested that the concourse may be studied to discover the changing nature of meaning attached to statements, objects, or events, at all levels from the individual to the cultural (Brown, 1993). Such meaning is subjective in nature and so the concourse is considered to enable the expression of subjectivity using objective, external units encompassing that meaning.

4.2.3.4 The modern context: pragmatism and mixed methods research.

Despite having an initially statistical basis, the differences between Q and R methodologies concern the '-ology' rather than the methods. R methodology is rooted in the logic of deduction, proceeding from generalised theories to specific hypotheses tested using specific observations (see section 4.1.1; Brown, 1980). During the latter half of the twentieth century, this approach stood in contrast to the inductive logic associated with qualitative research, whereby specific observations were used to develop generalised theories (Tashakkori & Teddlie, 2003). Stephenson suggested that Q methodology was grounded in the third form of logic associated

with the pragmatic mixed methods approach, namely that of abduction (Brown, 1980; Stephenson, 1986a).

Abduction aims to discover and understand something in detail in its unabridged form, rather than reduce it to objectively verifiable features or facts (Stephenson, 1986a). With the benefit of hindsight, this view may be deemed to have foreshadowed the mixed methods movement, as Stephenson recognised the precedent for Q and R methodologies to work in tandem to answer research questions and enable abduction. Stephenson suggested that Q methodology may be used initially to discover and understand a phenomenon in its intra-individual context, before then applying R methods to explaining that phenomenon in a way that would enable inter-individual generalisations (Stephenson, 1986a, 1986b).

This resembles the mixed methods approach of the pragmatic paradigm, which is perhaps understandable when considering that they both stem from the fundamental rejection of dualist psychology. Pragmatism rejects the incompatibility thesis that the qualitative and quantitative may never co-exist, and Q methodology rejects the dichotomy of the subjective and objective realms of existence. Although not developed within the same context, Q methodology is viewed on a post-hoc basis as “one of few truly mixed methodologies” (p.38, Baker et al., 2010; McKeown & Thomas, 2013; Newman & Ramlo, 2010; Stenner & Stainton Rogers, 2004). It is argued that due to the relative infancy of MMR, new techniques and procedures are needed, but the contribution of Q as a methodology pre-dating MMR is often overlooked (Creswell, 2010; Newman & Ramlo, 2010; Stainton Rogers, 1995). Q methodology was built upon the dual foundations of quantitative statistics and non-generalisable data, thus refuting from the outset any claims that mixed methods fail to achieve such fundamental integration (Newman & Ramlo, 2010). Aspects of pragmatism are evident throughout Q methodology, from the interpretive, abductive framework, through the application of FA to qualitative data and the priority given to the research question and aims in guiding operational decisions, to the practical development of a research method specifically designed to meet its own requirements (Baker et al., 2010; Brown, 1996).

4.2.3.5 The Q sort.

The Q sort is a bespoke research method designed to meet the requirements of by-person FA (Block, 2008; Brown, 1980; Watts & Stenner, 2012). The Q sort asks participants to place a series of stimuli into rank order according to the importance or relevance they hold personally, in relation to a given topic (Brown, 1980; van Exel & de Graaf, 2005). The following sections describe and discuss Q sort procedure with reference to key issues in ensuring integrity of application.

4.2.3.5.1 Q sort terminology and procedure.

The Q sort may be used with single or multiple participants without risk of losing richness or detail in the latter case (Watts & Stenner, 2007, 2012). It requires participants to rank a set of stimuli according to a dimension provided by the researcher, a fully worked research example of which may be found in chapter 7 (Brown, 1980; Brown, 1993). In doing so, participants render their subjective viewpoints objective, in a form conducive to analysis and interpretation by the researcher (Brown, 1980; Baker et al., 2010; McKeown & Thomas, 2013). A single completed Q sort creates a gestaltian expression of that participant's viewpoint (Brown, 1980; Watts & Stenner, 2012), providing a holistic representation of subjectivity to be understood in this gestalt form, rather than being broken down into components for separate analysis (Block, 2008; Brouwer, 1999; McKeown & Thomas, 2013; Stainton Rogers, 1995; Stainton Rogers, 1991; Stephenson, 1986b; van Exel & de Graaf, 2005; Watts & Stenner, 2007, 2012). Scholte, van Lieshout and de Wit (2005; Block, 2008) summarised this by suggesting that R methodology is to Q methodology as bricks are to a building; R methodology correlates the building blocks of a person while Q methodology explores the building as a whole. There are five broad phases to the Q sort, which are discussed below.

Phase 1 - deriving the Q set: The Q set is the stimuli that participants are asked to sort or rank (Brown, 1993; McKeown & Thomas, 2013; van Exel & de Graaf, 2005; Watts & Stenner, 2012). It may take a variety of forms depending on the needs of the researcher and/or participants, most commonly a series of statements about a given topic (Stephenson, 1952; van Exel & de Graaf, 2005; Watts & Stenner,

2012). The Q set may be tailored in size, complexity, or medium to suit the topic under study or the communication or capacity-related needs of participants (McKeown & Thomas, 2013).

The Q set is sampled from the concourse. This refers to all statements that may be made about the topic being studied in any place and at any time (Brown, 1980, 1993; McKeown & Thomas, 2013; Newman & Ramlo, 2010; Stainton rogers, 1991; Stephenson, 1986; van Exel & de Graaf, 2005; Watts & Stenner, 2012). As noted earlier, Q methodology may crudely be considered an inversion of R methodology, where the sample of individuals represents the population of interest. In Q methodology, the sample of statements (Q set) represents the population of statements of interest (the concourse; McKeown & Thomas, 2013). The Q set must therefore be representative of the population from which it is drawn (Brown, 1993; McKeown & Thomas, 2013; van Exel & de Graaf, 2005). This means that deriving the Q set is crucial in Q methodology and, as a result, is often the most time consuming and resource intensive phase when undertaken thoroughly and rigorously (Watts & Stenner, 2012). In contrast to the R methodology population, the concourse is of undetermined and potentially infinite size, meaning that sampling theory cannot be applied to determine the number or nature of statements to be sampled to deliver a representative Q set (Brown, 1980). Q methodologists should therefore employ theoretical sampling techniques (Brown, 1970, 1980; Stephenson, 1953b). In this regard, deriving a Q set is a qualitative and exploratory endeavour, which stands in contrast to the later systematicity of the quantitative aspects of Q methodology (Brown, 1980; Watts & Stenner, 2012).

A representative Q set demonstrates balance and coverage (Stainton Rogers, 1995; Watts & Stenner, 2012). The former speaks to the value-laden nature of statements, rather than their valence. For example, statements about colours may be positively valenced (“the nicest colours are cool toned”) or negatively valenced (“the worst colours are warm toned”) while maintaining the same underlying value (e.g. cool tones are preferred to warm). A balance in value rather than valence is required within a Q set. This researcher should therefore endeavour to balance statements preferring cool toned colours versus warm with those preferring the opposite, although the way these values are phrased may vary. This ensures that the Q set represents the entire population of opinion (Watts & Stenner, 2012), enabling

participants to express their viewpoints fully and without difficulty (Watts & Stenner, 2012).

Developing a Q set requires the gathering of as many representations of the concourse as possible, which may involve large scale and systematic searches and literature reviews covering everything from academic to grey literature, classical texts to social media posts, or even an entirely separate pre-study to explore opinions where other sources are not readily available (Brown, 1993; McKeown & Thomas, 2013; Newman & Ramlo, 2010; Watts & Stenner, 2012). These examples are then narrowed and reworded into a manageable Q set (Watts & Stenner, 2012). Identifying and selecting statements may be more or less structured depending on the aims of the research (van Exel & de Graaf, 2005). An unstructured Q set is based on researcher judgement and formed to reflect the overall essence of the concourse (McKeown & Thomas, 2013; Watts & Stenner, 2012), but can be labour-intensive if balance and coverage are to be maximised and researcher effects minimised (Watts & Stenner, 2012). A structured Q set, such as one derived using thematic analysis (McKeown & Thomas, 2013; Newman & Ramlo, 2010; Stephenson, 1952; van Exel & de Graaf, 2005; Watts & Stenner, 2012), may therefore be preferable for practical reasons where there is no specific rationale for an unstructured approach. In response to claims that structured Q sets are limited in flexibility and ability to fully represent a nuanced and gestalt picture of individual subjectivity, counter-arguments state that a participant is free to impress their own meaning upon statements while undertaking a Q sort regardless of how the Q set is derived, and that the presence of a structured Q set does not actually change the structure of the concourse itself (Brown, 1980, 1993; McKeown & Thomas, 2013; Stephenson, 1953b). It is therefore argued that by structuring the Q set in a way which makes Q methodology more practicable there is little likelihood that anything is lost and a high likelihood that greater balance and coverage will be gained. A structured approach to Q set derivation is therefore recommended (Brown, 1980, 1993).

Q sets vary in size depending on the needs and aims of the research. Broad guidelines suggest that a Q set of between ten and one hundred items is sufficient to ensure breadth and coverage without excessive participant burden (Bolland, 1985; Brown, 1980; Newman & Ramlo, 2010; Stainton Rogers, 1995; van Exel & de Graaf, 2005; Watts & Stenner, 2012).

Phase 2 - undertaking the Q sort: The participant sample in a Q sort study is known as the P set (McKeown & Thomas, 2013, van Exel & de Graaf, 2005) and in line with its pragmatic research aims, is usually identified in the most practical method to answer the research question. Participant numbers vary according to research aims, with the overriding concern being data saturation (van Exel & de Graaf, 2005). Guidelines are therefore tentative in nature, but a P set smaller than the Q set is advised (i.e. fewer participants than statements to be sorted), with around thirty to sixty participants deemed sufficient to reach data saturation (Brouwer, 1999; McKeown & Thomas, 2013; Watts & Stenner, 2007). A rule of thumb for data saturation is that three to five participants should be significantly correlated with each factor to provide confidence that that factor truly exists (for further details, see phase 4a - the analytical approach; Brown, 1980; van Exel & de Graaf, 2005; Watts & Stenner, 2012).

Random sampling of participants constitutes the gold standard in R methodology, but in Q methodology, participants are akin to the variables in an R methodological study. In the same way that selecting random variables in an R study would make little sense, randomly sampling participants in a Q study is also meaningless (Brown, 1980; Watts & Stenner, 2012). Instead, it is recommended that the P set be identified according to their relevance to the research question, using theoretical sampling (Brown, 1980; Glaser & Strauss, 1967; McKeown & Thomas, 2013; Stainton Rogers, 1995; Watts & Stenner, 2012). For example, if a study seeks to understand perceptions of management styles within a given organisation, it would be logical to sample all roles who are managed, manage others, or set the direction or procedures for the management of staff within that organisation.

Once the P set has been identified, each participant is presented with the Q set and asked to sort the statements according to a *condition of instruction* derived from the research question (Brown, 1993; McKeown & Thomas, 2013; Newman & Ramlo, 2010; van Exel & de Graaf, 2005; Watts & Stenner, 2012). Example conditions of instruction include asking participants to sort stimuli according to agreement, importance, or how pleasant they find the statements personally. The condition of instruction is tailored to maximise the meaningfulness of data, either by targeting the research question or by increasing participant comprehension of or engagement with the task (Stephenson, 1952; Watts & Stenner, 2012). Stephenson

suggested that the R methodological concept of statistical significance was of little value in exploring subjective viewpoints, instead suggesting that stimuli be sorted according to the then radical criterion of psychological significance (Brown, 1972; Watts & Stenner, 2012). Simply put, this means that the individual undertakes a Q sort based on what meaning the stimuli carry for them personally, such as how strongly they feel about them or how important they are to their viewpoint.

Most Q sorts ask participants to sort stimuli into a consistent configuration, which most commonly takes the shape of an inverted normal distribution (see figure 4.2; Stainton Rogers, 1995; van Exel & de Graaf, 2005; Watts & Stenner, 2007, 2012). Although Q methodology allows for different ranking configurations, including a single sequential ranking of all items, research suggests that this is unnecessary to achieve its aims and overly burdensome to the participant (Brown, 1980; Stainton Rogers, 1995). As a result, contemporary Q methodologists recommend a quasi-normal distribution to maximise comparability and minimise participant burden (Block, 1956, 2008; Stainton Rogers, 1995; Watts & Stenner, 2012, 2007).

Figure 4.2. *Example inverted quasi-normal distribution Q sort template.*

Most unimportant	Neutral/undecided						Most important
-4	-3	-2	-1	+1	+2	+3	+4

Phase 3 - post-sort interview: Following the Q sort, it is recommended that a post-sort interview is undertaken (Brown, 1980; Brown, 1993; Newman & Ramlo, 2010; van Exel & de Graaf, 2005). Post-sort interviews explore Q sorts further to understand the reasoning and thought processes underlying the placement of statements (Brown, 1980; Brown, 1993; van Exel & de Graaf, 2005). They may also

be used to gather other information required by the researcher, such as relevant demographic information (Watts & Stenner, 2012).

The results of the post-sort interview are later used to inform the interpretation of FA results. It can be used to explore areas of difference between researcher and participant viewpoints, adding an element of reflexivity to the process enabling the examination of potential researcher bias and maximising the clarity of shared understanding between both parties (Gallagher & Porock, 2010). Questioning and clarifying the placement of statements reduces the level of interpretation undertaken by the researcher, allowing participants to speak for their own subjectivity (Brown, 1980). The post-sort interview also provides an opportunity to explore the integrity of the Q sort, providing an indication of the level of participant engagement in the task (Brown, 1980).

Where depth of understanding is the primary research aim, the most effective post-sort interviews are undertaken face to face, as these provide greater richness of data (Gallagher & Porock, 2010; Watts & Stenner, 2012). However, in multiple participant studies where the need for deeper idiographic understanding is less, electronic interviews using open ended, free text entry questions is more common, for reasons of practicality (Watts & Stenner, 2012). It is most important to ensure that the additional information gathered during post-sort interviews is fit for the purpose of informing the interpretation of factors in line with the researcher question.

Phase 4a - the analytical approach: Q sort data from multiple participants is factored to reveal the major viewpoints shared by the P set (Brown, 1993; McKeown & Thomas, 2013; van Exel & de Graaf, 2005; Watts & Stenner, 2012). FA is a data reduction technique that explores the major dimensions or clusters of similarity amongst a data set (McKeown & Thomas, 2013; Newman & Ramlo, 2010). FA takes the individual viewpoints of each Q sort and reveals clusters of similarity amongst them, known as the factor structure. Decisions as to the final factor structure and how many factors to extract, are made by the researcher (Brown, 1980; Brown, 1993; McKeown & Thomas, 2013). Although Q methodology FA is predominantly exploratory, the logic of abduction also acknowledges that researchers are likely to be working from pre-conceived hypotheses or theories as to

the likely factor structure (Watts & Stenner, 2012). The overriding aim in factoring Q data is therefore to ensure balance between the statistically suggested factor structure and theoretical aspects of its interpretation.

FA is undertaken by first correlating all Q sorts with each other to create a matrix describing the relationships between them. This matrix is then factored to reveal clusters of similarity or high correlation (Brown, 1980; Stainton Rogers, 1995; van Exel & de Graaf, 2005; Watts & Stenner, 2012). Various methods are available to undertake FA but the most commonly used in Q methodology are centroid and principal components analysis (PCA). Centroid FA (Brown, 1980; Burt, 1940; McKeown & Thomas, 2013) was the earliest FA method developed and as a result, involves relatively simple statistical calculations (Brown, 1980). Stephenson also advocated the use of centroid FA for theoretical reasons, claiming that it was more conducive to abductive reasoning than other approaches (McKeown & Thomas, 2013; Newman & Ramlo, 2010; Stephenson, 1953b). Unlike PCA, centroid FA has no statistically ‘correct’ solution to the issue of factor structure, so researchers are free to explore the structure that appears most theoretically relevant and meaningful to their research question (Brown, 1980). Since the original development of Q methodology, however, advances in FA have been significant. Centroid FA is more recently viewed as providing merely an approximation of the more precise methods that include PCA and offer guidance as to the ‘correct’ factor structure, meaning that it is no longer offered as standard by some common statistical software (Brown, 1980). In addition, data has shown the results elicited using centroid versus PCA in Q methodology to be very similar (Brown, 1980; McKeown & Thomas, 2013), and so the more modern technique is generally preferred for reasons of accuracy and practicality (McKeown & Thomas, 2013).

Although PCA offers an indication of the best mathematical factor structure to extract from the data (the statistically ‘correct’ solution), the decision as to the structure rests with the researcher. There is much debate as to the criteria on which to base this decision, acknowledging the need to ensure theoretical meaning is not lost in the face of the statistical characteristics of the data. Consequently, seminal writings advocate that judgements based on the researcher knowledge and experience of related theory should be the ultimate guiding principle (Brown, 1980; Watts & Stenner, 2012). Within this broad guidance, however, there are some

objective guidelines that may also contribute. Brown (1980) suggests that factors be extracted where data saturation suggests that they truly exist, denoted by more than two Q sorts exhibiting significant factor loadings to that factor (Watts & Stenner, 2012). A factor loading is a correlation indicating the degree of similarity between an individual Q sort and a factor (McKeown & Thomas, 2013; Watts & Stenner, 2012). The higher a Q sort loads onto a factor, the more similar the individual's view is to the viewpoint summarised by that factor, with a perfect correlation indicating that a Q sort and factor are identical. Low factor loadings suggest that a Q sort shares some similarity with the underlying factor, but not as much as other Q sorts. Although Brown (1980) suggests that any factor with more than two significantly loading Q sorts be extracted, it is more common that four to five significant loaders is used today, to maximise the likelihood that the viewpoint truly exists (Brown, 1980; van Exel & de Graaf, 2005; Watts & Stenner, 2012). Another general guideline to inform factor extraction is Humphrey's rule, as described by Brown (1980). This technical rule is generally used to guide initial factor extraction, before theoretical judgements are made as to the final factor structure (Brown, 1980).

At this point it is worth noting that within R methodology FA, one of the most commonly used extraction criteria is that each factor must account for a minimum amount of variance, ensured by extracting only those factors with an eigenvalue exceeding one (the Kaiser-Guttman criteria; Auerwald & Moshagen, 2019; Yanai & Ichikawa, 2007). This rule is not recommended for use in Q methodology (Block, 2008), because eigenvalues speak only to the statistical characteristics of the data, rendering them virtually meaningless in terms of subjective significance (McKeown & Thomas, 2013). The Kaiser-Guttman criteria serves the R methodology objective of providing generalisable findings and so can lead to factors accounting for small amounts of variance being overlooked (Brown, 1980; McKeown & Thomas, 2013). However, the size of factors has no relevance in Q methodology as the divergent view of a single individual may be of specific interest to the research question (Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2012). For example, where a Q study seeks to explore differences between the views of individuals holding maximum and minimum responsibility roles within an organisation, the singular viewpoint of the chief executive officer is extremely important as the job carrying the highest weight of responsibility. Although it

accounts for minimal variance and would likely therefore be overlooked by the Kaiser-Guttman criteria, this factor is particularly pertinent to the research question under study and should not be overlooked in this case.

Although quantitative criteria such as those listed above may be useful in guiding iterative factor extraction, the ultimate decision as to the final factor structure should be based on a more global assessment of its meaning. While statistical considerations can highlight the structure accounting for maximum variance amongst the P set, this should not impede the more qualitative aims of the research question (McKeown & Thomas, 2013; Watts & Stenner, 2012). As a result, it is recommended that once the factor structure recommended statistically is established, additional factors are extracted as a starting point from which the researcher can work backwards to ensure that no meaningful content is lost by settling immediately upon the structure accounting for maximum variance (Brown, 1980; van Exel & de Graaf, 2005). The assessment of this potentially meaningful content should be considered within a framework of abduction, enabling both the discovery and confirmation of major viewpoints amongst the P set (Block, 2008). Overall, Q methodology factor extraction is recommended to be based on balanced consideration of a range of criteria in the context of both existing theory and the discovery associated with abduction. The criteria described above may inform the factor structure, but it is ultimately the researcher who decides it (McKeown & Thomas, 2013).

Phase 4b - factor rotation and factor arrays: Q sorts are plotted against factors to provide a two-dimensional graphical representation (Watts & Stenner, 2012). Within this graph, each Q sort is plotted against two of the extracted factors, which enables visual inspection of the clusters of similarity. Theoretically, the plot is viewed from the perspective of a third dimension perpendicular to the others (Watts & Stenner, 2012). Rotating factors allows the viewer to change their position relative to the datapoints to view them from different perspectives, to aid interpretation of factors (Brown, 1980). Rotation enables greater understanding of where clusters sit in relation to each other in terms of similarity and can demonstrate what aspects of factors generate particular consensus or divergence. While rotation changes the perspective of the viewer, the data and clusters themselves remain unchanged;

rotation is simply an aid to understanding and interpretation through a change of perspective (Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2012).

As described in chapter 7, section 7.2.4, factor rotation may be undertaken in a number of ways, those most relevant to Q methodology being by hand or judgement versus varimax rotation (Stainton Rogers, 1995; van Exel & de Graaf, 2005). Hand rotation enables maximum flexibility for researchers and maximises opportunities for abduction where there is a specific interest that may not be as relevant in statistical terms (Brown, 1980; McKeown & Thomas, 2013; Stainton Rogers, 1995). Using the previous example of a study exploring differences between high and low responsibility jobs, the researcher could choose to view the data from the perspective of the chief executive officer, to assess the similarities and differences between the viewpoints held by their colleagues within other roles from the viewpoint of their subjective position. Varimax rotation rotates data automatically to find the perspective that accounts for the maximum possible variance in statistical terms (Stainton Rogers, 1995; van Exel & de Graaf, 2005), seeking to maximise the amount of shared subjectivity accounted for by the factor solution, and is therefore well-suited to multiple participant studies with related aims.

Both methods of rotation are deemed to be acceptable as long as their use is justified in the context of the research question (Stainton Rogers, 1995). Varimax has become the favoured option over more recent years due to it providing a clearer representation of factors, thus maximising ease of interpretation, and its utility with larger datasets and research concerned with shared subjectivity (Block, 2008; McKeown & Thomas, 2013; Watts & Stenner, 2012). Hand and varimax rotation may also be combined where maximising understanding of an issue from both individual and shared viewpoints is advantageous, offering a compromise between the historical intentions of Q methodology and more modern developments (Block, 2008; Watts & Stenner, 2012).

Factor rotation may be orthogonal or oblique (Block, 2008; Watts & Stenner, 2012). This refers to whether the two plot axes representing the factors remain static in their distance from one another or are allowed to move. Orthogonal rotation preserves ninety-degree angles between axes but oblique rotation allows the axes to

move to indicate similarity or correlations between factors (Watts & Stenner, 2012). Although it is suggested that no data is orthogonal in nature as all factors are likely to correlate to some degree, orthogonal rotation remains the favoured option within Q methodology (Stainton Rogers, 1995). The aim of Q methodology is to explore and clarify subjective viewpoints, which requires that they be separated as far as possible to enable unobstructed interpretation (Stainton Rogers, 1995; Stainton Rogers, 1991). Orthogonal rotation is therefore theoretically justified to maximise separation between and clarity of extracted factors (Stainton Rogers, 1995). Q research questions usually target viewpoints themselves rather than the way that they relate to other viewpoints, meaning that orthogonal rotation is recommended.

Following rotation, factor loadings are re-calculated and used to identify Q sorts significantly associated with each factor (Brown, 1980; Newman & Ramlo, 2010; Stephenson, 1953b). Rotated factor loadings indicate the degree of similarity between each Q sort and each factor with those most closely associated used to interpret the content of factors (Watts & Stenner, 2012). *Factor arrays* are also created to provide a visual description of each factor to further support interpretation, as described in chapter 7, section 7.2.4. By representing a factor as an array, researchers can describe and interpret it holistically; the viewpoint can be interpreted as a whole with each statement's location being understood relative to the location of all others (Block, 2008; Brown, 1980; McKeown & Thomas, 2013; Stainton Rogers, 1995; Stephenson, 1986b; van Exel & de Graaf, 2005; Watts & Stenner, 2007, 2012). This ensures that factor interpretation meets the Q methodology aim of providing holistic understanding of discrete subjective viewpoints.

Best practice recommendations for Q analysis are flexible but there are some general consistencies. Centroid FA works well with theoretical factor extraction and hand rotation, with this combination being preferable when seeking in-depth understanding of individual viewpoints using an abductively-driven approach. PCA is more commonly used with varimax rotation to maximise the variance accounted for, in line with research aims targeting shared subjectivity. However, each combination may also be used with elements of the other according to the research question, in line with the pragmatic approach.

Phase 5 – interpreting factors results: The aim of interpreting Q factors is to describe the viewpoint encompassed by each factor. Factor descriptions communicate viewpoints as discrete entities using long-form narrative descriptions, to avoid disconnected descriptions of the locations of individual statements (Watts & Stenner, 2012). Factors are described according to the *contextuality principle*, which relates to the relativity of statements to each other rather than in isolation (McKeown & Thomas, 2013).

Factor descriptions in both Q and R methodology rely on researchers to interpret and explain the underlying meaning of each extracted factor (McKeown & Thomas, 2013). Resources supporting this in Q methodology are the factor array for that factor, post-sort interview content from participants who were significantly associated with it, and the distinguishing and consensus statements identified (van Exel & de Graaf, 2005; Watts & Stenner, 2012). Methods for achieving holistic factor interpretation are flexible and pragmatic in that they must meet the needs of the research question, but it is important that researchers, particularly those new to Q, not rely solely on distinguishing statements to inform their interpretations. Reliance upon distinguishing statements can lead to descriptions lacking meaningful coherence (Watts & Stenner, 2012). Consequently, best practice recommends that novices in particular use more structure methods of factor interpretation, such as the framework described by Watts and Stenner (2012) in their seminal Q methodology instructive text.

This framework involves creating crib sheets describing characteristics of factors, followed by detailed consideration of the potential significance of all statements (Watts & Stenner, 2012). The crib sheet identifies statements with statistical and/or abductive significance as well as those that carry some potential meaning worthy of further scrutiny and provides a structured approach to ensure no statement is overlooked or considered in isolation. Once the crib sheet is complete, comments of significantly loading participants are used to confirm, challenge, and elaborate the picture it describes (Watts & Stenner, 2012).

The abductive nature of Q factor interpretation receives criticism for being highly subjective and therefore subject to researcher bias (Brown, 1980; McKeown & Thomas, 2013). However, as previously discussed, pragmatism, MMR, and Q

methodology do not consider subjectivity the enemy of discovery. Stephenson's background in physics influenced his response to claims of researcher subjectivity, which he grounded in quantum theory (Brown, 1993). Stephenson stated that as socially constructed phenomena, viewpoints can only ever exist through the interpretation of another individual, in this case a researcher (Brown, 1993). Stephenson argued that all factor analytical approaches, whether applied under positivism, constructivism, or pragmatism, involve subjective interpretation of results, and that the more pressing concern is where this is ignored or overlooked in favour of false claims of objectivity (Brown, 1980). Factor interpretation within Q methodology is acknowledged as a subjective endeavour, although using more structured approaches and exploring researcher reflexivity can increase the transparency with which it is reported and communicated.

4.2.3.5.2 The roles of triangulation and reflexivity in intersubjectivity.

Triangulation has been claimed as qualitative research's version of validity (Denzin & Lincoln, 2008). It involves gathering data regarding the same phenomenon using different methods to provide assurance that the initial findings are correct (Burke Johnson et al., 2007). Triangulation enhances the rigour of qualitative research and provides additional depth to data increasing the richness of understanding generated (Burke Johnson et al., 2007; Denzin & Lincoln, 2008). However, this definition attracts controversy as it constitutes the application of positivist concepts to constructivist research methods. It is claimed that validity in any form is meaningless in qualitative research, where the objective truth or 'correctness' of findings is not of concern (Burke Johnson & Onwuegbuzie, 2004; Burke Johnson et al., 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Guba, 1990; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003). Alternatively, some argue that in the context of qualitative research, validity does not relate to the objectivity of findings but instead the trust that may be placed in their interpretation (Guba & Lincoln, 2008). Viewed this way, triangulation relates more to the authenticity of interpretations than their objective truth (Guba & Lincoln, 2008).

Reflexivity is proposed as an alternative to triangulation requiring researchers to critically reflect on their own role in, and influence over, execution and interpretation in research (Creswell & Creswell, 2018; Guba & Lincoln, 2008; Teddlie & Tashakkori, 2009). Although related to triangulation, reflexivity is conceptualised as distinct, as rather than using methodological controls to assure validity, reflexivity not only accepts researcher influence, but acknowledges it as a vital and inescapable aspect of the qualitative research process (May & Perry, 2017). Constructivism states that knowledge is created within the context of the unique perspective of the researcher and it is therefore neither possible nor desirable to remove their influence in pursuit of objective truths (Burke Johnson & Onwuegbuzie, 2004; Burke Johnson et al., 2007; Creswell & Plano Clark, 2011; Denzin & Lincoln, 2008; Feilzer, 2010; Guba, 1990; Tashakkori & Teddlie, 2003; Teddlie & Tashakkori, 2003). Instead, reflexivity is called for, during which researchers reflect on factors shaping their interpretations of data and how they might differ from the interpretations of others as a result (Creswell & Creswell, 2018). This reflection does not intend to minimise bias, but rather to explore it in a transparent way for the benefit of readers.

Intersubjectivity pertains to claims of a false distinction between objectivity and subjectivity, suggesting that all phenomena have an aspect of both: their truth in the reality of the external world and their subjective construction in the eye of the beholder (Gauttier, 2017; Morgan, 2007). Intersubjectivity acknowledges that there is an objective external reality that may be explored using triangulation, but also that the information humans receive about that reality can only ever be viewed through the unique lens of their individual subjectivity (Morgan, 2007). Intersubjectivity is recognised in research by seeking shared meaning between the subjectivity of researchers and participants regarding the objective reality that they are both exploring. Intersubjectivity draws on both triangulation and reflexivity; it seeks the clarity to increase trust in interpretations by ensuring they are shared by researchers and participants, but simultaneously acknowledges that this clarity will always remain to some extent at the mercy of individual subjectivity (Morgan, 2007).

Q methodology speaks keenly to the issue of intersubjectivity in that it seeks to capture the subjective in an objective manner (Brown, 1972; Newman & Ramlo, 2010; Stephenson, 1953a; Watts & Stenner, 2012). Its grounding in pragmatism

means that Q research seeks to exploit the understanding enabled by divergent findings, or in the case of Q, viewpoints (Teddle & Tashakkori, 2010). This pursuit of divergence suggests that the convergence sought by triangulation is not entirely congruent with its aims. Triangulation also speaks to the comparison of findings with independent criteria, and as no external criterion exists for individual subjectivity, triangulation would not appear entirely appropriate to Q (Brown, 1980). However, Q methodology does involve triangulation as one seeks to ensure factor interpretations are as authentic as possible. In this regard, triangulation becomes relevant once more. Similarly, reflexivity appears both relevant and irrelevant to Q in different aspects, such as the welcome acceptance of subjectivity, but then the aim of rendering that subjectivity objective respectively. It is in the cyclical balancing of the subjective and objective that intersubjectivity takes a central role in Q methodology. It seeks to maximise subjective meaning of interpretations while also capturing the objective reality of shared subjectivity with as much authenticity as the researcher's own subjectivity will allow (Brown, 1980, 1993; McKeown & Thomas, 2013; Watts & Stenner, 2012). In order to ensure methodological integrity, Q researchers are therefore expected to engage in some form of intersubjectivity exercise designed to meet the aims of the study. This may involve triangulation or reflexivity depending on the planned inferences to be made, in line with the pragmatist epistemology.

4.2.3.6 Strengths and applications.

Q methodology combines the systematicity of quantitative methodologies with a richness of data usually precluded by quantitative methods (Watts & Stenner, 2007, 2012). It also preserves the gestaltian nature of subjectivity, while being accessible to a field dominated by quantitative methods that prizes the compartmentalisation and highly controlled study of psychological concepts in isolation (Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2007, 2012). Q methodology therefore offers advantages where there is a need to balance systematicity of approach, such as where results will be used for important decision-making, with the need to understand complex, socially constructed phenomena in a high level of detail.

Q methodology enables unique understanding of subjectivity with a high level of flexibility (Brown, 1980; Watts & Stenner, 2012). Opportunities to apply any condition of instruction to stimuli taking such a wide variety of possible forms enables applications of Q methodology in, amongst others, the fields of health psychology (Eccleston, Williams & Stainton Rogers, 1997; Jordan, Capdevila & Johnson, 2005; Stainton Rogers, 1991; Stenner, Cooper & Skevington, 2003; Stenner, Dancey & Watts, 2000; Watts & Stenner, 2007) and social psychology (Stenner et al., 2006; Stenner & Marshall, 1999; Stenner & Stainton Rogers, 1998; Watts & Stenner, 2005, 2007), and to support understanding in sociological and political studies (Brown, 1971, 1993).

Q methodology is particularly suited to accessing tacit, self-referent knowledge, of which participants may not have a detailed, considered, and clearly articulated version readily available, or where they find their subjectivity ineffable without a stimulus to support its expression (Baker et al., 2010). The structure of the Q sort supports participants to express such aspects of subjectivity with ease, using a process often reported afterwards as being enjoyable and productive in enabling reflection and insight (van Exel & de Graaf, 2005). This aspect of Q methodology is particularly advantageous for exploring topics that individuals may not have considered in detail before from their own personal viewpoint, such as those received from others, perhaps during socialisation to an organisation or professional culture, for example. It offers opportunities to systematically disentangle subjectivity from received knowledge to provide a holistic picture inaccessible to other research methods.

4.2.3.7 Criticism and controversies.

Q methodology is underused and under recognised within psychology (Watts & Stenner, 2007). One reason for this may be that it arrived at the wrong time. Q methodology is lauded as a truly mixed methodology (Stainton Rogers, 1995) but as a relatively new development, the popularity of MMR and pragmatism arrived much later than Stephenson's original writings (1935). At the time of publication, Stephenson's blend of subjectivity and objectivity, and their discovery through abduction, were not well received, particularly in Europe (Brouwer, 1999; Stainton

Rogers, 1995). Although the issues raised at the time are now outmoded, as indicated by the dawn of pragmatism and MMR (for full discussion, see section 4.2.3.1), these issues remain pervasive and have since evolved into other criticisms built largely upon misapprehension of the methodology (Brown, 1972; Stainton Rogers, 1995), such as those borne out of applying the R methodological concepts of reliability and validity to the Q sort method (Watts & Stenner, 2012).

Applying R methodology criteria to the appraisal of Q research errantly imposes the need for generalisable conclusions onto the methodology (Thomas & Baas, 1992; van Exel & de Graaf, 2005). Stephenson's Q methodology does not intend to provide findings that may be generalised to other individuals, times, or contexts, instead seeking to provide information about the major viewpoints held within a specific P set at a specific time under specific circumstances (Stephenson, 1952; Thomas & Baas, 1992; van Exel & de Graaf, 2005). While R methodology seeks to generalise from person sample to person population, Q seeks to understand a sample of subjectivity as part of the subjectivity population rather than as a more generalised rule (Brown, 1980).

Critics of Q methodology often cite the subjectivity of researchers as a barrier to its aims, stating that the Q sort is designed, administered, and interpreted from this perspective rather than those of participants (van Exel & de Graaf, 2005). For example, when deriving the Q set, researchers are free to sample according to their own preference or judgement, which filters down to the participant through the items they are asked to sort. It is argued that different researchers would select different Q sets and that this would lead to different Q sorts as a result. This would mean that the resulting Q sorts are actually a product of the Q set, rather than a reflection of the participant's subjectivity (Block, 2008). However, according to Q methodology, there is no single correct Q set to be derived from a particular concourse, and as long as the Q set is broadly representative of that concourse, participants will still be free to imprint their own subjective meanings through the sorting process. For example, an individual presented with different sets of colouring pencils will still be able to express their favourite colour by choosing those closest to it to rank most highly. Although studies exploring this are few, research does suggest that undertaking separate Q sorts using different Q sets but derived from the same concourse results in fairly consistent findings (Thomas & Baas, 1992; van Exel & de Graaf, 2005).

Q methodology is also criticised for the role of researcher subjectivity in factor interpretation (Brown, 1980; Brown, 1993; McKeown & Thomas, 2013). However, as discussed in section 4.2.3.5.1, this is the case with all FA, whether undertaken within Q or R methodology. It is pertinent to note that while R methodological studies claim to provide objective and generalisable factor interpretations, this is not the case in Q methodology (Stephenson, 1952). It may therefore be argued that in acknowledging the role of researcher subjectivity in factor interpretation and its role in precluding generalisation, this criticism actually becomes a relative strength of Q methodology over R factor analytical approaches.

In more practical terms, the Q sort method itself is criticised on the grounds that the task requires significant sustained cognitive effort to complete, such that some participants are unable to engage with it fully, especially where it is undertaken at a distance (Bolland, 1985; McKeown & Thomas, 2013; van Exel & de Graaf, 2005). This criticism has been largely dismissed by the Q community, with the exponentially increasing number of Q publications demonstrating the accessibility of the task for a range of participants of all ages (McKeown & Thomas, 2013). Research finds a high level of congruence between Q sorts undertaken face-to-face and using distance methods such as mail or online data collection (Reber, Kaufman & Cropp, 2000; van Exel & de Graaf, 2005) and, contrary to what might be expected of participants who had undertaken a task out of their cognitive comfort zone, many participants report enjoyment and acknowledgement upon completing a Q sort, suggesting that the process is actually helpful and enjoyable rather than overly complex and taxing (van Exel & de Graaf, 2005). Controversy also surrounds arguments as to the best way to execute the Q sort. Some question the use of forced-choice distribution templates, arguing that they contravene the fundamental assumptions of Q methodology by limiting its ability to capture participant subjectivity (Block, 2008). This controversy remains unresolved, however, with some claiming that a forced-choice distribution has no effect on results (Block, 2008; McKeown & Thomas, 2013) and others asserting that this is simply untrue (Block, 2008).

Advocates of unconstrained sorting state that forced-choice distributions overlook aspects of participants' viewpoints as they are constrained in expressing it. This argument assumes that there is some meaningful characteristic of participant

viewpoints that prompt the choice of a specific distribution, rather than it being an artefact of individual differences, but empirical evidence refutes this (Block, 2008). Where Q sorters were asked to undertake a series of sorts regarding the personalities of others, first in an unconstrained manner and then using forced distributions, differences in chosen distribution were found to be almost entirely the result of individual differences (Block, 1956, 2008). As such, it appears that there is little qualitative nuance lost when choosing forced-choice sorting distributions over unconstrained.

Q researchers also generally overlook objections to using forced-choice distributions when they are balanced against the relative advantages offered by this approach. Forced-choice distributions require all sorters to make the same level of distinction between statements (Block, 2008). Without this, individuals demonstrating a natural tendency to express themselves in absolute terms who prefer extremes of opinion may use a dichotomous approach to their sorts, such as a simple agree/disagree distinction. Such extremity may be mistaken for strength of viewpoint or even as indicating expertise whereby assuredness or confidence is associated with ‘correctness’ (Block, 2008; McNeil, Sandberg & Binder, 1998). These mistaken interpretations may lead to these sorts being perceived as carrying more weight than those undertaken by more moderate thinkers who place more of their statements in the middle ground. While the latter group may possess the same level of ‘expertise’ as the first, and indeed may largely share their viewpoint, their Q sorts would appear quite different due simply to their individual differences in confidence or cognitive/communicative preference (Block, 2008; McNeil et al., 1998). In addition, differences in sorting configuration impedes the use of FA during analysis, arbitrarily reducing the level of correlation between sorts regardless of whether there is truly a substantive difference in viewpoint (Block, 2008).

A final reason for the favour of forced-choice distributions is that participants have reported that the decreased structure of unconstrained sorting makes it more difficult to manage the cognitive demands of rendering their subjectivity objective. In cases where concepts are complex or not yet fully articulated, the additional structure of forced-choice distributions aid clarity of thought and minimise participant burden and risks of random sorting (Block, 2008). Overall, it appears that

forced-choice distributions may be preferable as a means of maximising integrity in viewpoint comparisons and supporting maximum engagement from the P set.

The criticisms levelled against Q methodology discussed in this section have been effectively responded to by the Q community. Where the methodology is considered appropriately in relation to its aims to provide contextually bound understanding, R methodological critique is rendered meaningless. Where shared subjectivity is targeted, meaning that comparing the Q sorts of multiple participants is important, forced-choice distributions are deemed to have little impact on findings and actually serve to improve participant engagement and the derivation of meaningful conclusions. All things considered, Q researchers tend to reach consensus that the criticisms of the methodology and its method are largely unfounded, even where they disagree on the particulars of its philosophy or execution (Block, 2008).

4.2.3.8 Implications of Q methodology for the study of professionalism.

Q methodology is a well-established and detailed approach to the systematic study of subjectivity. Although not universal, there is significant consensus that it sits comfortably within the pragmatic research paradigm with the Q sort constituting a dedicated MMR tool. Q methodology makes a strong case that objective research methods are ill-equipped to handle the complexities of subjectivity meaning that the Q sort is unique in its ability to gather rich, detailed, and meaningful data pertaining to individual perspectives. Q methodology is particularly helpful in cases where the range of viewpoints is sufficiently divergent to impede consensus, such as is the case with professionalism. Q methodology offers a non-reductionist, flexible approach that is particularly advantageous in studying complex, socially constructed phenomena where viewpoints may be more tacit in nature. As a result, Q methodology is particularly well suited to understanding how professionalism is viewed by different individuals and where viewpoints are shared, providing an effective medium for rendering such nebulous subjectivity into a more objectively comprehensible form.

4.2.4 Miscellaneous methods.

A number of other methods used in this thesis were chosen based on the principles of pragmatism that underlie MMR, although they are purely qualitative in nature. These methods were chosen to answer specific research questions, as recommended by the pragmatic paradigm, but were not integrated with quantitative elements for reasons specific to the aims of the research study within which they were employed (for further details, please see chapters 5 and 8). The methods described in the following sections should be considered as contributing to a programme-level mixture of methods within this thesis.

4.2.4.1 Systematic and methodological reviews.

Systematic reviews involve appraising the research relating to a given topic to enable overall summary statements and conclusions to be drawn based on a comprehensive assessment of the available literature (Gough, Oliver & Thomas, 2012; Petticrew & Roberts, 2006). The key difference between systematic reviews and narrative literature reviews, is that the former is conducted according to scientific principles that aim to minimise bias (Centre for Reviews and Dissemination, 2009; Petticrew & Roberts, 2006). The systematicity of a review lies in the process used to search for and identify relevant literature and the methods used to synthesise the evidence contributed by individual sources into an overriding set of conclusions (Petticrew & Roberts, 2006). Systematic reviews are undertaken according to a detailed protocol developed prior to the commencement of searching or data collection, to ensure rigour and enable researchers to be held accountable for their practice (Gough et al., 2012; Petticrew & Roberts, 2006). In this sense, systematic reviews are more similar to research studies than the discussion of literature expected in narrative literature reviews; just as research methods are transparently stated in sufficient detail to enable replication before data collection commences, the same is true of systematic reviews (Bearman et al., 2012; Petticrew & Roberts, 2006).

Narrative literature review is an umbrella term for all items that provide an overview of more than one research study (Dickson, Cherry & Boland, 2017). They

are often considered to convey expert opinions using a sample of the literature sourced to the author's ends (Dickson et al., 2017; Torgerson, 2003). As such, the literature included may be skewed by confirmatory bias, whereby researchers select only those sources that confirm their *a priori* viewpoint (Torgerson, 2003), and usually no claims are made to offering a comprehensive or unbiased summary of the complete literature (Dickson et al., 2017). Conversely, systematic reviews stipulate the criteria for inclusion rather than the conclusions to be drawn on an *a priori* basis, therefore increasing the rigour associated with the research findings (Torgerson, 2003). Additional rigour is provided by the requirement to publish the protocol upon which a systematic review is based, to enable other authors to appraise and critique the research openly (Dickson et al., 2017; Torgerson, 2003).

Systematic reviews involve creating a protocol to dictate the search and analytical strategy prior to data collection, systematically defining searching of multiple named literature sources, screening results to remove duplicates and apply pre-defined inclusion and exclusion criteria, full extraction of data identified as meeting the study aims, and analysis of this data resulting in a synthesised statement of findings (Centre for Reviews and Dissemination, 2009; Dickson et al., 2017; Gough et al., 2012; Torgerson, 2003). A systematic search strategy balances sensitivity and precision/specificity (Gough et al., 2012; Torgerson, 2003). Sensitivity relates to the strategy's ability to detect all relevant studies without missing any, while precision relates to its ability to detect only those with relevance to the review's aims. Sensitivity and precision are balanced by undertaking precise database searches supplemented by sensitive methods such as hand searching (Centre for Reviews and Dissemination, 2009; Gough et al., 2012).

Although the early stages of this process are fairly standardised, data extraction and synthesis are more variable. Reviews seeking to synthesise quantitative findings may use meta-analyses to statistically summarise findings (Bearman et al., 2012; Torgerson, 2003). However, meta-analytical techniques require certain assumptions to be met, such as including homogenous studies using similar methods and targeting similar constructs, identified using narrowly defined inclusion and exclusion criteria (Centre for Reviews and Dissemination, 2009). Following searching, it may be found that these assumptions are not met by the identified literature, particularly in the social sciences where paradigms,

methodologies, and methods vary (Centre for Reviews and Dissemination, 2009; Petticrew & Roberts, 2006). In such cases, or where the research aim does not target quantitative data, qualitative analytical techniques are used, such as narrative synthesis (Centre for Reviews and Dissemination, 2009; Petticrew & Roberts, 2006). Not to be confused with narrative reviews, narrative synthesis takes place within a systematic review as a means of qualitatively integrating results into a complete summary of the body of literature identified (Centre for Reviews and Dissemination, 2009).

Methodological review using systematic principles is a form of systematic review. Methodological reviews employ systematic search strategies and report narrative synthesis, but specifically target methodological aspects of studies identified. Methodological reviews may be used to report the general quality of a field of study in terms of the methodological rigour, such as that reported by Windle, Bennett and Noyes (2011) appraising studies describing scales for measuring resilience. Although largely based on the same principles as other systematic reviews, methodological reviews are distinguished by the evidence targeted and conclusions drawn. Traditional systematic reviews target the best quality evidence available to support the best quality inferences (Centre for Reviews and Dissemination, 2009; Gough et al., 2012; Petticrew & Roberts, 2006). Best quality evidence is identified by applying principles such as the weight of evidence criteria and Cochrane principles to assess quality and disregard studies likely to undermine robust conclusions (Bearman et al., 2012; Gough, 2007; Gough et al., 2012). This is not the case within methodological reviews, which instead seek to gather a range of literature in totality and use the review process to synthesise findings regarding its collective methodological quality. Methodological reviews might therefore be used to establish the methodological rigour of a field before making decisions as to the further use of its data and/or conclusions.

Systematic review methods are widely used in a range of fields as they deliver advantages over single studies in isolation. They help to identify gaps in extant literature and indicate avenues for future research (Bearman et al., 2012; Petticrew & Roberts, 2006). As the methods are stipulated in advance, systematic review conclusions are also based on the entire relevant literature and not just those serving the author's purpose, meaning that they can be powerful in challenging

errant received wisdom or the assumptions of experts reported as fact (Petticrew & Roberts, 2006). In combining the findings of multiple studies, systematic reviews also mitigate the flaws of individual studies alone (Centre for Reviews and Dissemination, 2009; Dickson et al., 2017; Gough et al., 2012; Petticrew & Roberts, 2006; Torgerson, 2003). A systematic summary of research provides an overall indication of the strength and direction of findings regarding a given topic, in a way that enables research users to make decisions in the face of individual studies that otherwise might appear contradictory (Centre for Reviews and Dissemination, 2009; Petticrew & Roberts, 2006).

Systematic reviews are not without criticism. Proponents of the positivist paradigm criticise them as atheoretical, claiming that without a hypothesis generated from pre-existing theory, results are meaningless (Gough et al., 2012). However, such arguments only stand up to scrutiny if that scrutiny comes from within the authors chosen paradigm (Bearman et al., 2012). When considered from a pragmatic standpoint, this argument appears irrelevant to a mixed methods programme of research. Research paradigms have further implications, however, particularly within the social sciences. Social science research may be grounded in positivism, constructivism, and/or pragmatism, with the use of their associated methodologies and methods ebbing and flowing with a sense of fashionableness over time (Petticrew & Roberts, 2006). One implication of this is little agreement regarding the correct way to undertake systematic reviews (Gough et al., 2012). This issue speaks to analyses rather than search techniques, with the latter attracting more consensus. Regarding analysis, however, there is inconsistency in terminology and methods, and the nature of conclusions deemed appropriate (Gough et al., 2012). A second implication is that such wide-ranging research targeting the same topic is challenging or even impossible to synthesise (Bearman et al., 2012; Petticrew & Roberts, 2006), and in the aforementioned absence of consensus regarding how this might be best achieved, this issue is further compounded. However, when viewed in the context of the fundamental definition of a systematic review, these issues become more concerned with logistics than the value of such studies. If a systematic review requires the undertaking of a pre-planned and rigorous search and analytical synthesis of research that is transparently reported to enable replication, this does not require all such reviews to use the same methods of synthesis. As long as the

methods used are pre-planned, executed with integrity, and reported transparently, a systematic review would meet its fundamental aim effectively and appropriately.

Systematic reviews are also limited by a weakness not associated with the method itself, but rather grounded in the publication process. Publication bias refers to the tendency of publishers to favour certain types of research, resulting in a systematic difference between published and unpublished literature (Gough et al., 2012; Petticrew & Roberts, 2006; Torgerson, 2003). For example, when all else is equal, publishers favour studies with statistically significant quantitative results over those with non-significant results, with a greater proportion of the former reaching English-language publications than the latter (Centre for Reviews and Dissemination, 2009; Gough et al., 2012; Petticrew & Roberts, 2006). The implications of this are that where systematic reviews include only English-language publications and/or peer-reviewed research within the eligibility criteria, systematic bias associated with such publications is introduced to the study. This means that even when undertaken with a high degree of systematicity, reviews ought to be interpreted in light of the research they cannot include due to biases in availability.

4.2.4.1.1 The approach of this thesis.

This thesis employed a methodological review using the systematic principles discussed above. It was conducted according to a detailed, pre-planned protocol explicating the search strategy, eligibility criteria, data extraction approach, and method of synthesis, developed prior to the commencement of searching. A methodological review was chosen as a traditional systematic review would fail to address the research question, which specifically required the appraisal of the methodological quality of identified research studies, rather than an aggregation of their outcomes.

4.2.4.2 Consulting stakeholders.

The study described in chapter 8 reports stakeholder consultation regarding a proposed new theoretical model of professionalism based on existing psychological theory and synthesis of findings from chapters 6 and 7. The study was designed to

meet the aim of theory-building, but in the absence of existing research methods specifically dedicated to this end, a number of approaches were considered. These approaches, and decisions made regarding their implementation, are described in the following sections.

4.2.4.2.1 Member checking.

Member checking is used in qualitative research to enhance the credibility or trustworthiness of findings (Birt, Scott, Cavers, Campbell & Walter, 2016; Candela, 2019). It seeks to reduce the impact of researcher bias by returning to participants with the researcher's interpretations of data to gather their views regarding its accuracy (Birt et al., 2016; Candela, 2019; Harvey, 2015; Iivari, 2018; Madill & Sullivan, 2018; Naidu & Prose, 2018). Member checking ensures research receives 'buy-in' from its contributors and minimises the potential for errant researcher interpretations to overwhelm the true communications of participants (Birt et al., 2016; Harvey, 2015).

Member checking has been cited as providing the equivalent of validation evidence for qualitative research (Birt et al., 2016; Iivari, 2018). It is described as increasing trustworthiness and credibility of findings as an exercise in methodological rigour (Birt et al., 2016; Candela, 2019; Iivari, 2018). It also ensures ethically responsible reporting of research findings by ensuring they accurately reflect the voices of participants (Naidu & Prose, 2018). Member checking can be applied flexibly using a range of methods according to the needs of researchers or participants, making it popular in health and education-related research, where focus groups are one of the most popular methods used for reasons of practicality and accessibility (Birt et al., 2016; Naidu & Prose, 2018). Member checking has also been used to enhance the usability of research outputs where participants are the intended end users of a developing product, programme, or intervention (Iivari, 2018).

Despite widespread use, practice in member checking is variable, with inconsistent reporting constituting a major criticism (Candela, 2019). Many authors neglect to explain how member checking was undertaken and, crucially, what

changes were made as a result (Birt et al., 2016). This raises the potential for unjustified changes to research findings. Critics claim that member checking can lead to research outputs being led solely by the desire for credibility, at the expense of the integrity findings (Madill & Sullivan, 2018). Significant changes to findings following stakeholder consultation could therefore be governed more by the emotional reactions of participants than the ‘truth’ of their content (Madill & Sullivan, 2018). The most effective way to mitigate this issue is through clear and transparent reporting, so that readers may draw their own conclusions regarding the changes made to outputs following consultation (Madill & Sullivan, 2018). Overall, member checking is widely used and accepted in qualitative research, but attention should be applied to reporting its procedures and outcomes in ways allowing readers to evaluate its implications.

4.2.4.2.2 Stakeholder consultation.

Stakeholder consultation involves gathering and using the input of parties with an interest in a given topic, activity, or product as a means of supporting its development and chances for future success upon launch or implementation (Armstrong, Hearnshaw, Powell & Dale, 2007; Brereton et al., 2017). It can support understanding of stakeholder viewpoints and provide evidence for the development or improvement of products, programmes, or interventions (Mendenhall, Iachini & Anderson-Butcher, 2013). Stakeholder consultation aligns well with pragmatism, providing an additional viewpoint from which to view research but also ensuring research is ethically conscious and socially responsible by offering an evaluative voice to stakeholders who might otherwise be at the mercy of researcher interpretations (Barker & Pistrang, 2005). Stakeholder consultation may be a form of member checking where participants are the primary stakeholders but may also use different populations (Madill & Sullivan, 2018). Stakeholder consultation may be undertaken using a wide range of methods and in a variety of contexts, depending on the aim of the consultation and the research question (Brereton et al., 2017). Examples of popular approaches to stakeholder consultation include focus groups and the nominal group technique (Brereton et al., 2017).

Stakeholder consultation has been widely used for marketing purposes (Armstrong et al., 2007) and to enhance patient acceptance of and compliance with health interventions and improve the success of community interventions (e.g. Khan et al., 2015; Mendenhall et al., 2013). In research, stakeholder consultation is particularly important for applied research, enabling those designing or progressing projects or programmes to exploit greater understanding of the stakeholders intended to be served or affected by findings (Madill & Sullivan, 2018).

Stakeholder consultation has the potential to bridge the gap between academic research findings and the reality of lived experience for the participants it describes or stakeholders who will be its primary users (Madill & Sullivan, 2018). It formed a major component of the UK Research Excellence Framework through its links to the impact of research outside of higher education via knowledge transfer (Madill & Sullivan, 2018; Research Excellence Framework, 2011). Where the outputs of academic research are applied, stakeholder consultation and member checking may provide greater confidence that they will reach their full potential (Armstrong et al., 2007). However, stakeholder consultation is subject to the same criticisms discussed earlier regarding member checking. Stakeholder consultation undertaken or reported poorly can raise issues of acquiescence or allow research findings to become altered in ways that reflect the priorities of stakeholder intentions or politics, rather than those suggested by the data itself (Madill & Sullivan, 2018). Once again, transparent reporting addresses this issue, but it may also be tackled through methodological decisions.

4.2.4.2.3 Interacting focus groups.

Focus groups were introduced during the mid-twentieth century to offer a method of group interview that minimised researcher influence on the data collection process (Krueger & Casey, 2015; Morgan, 1997; Stewart, Shamdasani & Rook, 2007). However, they did not gain traction in the social sciences until the 1980's and 1990's, becoming known as *interacting focus groups* (Cyr, 2019; Krueger & Casey, 2015).

Interacting focus groups involve small groups of participants meeting face-to-face to engage in focused discussion, with multiple groups taking place until no new information is generated, known as data saturation (Carlson & Glenton, 2011; Guest, Namey & McKenna, 2016; Hennink, 2014; Krueger & Casey, 2015; Morgan, 1997). Estimates suggest that two to five groups are sufficient to enable data saturation, with empirical evidence suggesting that three to six groups are sufficient to identify 90% of the extant themes and all of the most prevalent ones (Cyr, 2019; Carlson & Glenton, 2011; Guest et al., 2016). Participants in interacting focus groups should be homogenous along one or more dimensions relevant to the interview focus or topic (Hennink, 2014; Krueger & Casey, 2015; Morgan, 1997; Stewart et al., 2007). For example, where the focus is experiences of road traffic accidents in the UK, all participants ought to have experienced a road traffic accident in the UK.

Interacting focus groups provide qualitative data and findings are therefore non-generalisable (Cyr, 2019; Krueger & Casey, 2015; Morgan, 1997; Stewart et al., 2007). As such, there is no advantage to participant randomisation as the research aims are better served by targeting participants purposefully (Krueger & Casey, 2015). Participants typically number between five and eight per group, with the overriding priority being to promote discussion through diversity of opinion but without introducing an unmanageable interpersonal dynamic due to too many voices in the room (Hennink, 2014; Krueger & Casey, 2015; Morgan, 1997). The outputs of interacting focus groups usually take the form of session transcripts that may be compared and contrasted to explore themes or patterns of opinion, but without seeking any specific conclusion, such as the contribution of ideas or the reaching of a decision; interacting focus groups target the processes taking place between participants and the content this elicits, rather than the destination at which they end up (Hennink, 2014; Krueger & Casey, 2015).

Interacting focus groups are helpful where the interaction of the group is likely to provide data that surpasses that expected from individual interviews (Krueger & Casey, 2015; Stewart et al., 2007). This is particularly advantageous where topics are complex or unclear, or relate to multiple dimensions such as behaviour, motivations, and attitudes, as interaction can prompt additional depth or breadth during discussions (Krueger & Casey, 2015; Morgan, 1997). Akin to the

rationale for MMR, interacting focus groups create synergy that elicits data greater than that expected in aggregate from each participant individually (Cyr, 2019; Krueger & Casey, 2015; Morgan, 1997). However, in the efforts to achieve synergy, researchers have minimal control over the direction of discussions during interacting focus groups. Allowing participants to lead discussions and set their direction enables them to feed off one another's contributions in ways that offers synergy, but with the accepted risk that participants may wander off topic and not necessarily provide the specific data sought (Cyr, 2019; Krueger & Casey, 2015; Morgan, 1997). This signifies the importance of matching the method appropriately to the research question; where research questions require specific outcomes or information, interacting focus groups may not be the appropriate approach.

Interacting focus group methods are widely used, particularly in health and social sciences (Carlson & Glenton, 2011; Guest et al., 2016; Hennink, 2014). They offer a highly flexible and effective way of eliciting data regarding subjective experiences and attitudes, stakeholder views, complex behaviours, group processes, and shared experiences and identities (Carlson & Glenton, 2011; Cyr, 2019; Hennink, 2014). Interacting focus groups are particularly effective when targeting socially constructed phenomena where multiple participants can share the burden of unpicking such complex and nuanced concepts (Cyr, 2019). Interacting focus groups provide insight into social contexts, raising the prominence of relevant cultural and social norms (Hennink, 2014). Interacting focus groups can empower participants to set the direction of new and emerging research agendas by identifying self-referentially salient points to be carried forward (Cyr, 2019; Hennink, 2014). Interacting focus groups also bring practical advantages, such as the social context providing a form of moderation whereby the interaction of participants can in itself identify outlying viewpoints or false contributions (Hennink, 2014). This provides less artificial moderation than in individual interviews, where researchers have greater influence on the ways in which contributions are reviewed and categorised (Hennink, 2014).

Interacting focus groups are criticised for overemphasising group dynamics that might interfere with the elicitation of authentic data. Critics claim that participants tend to intellectualise their contributions, focusing on portraying themselves as thoughtful and knowledgeable rather than transparently stating their

thoughts or opinions (Krueger & Casey, 2015). This could lead to participants contributing only ideas deemed to have intellectual merit, risking the omission of more experiential or emotional information. Moreover, evidence suggests that depending on the context, male participants may focus on adhering to gender stereotypes of machismo or dominance during interacting focus groups rather than the topic provided (Morgan, 1997; Stewart et al., 2007). Participants may also sanitise or manufacture false responses based on the social context that the group provides, such as to avoid stating that they do not have an answer in the presence of others (Krueger & Casey, 2015). Similarly, contributions may be inhibited by dominant personalities, perceived status, knowledge, or experience differentials, or by tendencies toward acquiescence, groupthink, or social conformity (Cyr, 2019; Hennink, 2014; Krueger & Casey, 2015; Morgan, 1997; Stewart et al., 2007). Interacting focus groups are also criticised as eliciting superficial data reflecting socially acceptable interactions rather than the depth of data associated with personal disclosures or contributions, or that expected from synergy (Hennink, 2014; Krueger & Casey, 2015; Morgan, 1997; Stewart et al., 2007).

All of these issues suggest that as a result of group dynamics, interacting focus group data may be systematically biased by their social context, potentially limiting the range and depth of information elicited (de Ruyter, 1996). However, these issues are only problematic where research aims to elicit information, because where the interpersonal issues listed above are of interest, their presence is an asset rather than limitation of interacting focus groups. Therefore, where research targets rich understanding of participant views or experiences, the social context of interacting focus groups may impede results, but where understanding social processes and dynamics are important, the method is recommended without issue (Morgan, 1997).

Another criticism of interacting focus groups lies in a lack of guidance against which to assess their rigour. Procedural guidance relating to interacting focus groups tends to focus solely on the practicalities of application, omitting methodological issues such as required sample sizes (Carlson & Glenton, 2011; Guest et al., 2016), and where guidance is provided, the picture remains confused. Some argue that the group itself is the unit of focus, so the number of groups is more important than the number of participants (Carlson & Glenton, 2011), but others

claim the opposite (e.g. Hennink, 2014). These conflicting claims means that many researchers rely solely on the controversial concept of data saturation to justify the number of groups undertaken (Carlson & Glenton, 2011).

Data saturation refers to the point at which no new information is generated by undertaking additional groups (Guest et al., 2016). It is derived from a comprehensive and detailed concept known as theoretical saturation, which is fundamental to grounded theory research (Carlson & Glenton, 2011; Guest et al., 2016). Theoretical saturation is established through an iterative process of comparing incoming data with developing theory until no new insights are identified (Guest et al., 2016). Critics claim that extrapolating the simplified concept of data saturation from the detailed concept of theoretical saturation renders the criterion ambiguous, without clear indication of what exactly is meant by ‘no new information’ (Carlson & Glenton, 2011). This ambiguity creates opportunities for misuse, with claims that poor reporting suggest it is often used as *post hoc* justification for small sample sizes, rather than as the iterative process intended (Carlson & Glenton, 2011). Having said this, it is recognised that not all qualitative research has a developing theory with which to compare incoming data, meaning that the derivation of data saturation for qualitative research more broadly has provided one of the most widely used methodological criteria of the field (Guest et al., 2016).

Focus groups are a widely used and accepted method particularly suited to exploring complex social issues such as professionalism. However, they should be used with caution and be applied only where social processes and group dynamics are of particular interest. As the primary aim of stakeholder consultation is generating specific information about a stimulus rather than exploring the social processes eliciting it, interacting focus groups may not be the appropriate methodological choice.

4.2.4.2.4 The nominal group technique.

The nominal group technique (NGT) is a method for establishing the level of consensus amongst a group of participants (Delbecq, Van de Ven & Gustafson, 1975; Harvey & Holmes, 2012; Humphrey-Murto, Varpio, Gonsalves & Wood,

2017; Hutchings, Rapport, Wright & Doel, 2013; McMillan, King & Tully, 2016; Rankin et al., 2016). It is used to elicit ideas or information regarding a topic and is particularly advantageous where viewpoints regarding that topic may be wide-ranging or the topic itself unclear (Carney, McIntosh & Worth, 1996; Humphrey-Murto et al., 2017; Reimer et al., 2019; Søndergaard, Ertmann, Reventlow & Lykke, 2018). The NGT is commonly used to aggregate expert viewpoints regarding specific topics to identify solutions or prioritise actions (Harvey & Holmes, 2012; Humphrey-Murto et al., 2017; McMillan et al., 2016; Søndergaard et al., 2018), and to explore the views of consumers and other stakeholders regarding a product, service, or intervention (Hutchings et al., 2013; McMillan et al., 2016; Rankin et al., 2016).

Although highly flexible with numerous variations described within the literature (McMillan et al., 2016), the NGT involves four core phases: silent generation of ideas, round robin recording of ideas, a clarification discussion, and anonymous voting or ranking of ideas (Carney et al., 1996; de Ruyter, 1996; Delbecq et al., 1975; Harvey & Holmes, 2012; Hutchings et al., 2013; Lloyd, 2011; McMillan et al., 2016; Rankin et al., 2016; Reimer et al., 2019; Van de Ven & Delbecq, 1971). First, participants are asked to work independently and silently to write down as many ideas regarding the topic as possible (Delbecq et al., 1975; Hutchings et al., 2013; Lloyd, 2011; Van de Ven & Delbecq, 1971), providing time for independent thought and reflection (Delbecq et al., 1975; Lloyd, 2011). Following this, participants are asked in turn to read out one of their ideas in a round robin approach repeated until all ideas have been contributed. Each idea is recorded for the group to see but without further discussion (de Ruyter, 1996; Delbecq et al., 1975; Harvey & Holmes, 2012; Hutchings et al., 2013; Lloyd, 2011; Van de Ven & Delbecq, 1971). Then follows a clarification discussion aiming to ensure participants have shared understanding of each idea according to its contributor's intention, including its meaning and the logic underlying it, but without assuming any consensus towards or agreement with it (Delbecq et al., 1975; Lloyd, 2011). This is achieved by taking each idea in turn and allowing the group to discuss it until clarity is achieved. The group may also combine or remove ideas agreed as duplicates (de Ruyter, 1996; Delbecq et al., 1975; Harvey & Holmes, 2012; Hutchings et al., 2013; McMillan et al., 2016). Finally, participants undertake a private vote in silence

(Delbecq et al., 1975; Hutchings et al., 2013; Lloyd, 2011), which involves choosing the ideas they feel are most important and assigning points to them to indicate a rank order from highest to lowest priority (de Ruyter, 1996; Lloyd, 2011; McMillan et al., 2016; Rankin et al., 2016). These points are then summed by the facilitator to indicate the level of consensus, or lack thereof, regarding the ideas (Lloyd, 2011; McMillan et al., 2016; Rankin et al., 2016; Van de Ven & Delbecq, 1971). The outputs of an NGT focus group are the total list of ideas generated and the aggregated votes/rankings indicating the level of consensus around them (Reimer et al., 2019; Van de Ven & Delbecq, 1971). These outputs provide a map of the group's thought and an indication of the level of consensus therein (Lloyd, 2011). The NGT is typically undertaken face-to-face with groups of between five and fourteen participants, although groups as small as two have proved successful (Delbecq et al., 1975; Humphrey-Murto et al., 2017; McMillan et al., 2016).

The NGT is well-established and has been applied successfully within a range of research, including that relating to professionalism (Hall & Ashcroft, 2011; Ho et al., 2011; Reimer et al., 2019; Søndergaard et al., 2018). It has also been applied successfully in business and marketing (Delbecq et al., 1975; Lloyd, 2011) and education (Harvey & Holmes, 2012), and has a particularly strong track record in health-related industries for exploring new practices or needs prioritisation (Carney et al., 1996; Harvey & Holmes, 2012; Hutchings et al., 2013; Rankin et al., 2016; Søndergaard et al., 2018). One reason for this track record is that in practical terms, the NGT offers a time-efficient approach to gathering ideas and consulting stakeholders, which is a particularly important consideration when seeking input from health professionals (Carney et al., 1996; Delbecq et al., 1975; Harvey & Holmes, 2012; McMillan et al., 2016). The flexibility of the NGT means that it can be applied in a broad range of contexts; it can be adapted according to the availability of resources and target participants, who may be defined by the particulars of the research question (McMillan et al., 2016).

In comparison to interacting focus groups, the NGT can accommodate larger groups due to the level of structure (Delbecq et al., 1975; Van de Ven & Delbecq, 1971). Interacting focus groups concentrate attention on interpersonal dynamics, exploring the nuances of social interaction (Lloyd, 2011). This can only be achieved with small numbers to capture the level of detail required and effectively manage

interpersonal dynamics (McMillan et al., 2016). However, the NGT requires minimal intervention from researchers and focuses on ideas generation rather than who generated them and how in relation to other participants (Carney et al., 1996; Lloyd, 2011; McMillan et al., 2016). As such, the NGT can accommodate larger groups without concerns of managing group dynamics (McMillan et al., 2016; Rankin et al., 2016). The NGT is also therefore suited to nebulous issues such as professionalism, because resources are directed towards generating information about the complexities of the topic rather than people-management (Rankin et al., 2016).

The structure of the NGT means that it is particularly effective in applied research targeting a specific problem or question; it provides practical data with relevance to the specific task at hand (Lloyd, 2011). In contrast to interacting focus groups that allow participants to direct discussions themselves, the NGT remains on topic throughout to ensure all ideas are systematically and equally explored, ensuring no angles are overlooked (Lloyd, 2011; Van de Ven & Delbecq, 1971). This task-focused approach also mitigates potential focus effects, where groups spend a disproportionate amount of time discussing one idea at the expense of all others (Delbecq et al., 1975). In fact, the structure of the NGT procedure also mitigates a number of other issues that must be managed interpersonally within interacting focus groups, issues that could inhibit the generation of information or ideas.

Silent generation ensures that ideas are generated without interaction, competition, or status inhibiting individual contributions and preventing minority or majority influences, or tendencies to conform, from unduly affecting participant contributions (Delbecq et al., 1975; Lloyd, 2011; Van de Ven & Delbecq, 1971). The round robin and voting phases ensure all participants have equal opportunity and responsibility to contribute by allowing for conflicting ideas and removing social desirability concerns associated with observed voting (Delbecq et al., 1975; Harvey & Holmes, 2012; Lloyd, 2011; McMillan et al., 2016; Søndergaard et al., 2018; Van de Ven & Delbecq, 1971). They also prevent unequal weighting of conflicting ideas by preventing premature discussion or dominance by individuals, and negate the potential impact of perceived status or expertise by separating ideas from the individuals contributing them (Carney et al., 1996; Delbecq et al., 1975; Lloyd, 2011; Rankin et al., 2016; Van de Ven & Delbecq, 1971). The structure of the NGT also ensures that the session remains on topic, focused on the question established by

the researcher rather than following the interpersonal flow of discussion (Delbecq et al., 1975; Lloyd, 2011). The clarification discussion ensures all ideas are fully understood with none being overlooked and due to its being based on the written record of the round robin, removes reliance on memory recall (Van de Ven & Delbecq, 1971). Delbecq and colleagues (1975) suggest that participants remember almost twice as much information when it is presented verbally and visually compared to verbally alone. The clarification discussion also prevents groups from reaching premature consensus before all ideas have received full attention (Delbecq et al., 1975). Finally, in procedural terms, Delbecq and colleagues (1975) cite evidence that suggests that mathematically summing votes results in fewer errors when aggregating individual judgements into a reflection of group consensus than the more interpretive, qualitative synthesis used with interacting focus groups.

Perhaps the most compelling rationale for the use of the NGT for ideas generation is that research suggests it is simply more effective: individuals generate more unique ideas when working as a group than the same number of individuals working independently, and groups are more effective at solving problems than individuals working alone (Delbecq et al., 1975; Laughlin, Bonner & Miner, 2006; Van de Ven & Delbecq, 1971). Moreover, nominal groups generate more unique ideas and higher quality ideas than interacting focus groups (de Ruyter, 1996; Delbecq et al., 1975; Langford, Schoenfeld & George, 2002; Reitzschel, Nijstad & Stroebe, 2006; Van de Ven & Delbecq, 1971). The NGT may therefore elicit more ideas or information than either individual or interacting focus group brainstorming (Delbecq et al., 1975; Humphrey-Murto et al., 2017). Nominal groups may work in two ways to deliver this: reducing inhibition of creativity from social interaction, and increasing creativity through social facilitation (Delbecq et al., 1975; Van de Ven & Delbecq, 1971). Social facilitation applies to silent generation, facilitating creativity by providing the constructive tension of observing the productivity of other participants, much like an exam hall (Delbecq et al., 1975; Van de Ven & Delbecq, 1971). Constructive tension also plays a role during the round robin, as individuals may be stimulated to add further ideas to their list based on the contributions of others, an effect known as hitch hiking (de Ruyter, 1996; Delbecq et al., 1975). Regarding social inhibition, Van de Ven and Delbecq (1971) suggest that it operates in various ways within interacting focus groups. Social inhibition may be manifested

as focus effects or premature consensus, participant unwillingness to disclose ideas verbally without first having had time to reflect privately on them or due to perceived status or capability differentials, majority/minority influence and tendencies to conform, the influence of dominant personalities, and the simple drain on resources invested in maintaining interpersonal interaction and momentum. As the NGT minimises these issues, it is recommended as a particularly effective tool for research seeking maximum information or ideas (Van de Ven & Delbecq, 1971).

Having said all this, the NGT also receives criticism, mainly around its execution. Critics claim that instructional texts fail to provide detail regarding measures to be taken to ensure methodological rigour, leading to poor NGT practice (Humphrey-Murto et al., 2017). This issue is further compounded by debate regarding the criteria for assessing the rigour of consensus methods such as the NGT. Some argue that mathematically summing votes constitutes a quantitative method that should demonstrate validity, while others claim that the non-generalisable nature of findings require that it demonstrate qualitative credibility instead. This debate continues but attempts to resolve it by providing guidelines for maximising rigour in the use of the consensus methods do tend to be closer to qualitative than quantitative criteria (e.g. Humphrey-Murto et al., 2017). However, as long as due care and attention is paid to ensuring transparent reporting and methodological integrity when using the NGT, its benefits for research seeking to consult stakeholders, elicit maximum feedback ideas, or prioritise ideas for action appear to outweigh its critique.

Overall, the NGT offers advantages over interacting focus groups when used in relation to certain research questions. Where research seeks to generate ideas or information and minimise the influences of researchers and group dynamics, the NGT is particularly effective. This may be the case where stakeholder consultation seeks to generate information regarding a product, programme, or intervention, rather than seek to understand the social processes that determine that information.

4.2.4.2.5 The approach of this thesis.

The discussions within the previous sections have informed the methodological approach used in chapter 8 of this thesis. As a summary, where the advantages of enhanced credibility and usability are sought, member checking is appropriate where participants contributing the original data are representative of its anticipated end users. However, where a broad participant group was used to gather data for use by a more specific population, it does not apply. In the context of this thesis, data was gathered from a general adult population, but the primary anticipated end users of the outputs were the professional groups to whom it most pertinently applied. As a result, the decision was taken not to undertake member checking with participants, but to apply its principles to stakeholder consultation with principal intended end users for whom the impact of research would be most pronounced.

This consultation was undertaken using focus groups for a number of reasons. Focus groups are an inherently social research method that capitalises on group settings to manage the cognitive requirements of topics as complex as the theoretical underpinnings of professionalism. However, as the research question required maximum information and necessitated that the group remain on-topic, the NGT was used because there was no advantage to understanding the social processes at play within interacting focus groups. The NGT also offered more structure that enabled a lower participant burden in terms of time. The target population of this study were professionals working in demanding environments and/or dispersed locations, so the NGT offered recruitment advantages. A stakeholder consultation was therefore undertaken informed by the principles of member checking using the NGT within focus groups.

4.3 Chapter Summary

This chapter outlined the assumptions underpinning the methodological approach of this thesis and the considerations made when selecting research methodologies. Decisions regarding research methods were made based on careful consideration of their appropriateness to the research questions and where those questions sat in relation to the philosophy of pragmatism. Systematic review

methodology was deemed appropriate to answering the first research question with maximum rigour when compared with a narrative literature review. In terms of mixed methods tools available, the RGT and Q methodology were perceived to offer advantages particularly relevant to the study of professionalism, including their ability to target unarticulated phenomena from a perspective of subjective construals and interpretations within an objective world. These methods also offered true or synergistic integration of qualitative and quantitative methods otherwise difficult to achieve when undertaking MMR. Finally, a stakeholder consultation using the NGT was selected to elicit maximum information to increase the credibility of the theoretical model developed based on the RGT and Q data. The following chapters describe the empirical research undertaken using these methods in the study of professionalism.

Part III: Empirical Research

Chapter 5: The Perennial Problem of Measuring Professionalism: A Methodological Review

Abstract

Professionalism is a requisite for graduate jobs, leading to a proliferation of research exploring its potential measurement to support teaching and learning. Despite growing interest from increasingly wide-ranging stakeholders, the psychometric quality of extant measures has yet to be thoroughly assessed against rigorous and demanding criteria encompassing the full breadth of validity theory. This methodological review aimed to identify published measures of professionalism and assess their rigour in relation to psychometric best practice. A systematic search of seven electronic databases was conducted to identify peer reviewed journal articles published between 2007 and 2017 using quantitative measures targeting professionalism as a complete construct. Psychometric quality was assessed using a combination of criteria reflecting best practice in testing and research. Forty-five measures were reviewed with none found to be methodologically sound. The overriding quality concern was pervasive evidence that validity was seriously and repeatedly undermined. No extant measures were recommended for use where meaningful conclusions were required. The first priority for future research is to establish a testable theoretical construct of professionalism to provide sound foundations for valid measurement.

5.1 Introduction

Holding underperformers to account for their professionalism has become more prominent for employers (Alcolta, Ruiz de Gauna & González, 2016; Blake & Gutierrez, 2011; Carter et al., 2015; Evans, 2008; Evetts, 2014). Professions such as medicine and dentistry have refocused their attention on the issue (Altirkawi, 2014; Arnold, 2002; Baernstein et al., 2009; Buck et al., 2015; Goldie, 2013; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004), with newer regulated professions including pharmacy, teaching, and nursing, following suit (Chisholm, Cobb, Duke, McDuffie & Kennedy, 2006; Evans, 2008; de Mendonça et al., 2016). Occupations previously unconnected to the professionalism debate, such as sales and journalism, have also expressed interest in related training and development activities (Arndt, Evans, Zahedi & Khan, 2019; Black et al., 2019). Attention to professionalism in training programmes is increasing (Blake & Gutierrez, 2011), but there are reports that individual performance still falls short of requirements. Around 20% of employers surveyed in 2017 by The Institute of Engineering and Technology reported non-technical professional skills as lacking in United Kingdom recruits. The professionalism of both student and qualified healthcare workers is also of growing concern (Anderson et al., 2014; Carter et al., 2015; Veloski et al., 2005), with consumers reporting loss of faith in what have been described as unprofessional, uncaring, and unethical staff (Hammer, 2006; Lynch et al., 2004; de Mendonça et al., 2016; Monrouxe & Rees, 2012). This places pressure on training and educational institutions to guarantee a higher quality of professionalism amongst forthcoming graduates (Mid Staffordshire NHS Foundation Trust Public Enquiry, 2013).

5.1.1 The case for effective measurement.

In order to assure professionalism, educators and regulators must be able to measure it effectively to track development, benchmark performance, and assign standards of achievement (Ben-David, Snadden & Hesketh, 2004; Bonke, 2006; Lynch et al., 2004; van Mook et al., 2009). Without adequate measurement, professionalism as it pertains to training and development is rendered irrelevant (Goldie, 2013; van Mook et al., 2009), because training institutions are unable to

make assurances required by labour market stakeholders (Arnold, 2002; Lynch et al., 2004; Veloski et al., 2005).

Attempts to measure professionalism have not proved fruitful, due to a lack of consensus around what their aims and methods should be (Aguilar, Stupans & Scutter, 2011; Arnold, 2002; Bonke, 2006). This means that even training institutions with long-standing interests in professionalism, such as medical and dental schools, are unable to guarantee the fitness of graduates to practice specifically in relation to professionalism (Altirkawi, 2014; Bonke, 2006). Robust psychological measurement enables both formal, norm-referenced assessment of inter-individual professionalism, and informal, developmentally focused intra-individual professionalism. The former enables the setting of acceptable standards and gauging of performance against them, while the latter could guide developmental support and broader practice in teaching and learning.

5.1.2 The case for a methodological review.

The extant literature is characterised by a proliferation of measures targeting professionalism claiming validation using psychometric theory (Arnold, 2002; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; van Mook et al., 2009; Veloski et al., 2005). However, no single measure has gained sufficient traction to be considered a leader in the field and so validation evidence generally comes from single studies, usually where validation is secondary to the major aims of the study (Veloski et al., 2005). This suggests that there may be no extant measures carrying sufficient evidence to satisfy the requirements of robust, multi-source validation (Kane, 1990, 2013; Messick, 1979; Newton & Shaw, 2014). When selecting a measure, practitioners must therefore base decisions on issues of practicality rather than validity, with many choosing to create bespoke measures rather than contribute to the validation of existing ones (Li et al., 2017).

In response, there have been numerous reviews seeking to evaluate the quality of measurement targeting professionalism (Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; de Mendonça et al., 2016; Veloski et al., 2005). These reviews all define professionalism only in the context of medical education or practice.

Moreover, although they speak to the quality of measurement, few include psychometric evaluation criteria. Those that do fail to apply comprehensive quality assessment frameworks grounded in detailed understanding of validity theory (e.g. Jha et al., 2007; Li et al., 2017; Veloski et al., 2005). Examples include Jha and colleagues (2007), who describe the number of studies reporting reliability and validity evidence, but without comment on methodological aspects of this evidence; Veloski and colleagues (2005), who take a more restricted view, including only certain types of validity evidence because their work is grounded in medical expertise rather than psychometry; and Li and colleagues (2017), who use a framework intended for assessing outcome measures rather than one grounded in psychometry, and so assess a limited range of psychometric characteristics.

Despite using limited psychometric quality criteria, these reviews identify shared concerns with measurement practice. Recurring themes include a lack of theoretical or empirical basis for creating measures, including the absence of a clearly defined construct of professionalism (Birden et al., 2014; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004), and inconsistent reporting of validation evidence (Jha et al., 2007; Lynch et al., 2004; Veloski et al., 2005). There is therefore a case to update and extend these findings to determine whether previously identified shortcomings have been addressed and offer robust conclusions by using a more demanding psychometric assessment framework.

Rigorous criteria for quality appraisal will ensure that the full range of psychometric evidence is considered, as recommended by the arguments-based approach to validation (Kane, 1990; Messick, 1979; Newton & Shaw, 2014). Several criteria have been published in research and by professional bodies tasked with evaluating and accrediting psychometric tools. Terwee and colleagues (2007) and Windle and colleagues (2011) developed comprehensive criteria for research purposes, and standards for psychological measurement have also been published by professional bodies, including the European Federation of Psychologists' Associations (Evers et al., 2013) and the Dutch Committee of Testing (Evers et al., 2010). Such criteria provide regulatory recommendations for researchers and practitioners using psychological measures (Evers et al., 2010; Evers et al., 2013). Taken together, these criteria are comprehensive and encompass the divergent range

of intentions psychometric test users have, including research and educational practice aims.

The present review appraises the measurement of professionalism in all occupational sectors, with reference to comprehensive assessment criteria grounded in psychometric theory. A methodological review was undertaken using a systematic search strategy and rigorous quality criteria to evaluate validation evidence pertaining to measure rationale, materials and documentation, interpretation guidelines, reproducibility, validity, and sensitivity of identified measures. The aims of this review were to identify published measures of professionalism, assess their rigour in relation to psychometric best practice, and discuss the implications of findings for educational and research practice.

5.2 Methods

5.2.1 Search strategy.

The search strategy was designed to maximise sensitivity while remaining focussed on the measurement of professionalism as a single, discrete construct. The literature search targeted electronic databases likely to return peer reviewed journal articles relevant to the study of professionalism. The databases selected were PsycINFO, Medline, ERIC, CINAHL, Web of Science, PubMed, and Scopus. The search terms used were (professionalism) AND (measur* OR questionnaire OR scale OR instrument), applied to article titles only and to publications within the 2007-2017 period. This time frame was chosen to provide updated evidence from the period since the publication of Jha and colleagues' (2007) systematic review. The search strategy aimed to update evidence from this review and broaden the criteria to include non-medical professions, in light of growing interest in professionalism from a range of professional, occupational, and educational sectors (Arndt et al., 2019; Black et al., 2019; Carter et al., 2015; Evans, 2008; Evetts, 2014; The Institute of Engineering and Technology, 2017). Initial search results were screened to identify measures used and databases then re-searched to identify, where relevant, publications detailing the original development of measures. The reference lists of all

results were hand searched to identify other relevant articles. Where papers were not freely available online, authors were contacted to request access.

5.2.2 Inclusion and exclusion criteria.

Articles were required to meet several inclusion criteria. They were required to have been published between 2007 and 2017, have undergone peer review during publication, use human participants, and use a quantitative measure of professionalism. Articles were excluded where the full article was inaccessible in English, where no measure of professionalism was used, including comment or review articles, or where professionalism was measured only as a component of a larger construct. Articles measuring only isolated components of professionalism or targeting attitudes towards, perceptions of, or values assumed to relate to aspects of professionalism, were also excluded.

5.2.3 Screening and data extraction procedure.

Identified articles were downloaded into EndNote X7, duplicates removed, and an identification number assigned. Articles were screened and the inclusion and exclusion criteria applied to abstracts. Where a clear screening decision could not be made based on the abstract alone, full papers were downloaded and screened. Papers failing to meet a single exclusion criterion, or all inclusion criteria, were excluded. Full texts identified for further review were downloaded and further exclusions made at full review where abstract screening had not identified inclusion/exclusion criteria violations. A cross-section of 20% of rejected and retained articles were re-screened by another member of the research team and no disagreements were identified.

5.2.4 Materials.

A data extraction form was developed to record details of reviewed papers and scores against a range of best practice psychometric criteria. The form provided qualitative descriptions for each score to be given for each criterion, based on a combination of those used by Evers and colleagues (2010) and Windle and

colleagues (2011). Methodological quality of measures was assessed and scored by the lead researcher, with 20% being sampled by another member of the research team. Discrepancies in scoring were discussed and clearer interpretations of scoring criteria agreed. For each paper, scores for each section were totalled, with overall totals calculated by summing section totals and a mean score per section produced. Overall mean scores per section were also calculated across all papers. As the applicability of some criteria was dependent upon the scores for previous questions, scores were used to support narrative review rather than as quantitative metrics. The scoring framework encompassed six sections: rationale, measure materials and documentation, norm groups utilised, reproducibility, validity, and floor/ceiling effects, as described in table 5.1.

Table 5.1*Quality assessment scoring criteria.*

Question	Score	Quality criteria
Section 1: Rationale		
1.1 Is the purpose of the measure specified?		
a Is the aim of the measure clearly defined?	0	No or insufficient information on which to base decision
	1	Description is unclear as to the aim of the measure
	2	The aim of the measure is briefly described but some aspects remain unclear or specific details are lacking
	3	Clear and detailed description provided of the aim of the measure
b Is the intended construct to be measured clearly defined?	0	No or insufficient information on which to base decision
	1	Target construct is vaguely described, but not operationally defined
	2	Target construct is defined, but only the authors and target users/population were involved in defining
	3	Target construct is clearly operationally defined with input from researchers, target users/population, and field experts
c Is (are) the group(s) the measure is intended for specified?	0	No or insufficient information on which to base decision
	1	Target population undefined or unclear
	2	Target population is defined but without inclusion/exclusion criteria
	3	Target population fully defined with clear inclusion/exclusion criteria to be applied
d Are measure administration details described?	0	No or insufficient information on which to base decision
	1	A description of conditions for administration are not included
	2	Conditions for inclusion are briefly described but without sufficient detail to enable replication
	3	Conditions for administration are fully described in sufficient detail to enable replication
1.2 Is the source of the construct described (i.e. is the construct based on existing theory)?	0	No or insufficient information on which to base decision
	1	No underlying theory described
	2	Theoretical basis is vague or ambiguous
	3	Theoretical basis clearly defined and described
1.3 Does the construction procedure justify the measure content in relation to the construct?	0	No or insufficient information on which to base decision
	1	Investigators only involved in item generation
	2	Investigators and field experts involved in item generation
	3	Investigators, experts, and target population/users involved in item generation
Section 2: Measure materials and documentation		
2.1 Are user and participant instructions AND items clearly defined, and free from ambiguity and unnecessary group-specific (e.g. culturally-bound) wording?	0	No or insufficient information on which to base decision
	1	No
	2	-
	3	Yes
2.2 Is an objective scoring system in place and clearly described?	0	No or insufficient information on which to base decision
	1	No
	2	-
	3	Yes

2.3	Where ratings are completed by an observer, are instructions sufficiently comprehensive and clear to maximise consistency of application?	0	No or insufficient information on which to base decision
		1	Instructions are brief/vague, failing to provide exemplars of rating levels
		2	Instructions are adequate, but fail to provide exemplars of rating levels
		3	Clear instructions provide exemplars of ratings at each level
2.4	Is the scoring/rating system standardised, to minimise scoring/rating errors?	0	No or insufficient information on which to base decision
		1	The scoring system is ambiguous and overly complex (e.g. different scoring conventions for different items, mixed scales of measurement, etc.)
		2	Scoring system is clearly described but overly complex
		3	The scoring system is standardised, simple, and clearly described
<hr/>			
Section 3: Interpretation guidelines/norms			
3.1	Are interpretation guidelines provided?	0	No or insufficient information on which to base decision
		1	No
		2	-
		3	Yes
3.2	Where norm-referenced interpretation guidelines are provided		
a	Quality of supplied group norms		
i	Are norms current?	0	No or insufficient information on which to base decision
		1	Norms were published more than 15 years ago
		2	Norms were published 5-15 years ago
		3	Norms were published fewer than 5 years ago
ii	Are sample sizes sufficient?	0	No or insufficient information on which to base decision
		1	Norm group is fewer than 200 participants AND continuous norming procedures were not used
		2	Norm group is 200-300 participants OR continuous norming procedures were used but without justification of the norm group size used
		3	Norm group is 300+ participants OR continuous norming procedures were used with full justification of the norm group size used
iii	Are the norm samples representative of the target population?	0	No or insufficient information on which to base decision
		1	Norm groups do not match the target population OR fail to cover all relevant groups
		2	Norm groups match the target population but fail to cover all relevant groups
		3	Norm groups match the target population and cover all relevant groups
b	Are relevant descriptive statistics provided regarding norm groups?	0	No or insufficient information on which to base decision
		1	No descriptive statistics for norm samples are included
		2	Descriptive statistics are included for norm samples, but some key details/groups are missing OR detail must be inferred e.g. from graphical representations
		3	Full and detailed descriptive statistics regarding norm samples are provided
c	Is information provided regarding potentially different or unexplored groups?	0	No or insufficient information on which to base decision
		1	Norm samples are mentioned but without sufficient detail to assess unexplored groups
		2	Norm samples are described, but without sufficient detail to identify potentially relevant omitted groups
		3	Norm samples both included and excluded are fully described

3.3	Domain-referenced interpretation		
a	Is there sufficient agreement between raters?	0	No or insufficient information on which to base decision
		1	Kappa is less than 0.6
		2	Kappa is 0.6-0.8
		3	Kappa is 0.8 or above
b	Are procedures for determining cut scores correct?	0	No or insufficient information on which to base decision
		1	Procedure for determining cut scores is arbitrary or not fully detailed
		2	A defined procedure for determining cut scores is described, but procedural adherence is limited
		3	A defined procedure for determining cut scores has been fully described in detail and followed fully
c	Were raters appropriately selected and trained?	0	No or insufficient information on which to base decision
		1	Investigators only were involved in determining cut scores
		2	Investigators and field experts were trained and involved in determining cut scores
		3	Investigators, field experts, and target population/users were trained and involved in determining cut scores
3.4	Criterion-referenced interpretation		
a	Do research findings justify the use of cut scores?	0	No or insufficient information on which to base decision
		1	Research findings used were from studies not intended to provide validity evidence
		2	Research findings used are from studies intended to provide validity evidence, but are of questionable quality
		3	Research findings used are from high quality studies aiming to provide validity evidence
b	Is the research sample appropriate to the intended measure purpose?	0	No or insufficient information on which to base decision
		1	Research samples do not match the target population OR fail to cover all relevant groups
		2	Research samples match the target population but some relevant groups are not included/detailed
		3	Research samples match the target population and cover all relevant groups
c	Is the research sample of sufficient size?	0	No or insufficient information on which to base decision
		1	Research studies using fewer than 200 participants
		2	Research studies using 200-300 participants
		3	Research studies using >300 participants
<hr/>			
Section 4: Reproducibility			
4.1	General		
a	Are procedures reported correct and correctly followed?	0	No or insufficient information on which to base decision
		1	Reliability assessment is lacking OR present but incorrectly/inappropriately applied OR present but fails to meet acceptable levels for the methodology of assessment used
		2	Reliability is assessed using limited techniques e.g. Cronbach's alpha only AND results meet acceptable levels for the methodology of assessment used
		3	Reliability is assessed using a range of approaches e.g. Cronbach's alpha AND factor analysis AND results meet acceptable levels for the methodology of assessment used
b	Are the samples used for reliability assessment appropriate to the purpose of the measure?	0	No or insufficient information on which to base decision
		1	No
		2	-
		3	Yes
c	Is sufficient reliability evidence provided to enable a robust overall assessment of reliability?	0	No or insufficient information on which to base decision
		1	No
		2	-
		3	Yes

4.2	Is internal consistency reliability correctly assessed and reported as meeting acceptable levels?	0	No or insufficient information on which to base decision
		1	Cronbach's alpha not reported OR reported below acceptable levels (<0.70)
		2	Cronbach's alpha only reported at acceptable levels (>0.70)
		3	Factor Analyses undertaken on an appropriate sample size (7*#items and >100) AND Cronbach's alpha reported for each dimension at acceptable levels (>0.70)
4.3	Is test-retest reliability correctly assessed and reported as meeting acceptable levels?	0	No or insufficient information on which to base decision
		1	Correlations not reported OR reported below acceptable levels (<0.70)
		2	Correlations reported at acceptable levels (>0.70)
		3	Correlations reported for each dimension at acceptable levels (>0.70)
4.4	Is parallel/alternate forms reliability correctly assessed and reported as meeting acceptable levels?	0	No or insufficient information on which to base decision
		1	Correlations not reported OR reported below acceptable levels (<0.70)
		2	Correlations reported at acceptable levels (>0.70)
		3	Correlations reported for each dimension at acceptable levels (>0.70)
4.5	Has item response theory (IRT) methodology been applied correctly and appropriately?	0	No or insufficient information on which to base decision
		1	No consideration given to IRT
		2	An appropriate IRT model is utilised and reported AND person fit issues are reported but not utilised/responded to during development
		3	An appropriate IRT model is utilised and reported AND utilised during development AND person fit issues are reported
4.6	Has generalisability theory methodology been applied correctly and appropriately?	0	No or insufficient information on which to base decision
		1	No consideration given to GT
		2	Generalisability coefficients are reported but not responded to during development
		3	Generalisability coefficients are reported and utilised during measure development
<hr/>			
Section 5: Validity			
5.1	Overall validity		
a	Has dimensionality been correctly assessed and appropriately interpreted and reported?	0	No or insufficient information on which to base decision
		1	No
		2	-
		3	Yes
5.2	Construct validity		
a	Has dimensionality been correctly assessed and appropriately interpreted and reported?	0	No or insufficient information on which to base decision
		1	Factor analysis not undertaken OR conducted on an inappropriate sample size (fewer than 7*#items and >100) OR interpretation is based on questionable methodology (e.g. significant post-hoc modifications without theoretical justification)
		2	Factor analysis undertaken on an appropriate sample size (7*#items and >100) AND modifications are theoretically justified but Eigenvalue>1 interpretation only used
		3	Factor analysis undertaken on an appropriate sample size (7*#items and >100) AND modifications are theoretically justified AND interpretation of structure undertaken based on more than Eigenvalue<1 rule e.g. scree plot interpretation
b	Item quality		
i	Is item wording clear, appropriate, and free from unnecessary jargon or group-specific wording?	0	No or insufficient information on which to base decision
		1	Items include vague or unnecessary group-specific wording, multiple meanings, or ambiguity
		2	Items are clear and free from double meanings and ambiguity, but some contain unnecessary group-specific wording or jargon
		3	All items are clear and free from double meanings and ambiguity, unnecessary group-specific wording and jargon

ii	Are item response options clear and consistent, and appropriate in number?	0 1 2 3	No or insufficient information on which to base decision Response options number >7 per item OR use of dichotomous/ipsative response options are not justified/appropriate OR response options are not qualitatively labelled OR response options/labelling is inconsistent Response options number 5-7 per item OR use of dichotomous/ipsative response options are justified/appropriate AND (all response options are not qualitatively labelled e.g. just endpoints are labelled OR response options/labelling is inconsistent) Response options number 5-7 per item OR use of dichotomous/ipsative response options are justified/appropriate AND all response options are qualitatively labelled AND response options/labelling is consistent throughout
c	Has convergent/discriminant validity been correctly assessed and appropriately interpreted and reported?	0 1 2 3	No or insufficient information on which to base decision Associative evidence is lacking OR present but inappropriately focussed (e.g. behavioural criterion for an affective measure) OR standardised associative evidence is present but fails to meet acceptable standards (correlations fail to match Multitrait-Multimethod Matrices (MTMM) predictions OR low (<0.7)/non-significant effect size for Quantifying Construct Validity (QCV)) Associative evidence used considers a range of associations but these are interpreted intuitively by the researchers Associative evidence uses standardised methods such as MTMM or QCV AND results meet acceptable levels (i.e. correlations match predicted pattern in MTMM OR significant high (>0.7) correlations for QCV)
d	General		
i	Are procedures reported correct and correctly followed?	0 1 2 3	No or insufficient information on which to base decision Construct validity assessment is lacking OR present but incorrectly/inappropriately applied Construct validity is assessed using limited techniques e.g. Eigenvalue>1 only, interpretive associative evidence Construct validity is assessed using best practice i.e. range of factor analysis interpretation techniques and standardised associative evidence
ii	Are the samples used for validity assessment appropriate to the purpose of the measure?	0 1 2 3	No or insufficient information on which to base decision Validation samples not included OR used incorrectly/inappropriately OR validity assessment provided is superficial e.g. face validity only Validation is conducted using limited techniques on appropriately sized and representative samples Validation is assessed using best practice techniques on appropriately sized and representative samples
iii	What is the quality of other relevant variables and their measures used in validity assessment?	0 1 2 3	No or insufficient information on which to base decision Construct validity not assessed OR relevant variables are poorly defined/theoretically justified OR no convincing evidence provided that other measures are psychometrically robust OR evidence relating to other measures indicates poor psychometric quality Relevant variables are defined but theoretical basis is questionable OR other measures are psychometrically average/questionable Relevant variables are clearly defined AND theoretically justified AND convincing evidence that other measures are psychometrically robust is reported
5.3	Criterion validity		
a	Quality of criterion research		
i	Are procedures reported correct and correctly followed?	0 1 2 3	No or insufficient information on which to base decision Criterion validity assessment is lacking OR present but incorrectly/inappropriately applied OR correctly applied but results fail to meet acceptable standard (<0.70) Criterion validity is assessed at acceptable levels (>0.70) but evidence utilised is questionable OR not fully justified Criterion validity is assessed at acceptable levels (>0.70 correlation) using evidence for which a strong case it made that it reflects the 'gold standard'

ii	Are the samples used for validity assessment appropriate to the purpose of the measure?	0	No or insufficient information on which to base decision
		1	Validation samples are not included OR used incorrectly/inappropriately
		2	Validation is conducted using appropriately sized and representative samples, but evidence utilised is questionable OR not fully justified
		3	Validation is assessed using acceptable levels (>0.70 correlation) using evidence for which a strong case it made that it reflects the 'gold standard' on appropriately sized and representative samples
iii	What is the quality of other relevant variables and their measures used in validity assessment?	0	No or insufficient information on which to base decision
		1	Criterion validity not assessed OR criterion is poorly defined/theoretically justified OR no convincing evidence provided that other measures are psychometrically robust OR evidence relating to other measures indicates poor psychometric quality
		2	Relevant variables are defined but theoretical basis is questionable OR other measures are psychometrically average/questionable
		3	Relevant variables are clearly defined AND theoretically justified AND convincing evidence that other measures are psychometrically robust is reported
5.4	Content validity		
a	Does the item set comprehensively encompass the full breadth of the construct?	0	No or insufficient information on which to base decision
		1	Measure development was undertaken by researchers alone OR construct was not fully/clearly operationally defined prior to item writing
		2	Measure development was undertaken by researchers AND (target users/population OR field experts) AND construct was fully and clearly operationally defined prior to item writing
		3	Measure development was undertaken by researchers AND target users/population AND field experts AND construct was fully and clearly operationally defined prior to item writing
<hr/>			
Section 6: Floor/ceiling effects			
6.1	To what extent do participant scores utilise the full range of the measure?	0	No or insufficient information on which to base decision
		1	>15% participant responses scored the highest/lowest possible scores
		2	<15% participant responses scored the highest/lowest possible scores but development methodology failed to take steps to minimise floor/ceiling effects
		3	<15% participant responses scored the highest/lowest possible scores AND steps were taken during development to address floor/ceiling effects

Scoring criteria were idiographic but followed the following general format: N/A – the criterion did not apply to this article. This option was used in such cases as where questions related to observer ratings, but papers described self-report measures; 0 (very poor) – no information regarding the criterion was provided on which to base a decision; 1 (poor) – there was mention of the criterion, but with serious shortcomings in its application/execution; 2 (adequate) – the criterion had been executed adequately, but with shortcomings in application or results; and 3 (good) – the criterion was adequately applied and executed, with satisfactory results.

5.2.5 Analysis.

Frequency tables were used to summarise descriptive characteristics of measures. The mean score per section was calculated for each paper to provide a single score out of three per section. Due to the level of heterogeneity found in definitions of professionalism utilised, study designs, and the nature of measures used, statistical integration and assessment were not possible. Consequently, findings were described in a narrative format structured to meet the aims of the study by reflecting the sections used within the scoring framework as a representation of best practice (Evers et al., 2010; Windle et al., 2011).

5.3 Results

A total of 247 papers were initially identified as potentially of interest (see figure 5.1). One hundred and sixty were identified through electronic database searches, 51 as original measure development papers, and 36 by hand searches. Ninety-nine duplicates were excluded prior to screening, where a further 97 papers were excluded for failing to meet the inclusion/exclusion criteria. Fifty-one papers were reviewed in full, with another six excluded at this stage based on information not picked up during screening that contravened the inclusion/exclusion criteria. Forty-five papers were retained for final review, the summary characteristics of which are included in table 5.2.

Figure 5.1. *Summary of review process.*

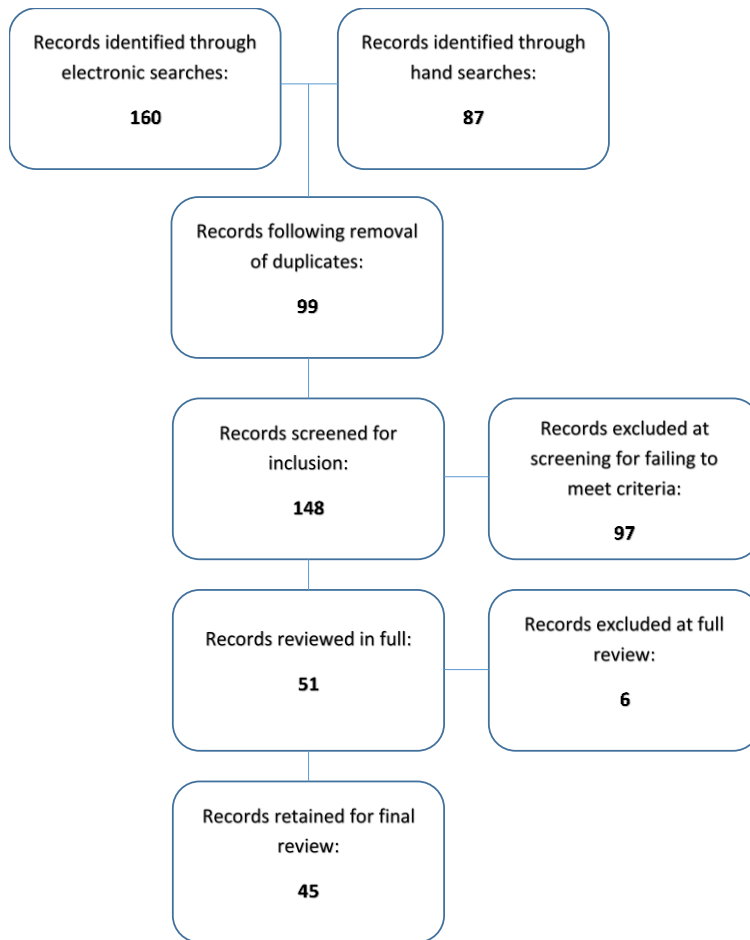


Table 5.2*Description of measures included in final review.*

Identification number	Name of measure	Author(s)	Target population	Mode of completion/ response	Intended administration context	Intended purpose	Notes regarding theory and item selection
1	ABIM's Patient Assessment survey questionnaire (Arabic version)	Abadel & Hattab (2014)	Practicing physicians	Patient completion using a 5-point likert-type rating scale	Clinical medical practice environment	To enable patient and physician peer assessment for maintenance of certification developed and used by the American Board Internal Medicine (ABIM; regulatory body)	Face validity assessed by Faculty of Medicine experts
2	Penn State College of Medicine Professionalism Questionnaire	Akhund, Shaikh & Ali (2014)	Not reported	Self-report using a 5-point likert-type rating scale	Not reported	Not reported	Not reported
3	Modified Physician Achievement Review	Al Ansari, Al Khalifa, Al Azzawi, Al Amer, Al Sharqi, Al-Mansoor & Munshi (2015)	Practicing physicians	Various observer reports (peer colleagues, senior medical colleagues, colleagues from other departments) using a 6-point likert-type rating scale	Clinical medical practice environment	To assess professionalism, communication skills, and collaboration	A working group and expert opinion were involved in developing the measure
4	ABIM questionnaire (Farsi translation)	Aramesh, Mohebbi, Jessri & Sanagou (2009)	Graduate medical trainees (residents)	Not reported	Not reported	To measure professionalism	Triangulated translation followed by triangulated back translation using independent experts Experts were also consulted regarding content validity
5	Adapted from ABIM Questionnaires	Brinkman, Geraghty, Lanphear, Khoury, Gonzalez, DeWitt & Britto (2007)	Graduate medical trainees	Patient/nurse ratings	Clinical medical training/assessment environment	For assessments undertaken by the USA professional regulatory body for registration purposes	Not reported

6	Not reported	Bumgarner, spies, Asbill & Prince (2007)	Pharmacy students	Self-report using a 4-point likert-type rating scale	Educational pharmacy training environment	Purpose-designed outcome measure for this study	Not reported
7	Institute on medicine as a profession survey on medical professionalism	Campbell, Regan, Gruen, Ferris, Rao, Cleary & Blumenthal (2007)	Practicing physicians regulated by the ABIM	Self-report using a mixture of likert-type rating scales of differing number of response options and labels, and yes/no response options	Clinical medical practice environment	To enhance understanding of physicians' attitudes towards and compliance with professional norms, and factors influencing professional behaviours	Researchers and experts were consulted using focus groups and cognitive interviews
8	GMC patient questionnaire and GMC colleague questionnaire	Campbell, Richards, Dickens, Greco, Narayanan & Brearley (2008)	Physicians registering/ revalidating with the GMC	Patient/colleague ratings using a 5-point likert-type rating scale with differing response options/labels, and yes/no response options	Clinical medical practice environment	To obtain summative multisource feedback on the performance of individual doctors, specifically for use in revalidation (registration)	Validation of existing measures
9	GMC patient and colleague questionnaires	Campbell, Roberts, Wright, Hill, Greco, Taylor & Richards (2011)	Practicing physicians in the UK	Patient/colleague report using a 7-point likert-type rating scale differing response options/labels, and additional response options for N/A and 'don't know'	Clinical medical practice environment in the UK	To support doctors in obtaining feedback from their patients and colleagues	Validation of an existing measure
10	Hall's Professionalism Scale	Chan, Chan & Scott (2007)	Not reported	Self-report using a 5-point likert-type rating scale, with some reverse-scored items	Not reported	To measure structural (professionalization process) and attitudinal aspects of professionalism	Validation of an existing measure
11	Social media professionalism scale	Chisholm-Burns, Spivey, Jaeger, Williams & George (2017)	First year pharmacy students	Self-report using a 5-point likert-type rating scale	Educational pharmacy training environment	To measure pharmacy student attitudes toward social media professionalism	Investigators generated items based on regulator guidance, with additional items from other questionnaires and those designed to meet University policy/guidance regarding social media use

12	Professional self identity questionnaire	Crossley & Vivekananda-Schmidt (2009)	Health and social care students in the UK	Self-report using a 7-point likert-type rating scale, with an additional 'N/A' response option	Educational training environment	To monitor the development of professional self-identity across different health and social care professions	Researcher item generation with input from professionals and students before refinement
13	Not reported	Dwyer, Takahashi, Hynes, Herold, Wasserstein, Nousiainen ... Ogilvie-Harris (2014)	Postgraduate orthopedic medical trainees	Faculty completed ratings using a 5-point likert-type rating scale	Educational assessment environment	To assess multiple intrinsic roles defined by the Royal College of Physicians and Surgeons of Canada (CanMEDS)	Researchers and other faculty members designed with input from a focus group of academic orthopedic specialists
14	Not reported	Dyrbye, Harper, Moutier, Durning, Power, Massie ... Shanafelt (2012)	Medical students in the USA	Self-report using a 5-point likert-type rating scale	Educational medical training environment	Purpose-designed to assess professional behaviours and beliefs within this study	Researchers combined items from a number of different previous studies, supplemented with researcher-written additional items
15	Not reported	Dyrbye, Massie, Eacker, Harper, Power, Durning ... Shanafelt (2010)	Medical students in the USA	Self-report using a mixture of yes/no and 5-point likert-type rating scale response options	Educational medical training environment	Purpose-designed to measure multiple dimensions of professionalism within this study	Items were taken from various existing measures
16	Not reported	Gillespie, Paik, Ark, Zabar & Kalet (2009)	Postgraduate medical trainees in clinical practice in the USA	Self-report using a 4-point likert-type rating scale	Clinical medical training environment	To enable self-report of professional competence	Researcher generated items based on the conceptual frameworks of various regulatory bodies, with input from curriculum steering committee
17	Finnish version of the Professionalism Mini-Evaluation Exercise	Karukivi, Korteganas-Savolainen, Saxén & Haapasalo-Pesu (2015)	Not reported	Multi-source ratings (self/tutor/"work community") using a 4-point likert-type rating scale	Used with medical students, but unclear as to whether in an educational and/or clinical context	To assess medical professionalism	Validation of an existing measure
18	Professionalism inventory scale	Kaufman & Ricci (2014)	Junior and senior level hospitality students in the USA	Self-report using a 5-point likert-type rating scale	Educational hospitality management training environment	To determine students' professional skills based on their own perspective	Researcher developed based on the results of a focus group with hospitality industry professionals

19	Professionalism Assessment Tool	Kelley, Stanke, Rabi, Kuba & Janke (2011)	Undergraduate pharmacy students in the USA	Self-report using a 5-point likert-type rating scale	Educational pharmacy training environment	To measure professionalism in pharmacy students in the USA	Researcher developed based on regulator definitions, and existing measures and tools
20	Professional Assessment Scale	Klemenc-Ketis & Vrecko (2014)	Medical students	Self-report using a 5-point likert-type rating scale	Educational medical training environment	To assess professionalism in medical students, based on their perceptions of and attitudes towards professionalism in medicine	Items generated using focus groups and the Delphi approach, followed by quantitative validation and review
21	Youth Worker Professionalism Scale	Krauss, Idris, Tamam, Suandi, Ismail, Bandar & Dahalan (2012)	Youth work practitioners	Self-report using a 5-point likert-type rating scale	Youth worker practice environment in Malaysia	Not reported	Researcher developed items by combining and adapting items from existing scales
22	Professionalism measurement tool for physicians and nurses	Lombarts, Plochg, Thompson & Arah (2014)	Practicing physicians and registered nurses	Self-report using a mixture of a 5-point likert-type rating scale and yes/no responses	Clinical medical environment	To measure the professionalism of both physicians and (registered) nurses	Researcher developed items by combining existing measures
23	Conscientiousness Index	McLachlan, Finn & Macnaughton (2009)	Undergraduate medical students	Awarding/deduction of conscientiousness points by staff	Educational medical training environment	Student assessment	Not reported
24	360 degree feedback	Meng, Metro & Patel (2009)	Graduate medical trainees (residents) on clinical rotation	Observer ratings using a 9-point likert-type rating scale	Clinical training environment	Workplace performance evaluation	Researcher developed based on key outcomes of training
25	Not reported	O'Sullivan & Toohey (2008)	Undergraduate medical students in an Australian university	Self-report using scenarios with three behavioural responses	Educational medical training environment	To assess the professionalism of undergraduate medical students	Researcher developed scenarios based on documented real-world example situations
26	Professionalism assessment scale for nurse educator	Pareek, Batra & Kalia (2016)	Nurse educators in various government and private nursing institutions in Punjab and Chandigarh	Self-report using a 6-point likert-type rating scale	Not reported	Not reported	Not reported

27	The Rasch Measurement Model instrument	Peeters & Stone (2009)	Recent pharmacy graduates of the University of Toledo	Self-report using 6 different rating scales encompassing a mixture of yes/no and likert type rating scales of varying length and response options/labels	Higher Education alumni communications	To objectively measure professional attitudes and the behaviours of recently graduated pharmacists	Not reported
28	Chisholm professionalism instrument	Poirier & Gupchup (2010)	Student and graduates of pharmacy in the USA	Uses a 5-point likert-type rating scale, but it is unclear who completes the ratings	Refers to original validation paper that violated inclusion/exclusion criteria	Refers to original validation paper that violated inclusion/exclusion criteria	Refers to original validation paper that violated inclusion/exclusion criteria
29	360 degree evaluation	Qu, Zhao & Sun (2010)	Graduate medical trainees (residents) in China	Ratings by various colleagues	Clinical medical training environment	To assess graduate (resident) professionalism	Not reported
30	360 degree feedback instrument from EOS	Qu, Zhao & Sun (2012)	Graduate medical trainees (residents) in China	Observer ratings using a 5-point likert-type rating scale	Clinical medical training environment in China	To assess performance against outcomes for residents	Not reported
31	360 degree evaluation of professionalism	Reed, West, Mueller, Ficalora, Engstler & Beckman (2008)	Graduate medical trainees (residents) in the USA	Observer ratings using a 5-point likert-type rating scale	Clinical medical training environment in the USA	Longitudinal evaluation of professionalism	Research team developed items
32	The German Professionalism Scale	Roos, Pfisterer, Krug, Ledig, Steinhauser, Sxecsenyi & Goetz (2016)	General Practitioner trainees within a specified training programme in Germany	Self-report using a 4-point likert-type rating scale	A specified General Practitioner training programme in Germany	To assess the professional behaviour of General Practitioner trainees in the Netherlands, validated in this study for use in Germany	Refers to original validation paper that violated inclusion/exclusion criteria
33	Professionalism Index	Sawdon, Whitehouse, Finn, McLachlan & Murray (2017)	Trainee anaesthetists in a UK medical school	Observer ratings using a 3-point likert-type rating scale	Clinical medical training environment	To capture subjective views on trainee anaesthetists' professionalism	Not reported
34	Professionalism score	Snider & Johnson (2014)	Osteopathic medical students	Staff award/deduct points based on self-report, observations, and peer evaluations	Educational medical training environment in the USA	Educational assessment	Not reported

35	Not reported	Symons, Swanson, McGuigan, Orrange & Akl (2009)	Graduate medical trainees (residents) in a USA hospital	Self-report using a 5-point likert-type rating scale	Clinical medical training environment	To enable self-assessment of communication skills and professionalism by graduate medical trainees (residents) in the USA	Researcher developed by modifying a regulator questionnaire from patient to self-report
36	The Barry Questionnaire	Tokuda, Barnett, Norisue, Konishi, Kudo & Miragi (2009)	Refers to original validation paper that violated inclusion/exclusion criteria	Self-report by selecting one of "4 or 5" responses to scenarios	Refers to original validation paper that violated inclusion/exclusion criteria	Refers to original validation paper that violated inclusion/exclusion criteria	Refers to original validation paper that violated inclusion/exclusion criteria
37	Amsterdam Attitudes and Communications Scale (adapted for use with non-Dutch medical graduates)	Tromp, Rademakers & Ten Cate (2007)	Not reported	Observer ratings using a 7-point likert-type rating scale	Not reported	Not reported	Not reported
38	Nijmegen Professionalism Scale	Tromp, Vernooij-Dassen, Kramer, Grol & Bottema (2010)	Graduate trainee General Practitioners in The Netherlands	Self and observer ratings using a 4-point likert-type rating scale	Clinical medical training environment	Formative assessment of trainee graduate General Practitioners in The Netherlands	Not reported
39	Professionalism Mini-Evaluation Exercise	Tsugawa et al. (2011)	Graduate clinical trainees (residents and fellows)	Observer ratings using a 4-point likert-type rating scale	Clinical medical training environment	To assess professionalism in residents	Refers to original validation paper that violated inclusion/exclusion criteria, researchers only involved
40	P-MEX (Japanese modified version)	Tsugawa et al. (2009)	Clinical trainees	Observer ratings using a 4-point likert-type rating scale	Clinical training environment	Not reported	Items from the original measure development study were supplemented by additional items developed using workshops involving doctors and nurses
41	Multisource feedback instrument	Violato, Lockyer & Fidler (2008)	Practicing psychiatrists in Canada	Self and observer ratings using a 5-point likert-type rating scale	Clinical psychiatry practice environment in Canada	To assess key competencies of practicing psychiatrists	Items developed using a working group of practicing clinicians from a range of disciplines

42	Not reported	Wang, He, Miao, Huang, Lu & Chen (2017)	Healthcare workers in China	Self-report using a 5-point likert-type rating scale	Healthcare practice environment	To assess medical professionalism in young healthcare workers in China	Not reported
43	Professionalism Assessment of Clinical Teachers	Young, Cruess, Cruess & Steinert (2014)	Clinical teachers in undergraduate medical education	Observer ratings using a 5-point likert-type rating scale	Clinical medical training environment in Canada	To assess the professional behaviours of clinical teachers	Researcher developed based on curriculum and extant student assessments
44	Standardised patients' scores of student professionalism	Zanetti et al. (2010)	Undergraduate medical students in the USA	Observer ratings but no information regarding scale	Educational medical training environment in the USA	To assess student professionalism in undergraduate medical education	Not reported
45	360 degree evaluation instrument for assessing surgery residents' competency in professionalism	Zhao, Zhang, Chang & Sun (2013)	Graduate surgical trainees (residents) in China	Observer ratings using a 5-point likert-type rating scale	Clinical medical training environment	To assess professionalism and interpersonal and communication skills in graduate surgical trainees (residents) in China	Not reported

5.3.1 Psychometric quality.

The number of papers receiving each score per section are included in table 5.3, along with mean question and section scores. No measures achieved the maximum possible score of 129. The highest scoring paper was Al Ansari et al. (2015) with a total score of 58 and the lowest Karukivi, Korteganas-Savolainen, Saxén, and Haapasalo-Pesu (2015) with a total score of 26. No sections achieved an 'adequate' (2 or above) average score. The section with the lowest average score was floor/ceiling effects ($M=0.11$) and the highest average scoring section was rationale ($M=1.49$).

5.3.1.1 Measure rationale.

No papers achieved the maximum possible score for this section and only two achieved an 'adequate' score across all questions (Krauss et al., 2012 and O'Sullivan & Toohey, 2008). The mean score across all papers was 1.49. Papers were strongest in describing administration conditions, both in terms of mean score per question ($M=1.96$) and the question eliciting the highest number of 'good' scores. Papers were weakest in using appropriate item generation procedures, both in terms of mean score per question ($M=0.98$) and the question eliciting the highest number of 'very poor' responses.

The most common reasons for deducting points were lack of clarity, including brief or vague/unclear descriptions of the aims of a measure, the definition of the construct, the way the measure was constructed, target participants, and conditions and procedures for administration. There were instances of a failure to provide any description of a given element, with key aspects of the rationale instead being implied and relying upon reader inference from context. Regarding theoretical rationales, the overwhelming majority of measures were not based on a construct grounded in existing theory. In the absence of theory, constructs themselves were also generally poorly defined and vaguely described.

5.3.1.2 Measure materials and documentation.

No papers achieved the maximum possible score for this section, and only one achieved an ‘adequate’ score across all questions in this section (Campbell et al., 2011). The mean score across all papers was 1.08. Papers were strongest in clearly describing objective scoring systems, both in terms of mean score per question (M=1.71) and the question eliciting the highest number of ‘good’ scores. Papers were weakest in issuing clear participant instructions and unambiguous items free from group-specific wording (highest number of ‘very poor’ scores) and providing clear and comprehensive observer instructions (lowest mean score; M=0.42).

The most common reason for deducting points in this section was failure to provide any information upon which to base an evaluative decision. For participant and observer instructions, this reflected the majority of papers. The majority of papers did clearly describe an objective and standardised scoring system, but the proportion of measures omitting this information altogether remained sizeable.

5.3.1.3 Norms.

No papers provided any information regarding interpretation norms. The mean score across all papers was 0.07.

5.3.1.4 Reproducibility.

No measures achieved an ‘adequate’ score across all questions in this section. The mean score across all papers was 1.32. Papers were strongest in undertaking Cronbach’s alpha assessment using appropriately sized participant samples (more ‘good’ scores than any other question) and reporting correctly followed procedures (highest mean score; M=1.76). The weakest areas were assessing parallel or alternate forms reliability (lowest mean score; M=0.98) and undertaking Cronbach’s alpha assessment using appropriately sized participant samples (more ‘very poor’ scores than any other question).

The most common reasons for deducting points in this section related to using limited methods to provide evidence of reproducibility, and incorrectly or

inappropriately applying reliability metrics where they were used. The most commonly reported statistic was Cronbach's alpha, but the majority of papers used this with insufficient participant numbers, applied it only partially such as to an overall scale where multiple scales were used, or returned results below the threshold stated as acceptable within the scoring criteria. Where factor analysis was reported, there were recurring issues of this being with too small a participant sample. The use of alternative approaches to assessing reproducibility were few, with no papers reporting test-retest or parallel/alternate forms evidence for all dimensions resulting in acceptable results. Only one paper reported using either Item Response Theory or Generalisability Theory respectively during validation.

5.3.1.5 Validity.

No measures achieved an 'adequate' score across all questions in this section. The mean score across all papers for validity was 1.15. The measures reviewed were strongest in using clear and consistent response options of an appropriate number, both in terms of mean score per question ($M=1.70$) and the question eliciting the highest number of 'good' scores. The weakest area was taking appropriate approaches to ensuring content validity during measure construction to ensure the full breadth of the construct was comprehensively captured, both in terms of mean score per question ($M=0.79$) and the question eliciting the highest number of 'very poor' scores.

The most common reasons for deducting points in this section were failure to undertake factor analysis to assess dimensionality, doing so with inappropriately sized samples, and/or using limited techniques for interpretation. No papers reported criterion validation evidence, and only one reported convergent/discriminant validity evidence at acceptable levels. Although many papers neglected to provide copies of measure items, the majority of those that did included ambiguous and group/profession-specific wording. The majority of measures also relied upon response options that were inconsistent and unjustified (in the case of ipsative response options).

5.3.1.6 Floor/ceiling effects.

Only one measure achieved an ‘adequate’ score in this section (Tromp, Rademakers & Ten Cate, 2007). The mean score across all measures was 0.11. Only four papers referred to floor/ceiling effects (Crossley & Vivekananda-Schmidt, 2009; Sawdon, Whitehouse, Finn, McLachlan & Murray, 2017; Symons, Swanson, McGuigan, Orrange & Akl, 2009; Tromp et al., 2007) with all others including no information on which to base decisions. Of those reporting floor/ceiling effects, only one paper described steps taken during development to address them (Tromp et al., 2007).

Table 5.3*Number of papers receiving each score for each quality criterion.*

Question	Number of papers receiving each score			
	0	1	2	3
Section 1: Rationale				
1.1 a)	3	8	27	7
b)	5	32	5	3
c)	0	11	30	4
d)	0	11	25	9
1.2	2	35	7	1
1.3 ^{*2}	20	9	9	5
Section 2: Measure materials and documentation				
2.1	28	12		5
2.2	10	14		21
2.3 ^{*19}	19	3	4	0
2.4 ^{*1}	16	5	4	19
Section 3: Norms				
3.1	5	40		0

3.2	a)	i) ^{*45}	0	0	0	0
		ii) ^{*45}	0	0	0	0
		iii) ^{*45}	0	0	0	0
	b) ^{*45}		0	0	0	0
	c) ^{*45}		0	0	0	0
3.3	a) ^{*45}		0	0	0	0
	b) ^{*45}		0	0	0	0
	c) ^{*45}		0	0	0	0
3.4	a) ^{*45}		0	0	0	0
	b) ^{*45}		0	0	0	0
	c) ^{*45}		0	0	0	0
<hr/>						
Section 4: Reproducibility						
4.1	a)		0	21	14	10
	b)		8	22		15
	c)		0	33		12
4.2			0	24	16	5
4.3			1	43	1	0
4.4			1	44	0	0
4.5			0	44	0	1
4.6			1	36	7	1
<hr/>						
Section 5: Validity						
5.1			0	37		8

5.2	a)		4	30	5	6
	b)	i)	14	14	12	5
		ii) ^{*1}	4	19	7	14
	c)		1	43	0	1
	d)	i)	2	32	11	0
		ii)	0	38	2	5
		iii)	0	44	1	0
5.3	a)	i)	0	45	0	0
		ii)	0	45	0	0
		iii)	0	45	0	0
5.4 ^{*2}			17	19	6	1
<hr/>						
Section 6: Floor/ceiling effects						
6.1			41	0	3	1
<hr/>						

Note. ^{*n}Indicates the number of papers to which this question was deemed not applicable.

5.4 Discussion

Developing professionalism is a major challenge within education and industry (Carter et al., 2015; Evans, 2008; Evetts, 2014), but research is limited by the availability of high quality, extensively validated measures from which to draw conclusions and inform practice (Goldie, 2013; Jha et al., 2007). The present study appraises the current state of professionalism measurement by comparing extant measures published between 2007 and 2017 to best practice psychometric criteria. A systematic search identified 45 publications using evidence from measures of professionalism to support their conclusions. These publications were subjected to a review of their methodological rigour specifically relating to psychometric validation. Within the current study, no measures were found to meet acceptable levels under the quality assessment framework.

5.4.1 Psychometric quality.

5.4.1.1 Measure rationale.

One of the most common failings in this area was relying on reader inference, such as where the aims of measures were implied through contextual information but not explicitly stated. The majority of papers neglected to define the construct of professionalism entirely, instead relying upon congruence between the definition used and those of readers. This is particularly problematic for professionalism as a concept plagued by controversy around its definition, even in areas such as medicine where its importance has been discussed for decades (Aguilar et al., 2011; Anderson et al., 2014; Blake & Gutierrez, 2011; Birden et al., 2014; Bonke, 2006; Buck et al., 2015; Evans, 2008; Finn et al., 2010; Goldie, 2013; Monrouxe et al., 2011; Wilkinson et al., 2009). Throughout the literature, definitions of professionalism vary widely and although it is possible to identify a number of core characteristics by their frequent occurrence, even these are without consensus. The implication of this is that the validity of measures is undermined from the outset, as readers cannot assess whether the inferences made from its data are appropriate, given that they remain naïve as to the target construct (Furr, 2011). Similar implications result from a lack of detail regarding measure construction, target participants, and

administration procedures, as readers are unable to assess whether conclusions are appropriate in light of these missing details. In practice this means that subsequent validation activities are undermined by inadequate evidence in this initial but crucial stage (Furr, 2011).

5.4.1.2 Measure materials and documentation.

The majority of papers failed to provide information regarding materials and documentation. Evaluating the clarity and objectivity of participant and observer instructions is an important step in appraising the validity of an instrument (Simms & Watson, 2007), as clear, unambiguous instructions equally relevant to all test users ensure interpretations of items do not vary inter- or intra-individually across different timepoints or contexts (Krosnick, Judd & Wittenbrink, 2005; Saville & MacIver, 2017). Without opportunities to review these aspects of measures, readers cannot ascertain whether interpretations or group-specific items have introduced systematic error into test scores and are therefore unable to assess whether conclusions drawn are sound (Furr, 2011).

A stronger aspect within in this section was clearly describing objective scoring systems. This minimises errors by reducing ambiguity in scoring protocols and the need for scorer interpretation (Furr, 2011), thus increasing validation evidence by ensuring integrity of data. Although a slight majority of papers included this information, the remainder were negligent in describing their scoring system. Although in practice these papers may have used clear, standardised, and objective scoring systems, this could not be assumed and therefore could not form part of the validation argument. Omitting such details has been criticised by previous reviews of professionalism measurement, with authors declaring it to be one of the major contributors to weak validation (Li et al., 2017).

5.4.1.3 Norms.

The lack of reported interpretation norms was concerning. Robust interpretation norms enable test users to interpret scores in the context of inter-individual differences (Evers et al., 2010). Inter-individual differences are of interest

where test users seek to benchmark performance against a given population, such as in educational assessment where students are required to meet the performance of a given percentile of the population in order to be considered proficient. This is of particular interest to the papers reviewed as many used professionalism assessment within educational contexts. The lack of interpretation norms relating to measuring professionalism is not new (Ben-David et al., 2004) and is perhaps unsurprising given that robust interpretation guidelines require the collection of large data sets, meaning it can be lengthy and expensive (Evers et al., 2010; Furr & Bacharach, 2014). Consequently, although the initial publication of a measure may not include interpretation norms, a discussion of their importance and subsequent research to provide them would be expected as part of an ongoing process of validation. The failure of all papers to discuss this issue leaves a hole in their validation arguments.

5.4.1.4 Reproducibility.

Reproducibility/reliability evidence supports appropriate interpretation of test scores (Cook & Beckman, 2006; Furr, 2011; Kane, 2013). Stronger reliability evidence enables more confident inferences as to the meaning of scores. Reliability is statistically appraised (Newton & Shaw, 2014; Simms & Watson, 2007) but there are limitations to all available metrics when used in isolation, so a robust validation argument requires a range of evidence (Kane, 1990; Messick, 1979; Newton & Shaw, 2014). The papers reviewed relied strongly upon Cronbach's alpha as the sole form of reproducibility evidence, which is a common finding within the broader psychometric literature (Borsboom, 2005; Simms & Watson, 2007). Cronbach's alpha is derived from classical test theory, which states that scores elicited by measures reflect a combination of true scores and error, and that these are uncorrelated so error must therefore be random (Barchard & Hakstian, 1997; Borsboom, 2005; Furr, 2011; Krosnick et al., 2005; Lumsden, 1976; Raykov, 1997; Sijtsma, 2012a; Starkweather, 2012; Yang & Green, 2011). The tenet of random error is a fundamental assumption of Cronbach's alpha, but there is consensus that measurement can never be free from systematic error, including that introduced by methods of measurement or participant bias (Borsboom, 2005; Maul, 2013; Yang & Green, 2011). This suggests that the fundamental assumption of Cronbach's alpha

that error is random is universally violated, meaning that its contribution to validation evidence is essentially flawed. Overreliance on Cronbach's alpha within the papers reviewed is compounded by failures to acknowledge this limitation throughout the psychometric literature (Borsboom, 2005; Lumsden, 1976; Yang & Green, 2011).

Results also indicated that Cronbach's alpha was often misused or misapplied. As was also found by Li and colleagues (2017) in their review of medical professionalism measures, authors failed to apply internal consistency statistics to all dimensions of measures, thus drawing conclusions based on incomplete empirical evidence. However, the most common issue was that they were used with data from too few participants. Similarly, where factor analyses were used as an additional reliability metric, this was also commonly with too small a sample. The application of these statistics to insufficient participant data limits the generalisability of findings and in the case of factor analysis, adversely impacts the stability and clarity of factor structures, meaning that while they may assess reliability of the particular measure when used with that particular sample, findings cannot be confidently interpreted as applying to different participant samples (Auerswald & Moshagen, 2019; Costello & Osborne, 2005; Furr, 2011; Igundunasse, 2016). Once again, this undermines the validity argument for measures as inferences cannot be made using data collected from different participants (Kane, 2013).

More modern techniques for addressing reproducibility were severely lacking in the papers reviewed, despite some evidence for their superiority. Evidence suggests that the results of reliability metrics based upon classical test theory, such as Cronbach's alpha, vary according to samples used to gather data (Barchard & Hakstian, 1997; Borsboom, 2005; Fan, 1998; Hambleton & Jones, 1993; MacDonald & Paunonan, 2002). Metrics based on more modern developments, however, such as item response theory (IRT) may be more stable and less influenced by such extraneous details (Fan, 1998; MacDonald & Paunonan, 2002; Rudner, 1983; Tinsley & Dawis, 1977). Using IRT may therefore eliminate concerns inherent to Cronbach's alpha. Having said this, the available evidence remains mixed, so it would be prudent for researchers to use both approaches in order to ensure the strongest validation argument (Cook et al., 1988; Fan, 1998; MacDonald &

Paunonan, 2002; Rudner, 1983; Sijtsma & Emons, 2013). In agreement with the findings of Li and colleagues (2017), the papers reviewed did not evidence this, and so the overall evidence for robust reliability assessment was weak, once again undermining arguments for validity.

5.4.1.5 Floor/ceiling effects.

Floor/ceiling effects are present where a sizeable proportion of individuals responding to a measure score the highest/lowest possible score (Garin 2014; Terwee et al., 2007). This may occur where items fail to represent extremes of a given behaviour (Garin, 2014), which is psychometrically relevant for several reasons. Firstly, if items within a measure fail to sample extreme behaviours, they also fail to sample the entire domain being measured, thus limiting evidence for content validity (Garin, 2014; Terwee et al., 2007). This is a particular issue in educational assessment where practice involves sampling curricula content to develop test items (American Psychological Association, 1954; Guion, 1980; Kane, 2013), so is particularly pertinent to the studies reviewed here given that so many of them had been developed and/or applied in educational or training contexts. Secondly, their derivation from classical test theory means that statistical reliability estimates speak to the precision of a measure in recording true scores versus random error (Borsboom, 2005; Sijtsma, 2012a; Yang & Green, 2011). This requires measures to detect the full extent of variance within samples, including that at the uppermost and lowermost score extremes. Floor/ceiling effects prevent this and therefore undermine the accuracy of reliability estimates and related validation arguments (Garin, 2014; Terwee et al., 2007). Finally, specific to intra-individual measurement, floor/ceiling effects limit the ability to track change over time for low/high performers (Garin, 2014; Terwee et al., 2007). Where intra-individual assessment is a test user's goal, inferences cannot be made based on this missing information, and so the test is not validated for its primary aim (Garin, 2014; Terwee et al., 2007).

Overall, the impact of floor/ceiling effects relates to validity and so the most important factor is that readers be informed as to their presence/extent or absence, in order to enable informed decisions regarding impact on inferences. Within this review, the absence of floor/ceiling effect reporting means that this aspect of

validation evidence was missing, and that none of the measures reviewed could therefore be assumed as free from the above limitations.

5.4.1.6 Validity.

Although viewed today as the most central issue in validation and psychometric theory (Cook et al., 1988; Embretson & Gorin, 2001; Hogan & Agnello, 2004; Kane, 1990), validity assessment in the papers reviewed was limited. Factor analysis was heavily relied upon to assess dimensionality, with such overreliance upon empirical validation representing an atheoretical approach (Newton & Shaw, 2014; Simms & Watson, 2007). This carries the risk of creating statistically robust measures that fail to hit their intended mark (Travers, 1951). Given the issues relating to the lack of theoretical understanding of professionalism discussed earlier, this concern is rendered even more significant. Factor analysis requires researchers to use their judgement in identifying and describing dimensions and so without sound grounding in *a priori* theory, such descriptions are without justification and subject to author subjectivity and bias (Bringmann & Eronen, 2016; Sijtsma, 2012a, 2012b; Sijtsma & Emons, 2013; Simms & Watson, 2007). As a result, best practice requires both empirical and logical validation evidence be presented.

It was concerning to find that logical form validation evidence, such as content validity evidence, was also lacking. This was surprising in the context of previous literature, which suggests educational measures are more likely to rely on content validation evidence than empirical (Kane, 2013; Newton & Shaw, 2014; Travers, 1951). This was not supported by the results of this study, with one potential reason lying in misunderstanding of the requirements for robust content validation. Best practice is to consult experts when developing measures, including expertise in both psychometry and the construct under study, and to include feedback from target participants regarding item clarity and the potential role of interpretation. This helps to identify biased, errant, or conflicting understanding of construct theory, enabling experts to discuss points of contention and ensure constructs remain clearly articulated and closely adhered to (Kane, 2013; Simms & Watson, 2007). Neglecting such evidence in developing measures of professionalism may stem from limited or

incorrect understanding of this, specifically of what constitutes sufficient expertise. Researchers unfamiliar with the detail of psychometric theory may rely upon their own knowledge of the literature as sufficient to assure content validity, therefore taking a convenience approach to sampling expertise rather than seeking more comprehensive input (Veloski et al., 2005). In reality, strong validation arguments require broader evidence from a range of experts (Kane, 1990; Messick, 1979; Veloski et al., 2005). Without this, validity arguments are fundamentally flawed.

5.4.1.7 Ethical considerations.

Although attracting controversy (Cizek et al., 2008; Furr & Bacharach, 2014; Zumbo, 2007), best practice guidance recommends that validation considers ethical issues in interpreting test scores (Messick, 1979, 1990, 1995). Although this did not form part of the appraisal process in this study, the findings warrant discussion in this context. In order for decisions based on psychometric tools to be ethical, the strength of their validation argument ought to correspond to the weight and consequences of those decisions (Jonson & Plake, 1998; Miller & Lovler, 2016; Newton & Shaw, 2014). In the context of professionalism, decisions based on the outcomes of measures have life-altering consequences, such as those relating to qualification, registration, or employment. The findings of the present study suggest that the validation arguments for extant measures of professionalism are weak, thus indicating that their use in this way is unethical.

5.4.2 The context and relevance of findings.

Although presenting clear cause for concern, the findings of this review are not unexpected, as previous research has suggested that the use of psychometric assessments suffers from long-standing misunderstandings of psychometric theory and validation best practice (Borsboom et al., 2004; Cizek et al., 2008; Duckor, 2017; Hogan & Agnello, 2004; Jonson & Plake, 1998; Messick, 1979; Wolming & Wikström, 2010). Previous reviews suggest that professionalism measurement contributes to this issue, even when more limited psychometric quality criteria are used (Goldie, 2013; Jha et al., 2007; Simms & Watson, 2007). This is congruent with

the findings of this review and in line with claims of persistent gaps between contemporary validation theory and practice (Hogan & Agnello, 2004; Jonson & Plake, 1998; Kane, 1990; Sijtsma, 2012a), particularly in educational contexts (Wolming & Wikström, 2010).

The importance of the findings reported here lies in their relation to validity. Validity is the single most important aspect determining the quality of measures (Cizek et al., 2008; Embretson & Gorin, 2001; Hogan & Agnello, 2004; Kane, 1990; Newton & Shaw, 2014), as underpinned by each of the quality criteria used within this review. Validity refers to the body of evidence supporting or refuting the appropriateness of inferences made from data gathered using a measure (Hogan & Agnello, 2004; Kane, 2001; Messick, 1990). Without validity, no meaningful conclusions may be drawn from that data (Cizek et al., 2008; Furr, 2011; Kane, 1990; Messick, 1990; Wolming & Wikström, 2010). The validation evidence required varies in strength according to the purpose of measures and how their results may be used (Bringmann & Eronen, 2016; Furr, 2011; Kane, 2001; 2013; Messick, 1990; Newton & Shaw, 2014). Measures intended to assess intra-individual professionalism to support development may require less strong bodies of validation evidence than those intended to make important interpersonal decisions, such as those regarding employment. As many of the measures reviewed report providing evidence for formative and summative educational assessment, institutions may be drawing inappropriate and unethical conclusions regarding the fitness of students to practice, thus increasing risks to stakeholders and the reputations of related professions (Carey & Ness, 2001).

5.4.3 Recommendations for practice and further research.

There is no extant gold standard measure of professionalism recommended for use. Individuals seeking to measure professionalism are recommended to do so with caution, giving full consideration to the issues discussed within this review. It is recommended that conclusions drawn using extant measures of professionalism be highly caveated and not used as a basis for meaningful decision making.

Regarding future research directions, although best practice in psychometric validation suggests that it is preferable to improve existing measures over creating new ones (Furr, 2011; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004), this argument carries an assumption that those measures target theoretically defined constructs. Given that all measures reviewed here fail to meet this fundamental criterion, it is recommended that further work is required before the construction of a new measure begins. Specifically, there is a gap in the literature requiring a theory-led definition of professionalism that may be clearly articulated and empirically tested before being used as the basis of measurement.

5.4.4 Strengths and limitations of this review.

This review is the first to appraise extant measures of professionalism from a methodological standpoint using a systematic search strategy encompassing all occupational sectors, and robust and comprehensive assessment criteria based on gold standard research and practice guidelines. Previous reviews have been undertaken in relation to specific occupational sectors (e.g. Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; Veloski et al., 2005). The present review improves upon current knowledge by removing this limitation and returning to fundamental criteria for best practice psychometric assessment, regardless of the context in which it is used.

The criteria used to assess psychometric quality in this review were highly demanding, so it may be argued that few measures could demonstrate the strength of validation evidence it required (Cizek et al., 2008; Hogan & Agnello, 2004; Jonson & Plake, 1998; Newton & Shaw, 2014; Sijtsma, 2012a; Wolming & Wikström, 2010). As the strength of validation arguments required varies with the purpose of measurement, applying such a standardised threshold for ‘acceptable’ validation failed to meet best practice in responding dynamically to the individual characteristics and aims of measures. Moreover, the lowest scores available within the quality assessment framework related to a lack of available information. This meant that where other papers validating the same measure were outside of the date range, low scores may be indicative not necessarily of poor psychometric quality, but of a failure to re-describe details potentially discussed previously elsewhere. This

means that low scores against the criteria used here may not accurately reflect validation practice.

All reviews are limited by their search strategy. Where such information is relevant, it is recommended that readers update the search used in this review to include measures published after 2017. The search criteria were also limited to measures published in English and those not measuring professionalism as a component of a larger construct or a single component of professionalism. If component measures or those published in other languages are of interest, the search would require amendment accordingly. Limiting the language of publication may also have limited global cultural representation within results. Most measures reviewed described target populations in Europe, Canada, or the United States who may be argued to define professionalism in similar ways relative to non-Western cultures. The level of difference even between the conceptualisations of professionalism used in the United Kingdom and United States of America (Hafferty, 2006) suggest that practice may be entirely different within different professional cultures (Monrouxe et al., 2011), and so higher quality measurement of professionalism may be found published in non-English languages.

In terms of procedure, searches, screening, and full extraction was largely undertaken by the lead researcher alone. Although a sample of decisions were checked by another member of the research team, best practice in undertaking systematic reviews recommends that all evaluations are undertaken independently by at least two individuals (Higgins & Green, 2011). It is therefore possible that screening and scoring decisions might have been altered if two independent scorers had been used throughout.

5.5 Chapter Summary and Conclusions

The research reported in this chapter identified no measures of professionalism with psychometric validation evidence sufficient to recommend them for either research or practice purposes. Validation evidence sources were lacking or misused in the overwhelming majority of cases, with none meeting acceptable psychometric standards overall. The first step in rectifying this is to

ensure future measures are grounded in a sound and testable theoretical construct of professionalism, and that during construction, all facets of psychometric theory are considered and responded to, whether by action or discussion, to ensure readers are able to appraise the validity argument fully. Constructing new measures should be done in full recognition of the demanding and resource-intensive nature of best practice, arguments-based validation, and within a framework to ensure this long-term endeavour is realised.

Chapter 6: An Exploration of Shared Construing in Judgements of Professionalism using the Repertory Grid Technique

Abstract

Controversy surrounding the definition of ‘professionalism’ has hindered efforts to support professional development through research. In response, previous research has elicited espoused theories of professionalism from various stakeholders by asking them to articulate their perception of professionalism objectively. This study examined professionalism as a subjectively construed concept by exploring beyond the pre-planned articulations elicited by previous research and accessing the basis of judgements regarding professionalism in real-world contexts. The repertory grid technique was applied as a new approach to exploring the issue of professionalism and as a research method particularly well suited to the research aims of revealing tacit person evaluation judgements. Thirty-six completed grids were subjected to multiple groups factor analysis, with three principal components extracted that accounted for 62.9% of the study variance. These components were interpreted under the headings of ‘violation of baseline expectations’, ‘logical ruthlessness’, and ‘personal likeability’. Findings suggest that person evaluation judgements regarding professionalism are made based on these components. Previous research in this area has elicited different espoused theories of professionalism; this study contributes new understanding as to how these differ from real-world judgements of professionalism. The data also suggest that gender and likeability may play distinct and, as yet, unexplored roles in the perception of professionalism.

6.1 Introduction

Understanding professionalism is a major focus for educational researchers, owing to high profile cases of unprofessional conduct reaching mainstream audiences and leading to calls for educators and employers to do more to guarantee the professionalism of their graduates and staff (de Mendonça et al., 2016; Hammer, 2006; Lynch et al., 2004; Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013; Monrouxe & Rees, 2012). However, efforts to improve professionalism have reportedly been inhibited by lack of agreement over its definition, which has prevented the development of valid measures for use in benchmarking and regulating conduct (Birden et al., 2014; de Mendonça et al., 2016; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; Veloski et al., 2005; see also the findings of chapter 5).

Attempts to generate consensus around the construct of professionalism have predominantly involved asking individuals to explain their concept of and attitudes towards professionalism through qualitative research methods (e.g. Carter et al., 2015; Finn et al., 2010; Monrouxe et al., 2011), and many published articles offer nothing more than personal opinion or descriptions of consensus amongst the personal opinions of others (Birden et al., 2014). Using such explicit approaches to understanding professionalism can lead to biased responses (Hill et al., 2016). That is not to say that participants actively seek to deceive researchers, rather that responses elicited by research methods such as interviews and focus groups are more likely to reflect participant perceptions of the 'correct' answer, based on their previous experiences (Easterby-Smith et al., 1996).

In the context of psychological research, this includes participants providing what they expect to be the most helpful responses to questions or tasks based on subtle contextual cues provided by the research situation, communications with the research team or other participants, and inferences from study literature (Orne, 1962; Rosnow, 2002; Sharpe & Whelton, 2016; Shaughnessy & Zechmeister, 2015). These cues, known as demand characteristics, relate to research situations, but it appears reasonable to expect similar effects outside research contexts. For example, relating to professionalism this might mean that individuals who have worked exclusively within one occupation are likely to base their concept of professionalism on a mixture of their lived experience of the culture of that sector and how to demonstrate

compliance with its policies and procedures, through both explicit guidance and more subtle cultural and contextual cues. The combination of explicit guidance and implicit demand characteristics may influence individual articulations of the concept of professionalism. Indeed, some report that due to ongoing reliance upon professional socialisation as a vehicle for professional development (Anderson et al., 2014; Gleeson, 2007; Hammer, 2006), individuals training for heavily regulated occupations, and specifically so around professionalism, rapidly adapt to demand characteristics, thus being even more prepared to deliver ‘correct’ answers upon demand (Bertolami, 2004; Marchalik, 2015). However, whether the same individuals would judge the professionalism of others in accordance with the distinct demand characteristics associated with their own occupation is unclear.

The impact of demand characteristics on existing research into professionalism ought not to be ignored. A failure to transparently account for their effects undermines the validity of findings and has been described as delivering overly simplistic and naïve results ignorant of the complexities of the social context of behaviour (Rosnow, 2002; Sharpe & Whelton, 2016; Shaughnessy & Zechmeister, 2015). This criticism would affect original research that fails to account for demand characteristics as research artefacts, but also any subsequent secondary research based on its findings. The latter would include, for example, research using the initial findings as a basis for further research, such as developing a measure of professionalism.

There is a gap in the literature requiring the influence of explicit research methods and the related potential for bias on conceptualisations of professionalism to be extricated from real-world experience. This may be achieved by comparing results with those elicited by alternative methodologies. Mixed methods research that blends qualitative and quantitative research approaches is particularly effective at tackling the impact of demand characteristics, due to its recognition of both the subjective and objective aspects of human experience (Sharpe & Whelton, 2016). In addition, methodologies targeting implicit knowledge are of particular interest in that they may be able to by-pass more biased, overtly articulated accounts of professionalism and instead focus in on those aspects of individual subjective perspective that are less accessible.

The repertory grid technique (RGT) is a unique research method that explores how individuals perceive or construe the world around them (Easterby-Smith et al., 1996; Jankowicz, 2004; Kelly, 1955; Smith, 1995) by specifically targeting the implicit knowledge systems used to navigate it (Fransella et al., 2004). The RGT is cited as one of few truly mixed methodologies that integrally blends qualitative and quantitative research approaches, and is therefore particularly suited to issues where demand characteristics may be of concern (Brewerton & Millward, 2001; Cooper, 2010; Hill et al., 2016; Rocco et al., 2003; Sharpe & Whelton, 2016; Smith, 1995). The RGT is based on George Kelly's personal construct theory (PCT; Brewerton & Millward, 2001; Catania & Randall, 2015; Easterby-Smith et al., 1996; Fransella et al., 2004; Hill et al., 2016; Kelly, 1955; Smith, 1995; Walker & Winter, 2007). PCT states that individuals navigate their environment through an active and ongoing process known as construing that involves generating, testing, and reviewing hypotheses to enable the prediction of events in the external world (Butler, 2009a; Cooper, 2010; Fransella et al., 2004; Kelly, 1955). Individuals construe the external environment through an organised system of bipolar constructs, a system that defines individuals' unique ways of viewing the world around them and events within it (Butler, 2009a; Kelly, 1955). Constructs are subjective and idiographic, as are the language-related labels individuals assign to them, but several individuals may share similarity in some or all of their construing (Butler, 2009a; Easterby-Smith et al., 1996; Procter, 2009; Walker & Winter, 2007). PCT suggests that individuals understand their own construing through the stories associated with the acquisition of constructs, rather than the constructs themselves (Procter, 2009). As a result, asking individuals to articulate their own construing, such as that pertaining to professionalism, is likely to result in data reflecting their personal story of professionalism, rather than the actual constructs through-which they experience it (Butler, 2009a; Catania & Randall, 2015; Jankowicz, 2004). The RGT offers a more valid means of exploring and mapping these constructs (Butler, 2009a; Catania & Randall, 2015).

The RGT is designed to explore individual construing with minimal researcher influence (Easterby-Smith et al., 1996; Jankowicz, 2004). It provides a framework allowing individuals to describe their own constructs regarding a topic in ways that support articulations of content they may have previously been unaware of

or unable to describe (Easterby-Smith et al., 1996). The RGT therefore provides opportunities to avoid the pitfalls of previous research seeking to define professionalism by offering a less biased account of the experience of professionalism as construed by the individual (Easterby-Smith et al., 1996; Procter, 2009). The present study aims to explore the constructs used by individuals when considering the professionalism of others, as a phenomenologically grounded concept. Using the RGT, this study explores the potential for individuals from different occupational experiences and backgrounds to share similar constructs as to the nature and content of professionalism.

6.2 Methods

6.2.1 Materials and the repertory grid technique.

The RGT involves three broad phases. First, the participant is asked to generate elements to sit across the top of the grid (Brewerton & Millward, 2001; Catania & Randall, 2015; Jankowicz, 2004). These elements are taken from participants' real-world experience and, within the current study, took the form of actual individuals they had previously encountered (Brewerton & Millward, 2001; Butler, 2009a; Catania & Randall, 2015; Easterby-Smith et al., 1996). Second, the elements are re-presented to participants to elicit constructs (Butler, 2009a; Easterby-Smith et al., 1996; Jankowicz, 2004). A selection of three elements are presented at random and participants asked to choose two that, according to their perception, share something in common (Brewerton & Millward, 2001; Cooper, 2010; Easterby-Smith et al., 1996; Jankowicz, 2004). This similarity is recorded in the left-hand column of a blank repertory grid as the emergent pole of the first construct (see figure 6.1; Butler, 2009a; Cooper, 2010; Jankowicz, 2004). The implicit pole of the construct is elicited by asking participants how the third, unselected element differs from the other two; this is recorded in the right-hand column of the blank repertory grid (Brewerton & Millward, 2001; Cooper, 2010; Jankowicz, 2004). This process is repeated until participants are unable to generate further constructs (Brewerton & Millward, 2001; Cooper, 2010; Easterby-Smith et al., 1996). The final phase requires participants to rate each element according to the perceived level of similarity with each pole of each construct, usually using a likert-type rating scale (Brewerton &

Millward, 2001; Catania & Randall, 2015; Cooper, 2010; Easterby-Smith et al., 1996).

Figure 6.1. *Blank repertory grid used in the main study.*

	Extremely professional	Somewhat professional	Neither professional nor unprofessional OR equally professional and unprofessional	Somewhat unprofessional	Extremely unprofessional	Yourself	Your ideal professional	

Where grids are completed by multiple participants, they may be analysed quantitatively using data reduction methods such as factor analysis to reveal the major dimensions of shared construing among the participants (Catania & Randall, 2015; Cooper, 2010; Fransella et al., 2004; Hill et al., 2016; Jankowicz, 2004).

The RGT has been applied in a wide variety of settings (Catania & Randall, 2015), particularly within educational and organisational contexts (Easterby-Smith et al., 1996; Jankowicz, 2004; Walker & Winter, 2007) and in exploring social relationships (Catania & Randall, 2015; Neimeyer & Neimeyer, 1985), including those within a professional context (Aranda & Finch, 2003; Catania & Randall, 2015; Easterby-Smith, 1980; Easterby-Smith et al., 1996; Fassin & Van Rossem, 2009; Fransella et al., 2004; Ginsberg, 1989; Jankowicz, 2004; Rogers & Ryals, 2007). It is particularly suited to targeting tacit viewpoints, where individuals may struggle to clearly articulate their construing (Catania & Randall, 2015). As such, the RGT is particularly relevant to the current study both in terms of precedent of use and the issues relating to the challenges of articulation that the concept of

professionalism inherently brings (Catania & Randall, 2015; Easterby-Smith et al., 1996).

6.2.2 Participants.

Convenience sampling was employed to recruit participants aged 18 years or over using existing professional networks and the social media platforms LinkedIn, Facebook, and Twitter. Participants received no incentive for taking part in the study and only data from participants returning fully completed repertory grids and demographic questionnaires were retained for analysis.

6.2.3 Procedure.

6.2.3.1 Pilot study.

A small pilot study was carried out and the findings used to finalise the format of materials used in the main study. During the pilot, seven participants took part in face-to-face RGT interviews. The mean age of pilot participants was 43.6 years ($SD=9.65$), with a range of 33-58 years. Participants were fairly equal in gender ($F=4$, $M=3$) and came from a range of occupational backgrounds including education and healthcare sectors, management/leadership, security services, and the construction industry. The majority of participants were educated to postgraduate level ($n=4$) with the remaining three having completed compulsory education (GCSE or equivalent), further education, or undergraduate level education respectively.

Following a brief introduction to the subject of professionalism, pilot participants were presented with either seven or nine elements (see table 6.1) and asked to assign real individuals from their experience to each element. Anywhere from six to twelve elements are deemed optimal within the RGT, so a higher and lower number of elements were trialled to explore experiential factors relevant to participant fatigue (Jankowicz, 2004). Participants were invited to record the identities of element individuals for their own use throughout the interview and were informed that they could use initials, codes, or aliases to protect identities where preferred. They were reminded that these should still refer to actual rather than

hypothetical individuals, with the exception of the ‘ideal professional’ element. Participants were informed that they could keep the page used to record their elements or return it to the researcher for destruction, and that this data would not be used for any other purpose following conclusion of the interview. At this point, they were invited to provide feedback or comments regarding element elicitation to enable improvements to future participant experience.

Table 6.1

Elements provided to participants during pilot and main study.

Nine elements (pilot study only)	Seven elements (pilot study and main study)
Extremely professional	Extremely professional
Moderately professional	Somewhat professional
Slightly professional	Neither professional nor unprofessional
Neither professional nor unprofessional	Somewhat unprofessional
Slightly unprofessional	Extremely unprofessional
Moderately unprofessional	Ideal professional
Extremely unprofessional	Self
Ideal professional	
Self	

Participants were presented with sets of three elements (triads), with each element written on an individual card. Cards were selected at random and displayed on the table in front of them. Participants were asked to read the cards and use them as prompts to think about the individuals assigned to each element. They were asked to choose two individuals that they felt had something in common and describe this shared attribute or characteristic.

The participant and researcher discussed the attribute to clarify its content and agree a single word or short phrase to accurately describe it. This description was written into the left-hand column of a blank repertory grid template (see figure 6.1). Participants were then asked to describe that which made the third element/individual different to the two already discussed. This need not necessarily be a logical opposite of the attribute already recorded, such as ‘tidy’ versus ‘neat’. Instead, the implicit pole should capture the contrast holding maximum significance to the participant when comparing individuals, such as ‘tidy’ versus ‘disorganised’, for example. Following clarification, this attribute was recorded in the right-hand

column of the repertory grid template to create the first bipolar construct. Triads continued to be presented until participants were unable to generate any new constructs. At this point, participants were invited to provide feedback or comments on construct elicitation, to enable improvements to future participant experience.

Finally, participants were asked to rate each element against each construct using a five or seven-point likert-type rating scale within the repertory grid template (see table 6.2). Low likert ratings indicated similarity between the element and the left-hand or emergent pole of a construct, and high ratings indicated similarity to the right-hand or implicit pole. At this point, participants were invited to provide feedback or comments on the rating process, to enable improvements to future participant experience.

Table 6.2*Likert-type rating scale and guidance for pilot and main study.*

Seven-point likert-type rating scale (pilot study only)		Five-point likert-type rating scale (pilot and main study)	
1	The person is very much like the word/phrase written in the left-most box on this row This word/phrase applies to them all or almost all of the time	1	The person is very much like the word/phrase written in the left-most box on this row This word/phrase applies to them all or almost all of the time
2	The person is moderately like the word/phrase written in the left-most box on this row This word/phrase applies to them most of the time	2	The person is somewhat like the word/phrase written in the left-most box on this row This word/phrase applies to them some of the time
3	The person is slightly like the word/phrase written in the left-most box on this row This word/phrase applies to them some of the time	3	The person is neutral to the words/phrases written in the left-most and right-most boxes on this row
4	The person is neutral to the words/phrases written in the left-most and right-most boxes on this row	4	The person is somewhat like the word/phrase written in the right-most box on this row This word/phrase applies to them some of the time
5	The person is slightly like the word/phrase written in the right-most box on this row This word/phrase applies to them some of the time	5	The person is very much like the word/phrase written in the right-most box on this row This word/phrase applies to them all or almost all of the time
6	The person is moderately like the word/phrase written in the right-most box on this row This word/phrase applies to them most of the time		
7	The person is very much like the word/phrase written in the right-most box on this row This word/phrase applies to them all or almost all of the time		

Finally, participants were presented with brief demographic questionnaires for completion. Throughout all phases of the interview, and in addition to the dedicated feedback sections, participants were invited to think critically aloud when completing the RGT process, specifically around procedural aspects. The interviewer took notes during sessions, which were also audio-recorded.

The primary findings of the pilot study related to the number of elements used and the likert-type rating scale and related instructions. Feedback suggested that a grid using seven elements was manageable for participants while still retaining the level of gradation required by the study aims. Where nine elements were presented, participants commented that it was difficult to separate out the finer gradings of professionalism, causing confusion. The final elements used in the main study are listed in table 6.1. Pilot participant feedback also resulted in the likert-type rating scale being refined from seven-points to five and supplemented by additional guidance. Participants reported difficulty recalling the procedure when working with seven points, but once this was reduced to five and additional guidance was introduced, such feedback ceased. The final likert-type rating scale used in the main study is listed in table 6.2.

6.2.3.2 Main study.

The study was advertised via posts to the social media platforms LinkedIn, Facebook, and Twitter. The post included a link to complete the study and an invitation to share the post amongst readers' own networks. The online survey was created using Qualtrics (<https://www.qualtrics.com/uk/>) and could be completed using participants' personal computers or devices. Participants could complete the survey in a single sitting or by closing and returning to the survey address. Upon clicking the link, an information sheet was displayed introducing the study and topic of professionalism, and a consent form required completion before the survey could be accessed.

Participants were presented with seven elements (see table 6.1) and asked to assign real individuals to each. They were invited to enter the identities of these individuals to use as a reminder and were invited to use initials, codes, or aliases to protect identities where preferred. They were then presented with randomly generated sets of three elements (triads) in turn, with each element represented as the name/code/alias participants had previously entered. They were asked to select two of the individuals that they felt had something in common and describe this shared attribute or characteristic within a free-text box. Participants were then asked to describe the attribute or characteristic that made the third, unselected individual

different to the two previously selected. They were asked to use single words or brief phrases to describe these attributes. This process was repeated until a maximum of fifteen constructs had been generated. If participants were unable to think of constructs for a triad presented, they could skip to the next, and after the presentation of five triads, participants were given the option to quit construct elicitation and move on to the next phase of the survey. Finally, participants were asked to rate each element against each construct using a five-point likert-type rating scale (see table 6.2) and to complete a brief demographic questionnaire.

6.2.4 Analyses.

6.2.4.1 Multiple groups factor analysis.

Data gathered using the RGT with multiple participants gathers standardised observations (element ratings) from each participant based on self-generated variables (constructs). As these constructs are idiographic, analysis must account for consistent observations but gathered from different participants and using different variables, each within a separate data table. This was achieved using an approach known as multiple groups factor analysis (MGFA; Abdi et al., 2013; see chapter 4, section 4.2.2.3.2). Standard analytical software is not equipped for MGFA so software specifically developed for multi-table data was used; completed grids were correlated and subjected to MGFA as described by Abdi and colleagues (2013), using the MExPosition package (Chin Fatt, Beaton & Abdi, 2013) within RStudio 1.0.143 (RStudio Team, 2016). Completed grids were correlated and subjected to MGFA as described by Abdi and colleagues (2013), using the MExPosition package (Chin Fatt, Beaton & Abdi, 2013) within RStudio 1.0.143 (RStudio Team, 2016). MGFA involves running principal component analysis on each individual participant grid to standardise the data and ensure no single grid receives disproportionate weighting resulting from a larger number of elicited constructs. The normalised datasets are then concatenated to produce a single grid known as the compromise. The compromise is subjected to principal component analysis to reveal the major dimensions and variance accounted for by each (Abdi et al., 2013).

Principal components were extracted based on qualitative consideration of their utility in contributing to the aims of the study. Element ratings most closely associated with each principal component were identified as defining that component and the qualitative content of their bipolar constructs and comments made during grid completion used to interpret it. As this study did not seek generalisable conclusions, it was not deemed necessary to calculate statistical significance to identify defining elements, but a loading of ± 0.5 was deemed to indicate sufficient proximity of association. All elements loading to a principal component at this level or above were used in interpretation.

6.2.4.2 Cluster analysis

MGFA uses construct ratings to explore similarity but takes no account of the semantic content of self-generated constructs. Hierarchical cluster analysis was also used to explore the semantic content of the data, to classify construct poles into clusters according to their similarity (Everitt et al., 2001; Han et al., 2011; Yim & Ramdeen, 2015). Cluster analysis involves grouping items based on their similarity or dissimilarity, to reduce data to fewer, more manageable clusters for description (Everitt et al., 2001; Han et al., 2011).

All constructs were entered into a single spreadsheet and described using word vectors or embeddings. Word embeddings are numerical representations of text developed within the field of machine learning that may be used to quantify semantic similarity between words and phrases (Mikolov, Sutskever, Chen, Corrado & Dea, 2013). Search engines, for example, use word embeddings to capture associations between words. When using a search engine, word embeddings enable the returning of results that may not contain the exact target word, such as ‘swimming’ for example, but may contain other words that often relate to it, such as the name of a famous swimmer. Words with embeddings that are close together in numerical terms have more similar meaning than those more distant.

The constructs gathered within the present study were analysed using the Hclust command for hierarchical clustering within RStudio 1.0.143 (RStudio Team, 2016), using Ward’s method of linkage. Ward’s method (1963, as cited in Everitt et

al., 2001) relates to the procedure used to calculate the distance between clusters and is one of the most popular linkage methods employed in the UK (Doreian, 2004; Everitt et al., 2001). Ward's method of linkage is agglomerative, meaning that it forms clusters by first considering each data point as an individual cluster and then merging the two closest or most similar (Everitt et al., 2001; Han et al., 2011). This process is repeated until there is only one large cluster. Divisive procedures take the opposing approach, with all objects initially forming a single cluster that is successively divided until each data point occupies a single cluster (Everitt et al., 2001; Han et al., 2011). Agglomerative procedures have been widely studied, are the most commonly used approaches to hierarchical clustering, and offer advantages of scalability when used with large datasets relative to the opposing divisive procedures, and so were chosen for use within this study (Everitt et al., 2001; Han et al., 2011). The network of clusters progressing from one to many is presented in a tree or dendrogram (Doreian, 2004; Han et al., 2011; Roux, 2018; Yim & Ramdeen, 2015). The number of clusters to be retained was determined via visual inspection of the dendrogram, as a graphical representation of the similarity between different words.

6.3 Results

Thirty-nine grids were received from participants but three were excluded from further analyses due to incomplete content. Based on 36 completed grids, the total number of constructs elicited was 399, with an average of 11-12 constructs elicited per grid. The mean participant age was 35.8 years ($SD=10.85$), with a range of 20-58 years. Participants were mainly female ($F=26$, $M=10$) and had spent an average of 13.1 years in total in the workplace throughout their lives ($SD=9.77$). Participants came from a range of educational and occupational backgrounds. For a full summary of participant characteristics, see table 6.3.

Table 6.3*Summary characteristics of participant sample (n=36).*

Characteristic		N	%
Age	18-30 years	15	42%
	31-45 years	12	33%
	46-60 years	9	25%
	61+ years	0	0%
	Declined	0	0%
Gender	Female	26	72%
	Male	10	28%
	Declined	0	0%
Sector of current/most recent occupation	Administration	2	6%
	Charity	2	6%
	Creative industries	1	3%
	Customer services	3	8%
	Dentistry	2	6%
	Education	8	22%
	Healthcare	4	11%
	Manufacturing	1	3%
	Medicine	1	3%
	Management/leadership	1	3%
	Science/engineering	7	19%
	Other	4	11%
	Declined	0	0%
Time spent in workplace	Early career (0-10 years)	21	58%
	Developing career (11-20 years)	7	19%
	Established career (21-30 years)	5	14%
	Late career (31+ years)	3	8%
	Declined	0	0%
Level of education completed	Levels 1-2 (compulsory high school e.g. GCSEs)	1	3%
	Levels 3-6 (further and undergraduate higher education)	13	36%
	Levels 7+ (postgraduate higher education)	22	61%
	Declined	0	0%

6.3.1 Multiple groups factor analysis.

Thirty-six grids were subjected to MGFA using the MExPosition package (Chin Fatt et al., 2013) within RStudio 1.0.143 (RStudio Team, 2016). A three-factor solution was extracted accounting for 62.9% of the variance observed within the participant sample. Table 6.4 displays the variance accounted for by each principal component and rotated factor loadings broken down by grid element, and identifies defining loadings, the content and associated comments of which were used to interpret the principal components (see appendix B).

Table 6.4

Rotated factor loadings and summary results of multiple groups factor analysis (MGFA).

P*	Extremely professional			Somewhat professional			Neither professional nor unprofessional			Somewhat unprofessional			Extremely unprofessional			Ideal professional			Self		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
1	-0.37	0.03	-0.12	-0.34	-0.06	-0.07	0.08	-0.26	-0.02	0.53	0.09	0.29	0.66	0.28	0.03	-0.30	-0.05	-0.08	-0.26	-0.02	-0.03
2	-0.52	0.11	0.24	0.14	-0.01	0.10	-0.06	-0.07	-0.17	0.13	0.03	0.25	0.69	0.06	-0.51	-0.25	-0.14	0.00	-0.13	0.01	0.07
3	-0.20	-0.01	0.11	-0.05	-0.16	0.25	-0.03	-0.04	0.03	-0.04	-0.13	0.22	0.80	0.62	-1.10	-0.22	-0.06	0.19	-0.26	-0.22	0.30
4	-0.36	-0.05	-0.21	0.10	-0.03	0.05	-0.19	-0.17	-0.16	0.54	0.18	0.38	0.54	0.18	0.34	-0.45	-0.09	-0.29	-0.17	-0.02	-0.11
5	-0.28	0.19	-0.00	-0.05	-0.06	0.19	0.01	-0.29	-0.20	0.12	0.02	0.45	0.66	0.13	-0.71	-0.55	0.09	0.47	0.09	-0.09	-0.20
6	-0.26	0.83	-0.44	0.18	-0.72	0.20	0.18	-0.59	-0.61	-0.05	-0.17	1.37	0.24	0.08	-1.20	-0.25	0.76	0.20	-0.04	-0.20	0.49
7	-0.61	0.29	-0.19	-0.01	0.01	0.44	0.15	-0.41	-0.14	0.46	-0.22	0.59	0.50	-0.04	-0.10	-0.29	0.20	-0.21	-0.20	0.18	-0.39
8	-0.23	-0.13	-0.24	-0.28	-0.02	0.12	-0.16	-0.71	-0.18	0.44	0.39	0.27	0.65	0.80	0.08	-0.19	0.19	0.20	-0.23	-0.52	-0.25
9	-0.28	0.03	0.06	-0.17	-0.10	-0.05	-0.19	-0.41	-0.09	0.40	0.37	0.21	0.77	0.64	-0.06	-0.31	-0.17	0.03	-0.23	-0.36	-0.09
10	-0.20	0.01	0.23	-0.31	-0.15	0.17	-0.06	-0.43	-0.30	0.25	0.41	0.19	0.84	0.58	-0.43	-0.31	-0.15	0.17	-0.21	-0.27	-0.03
11	-0.06	-0.01	-0.03	-0.02	-0.17	-0.00	0.08	-0.24	0.03	0.43	0.04	0.16	0.50	0.15	0.14	-0.54	0.12	-0.17	-0.38	0.11	-0.12
12	-0.38	0.04	0.24	-0.31	0.02	0.11	0.13	-0.25	-0.36	0.30	0.26	0.41	0.79	0.05	-0.46	-0.38	0.04	0.24	-0.16	-0.16	-0.19
13	-0.40	0.15	-0.15	-0.32	-0.10	-0.30	0.12	-0.07	0.00	0.46	0.15	0.52	0.29	-0.02	0.19	0.06	0.04	0.05	-0.22	-0.15	-0.30
14	-0.32	0.11	-0.03	-0.19	-0.06	0.04	0.07	-0.29	-0.08	0.49	0.08	0.33	0.71	0.24	-0.07	-0.46	0.01	-0.10	-0.30	-0.08	-0.08
15	-0.33	0.08	-0.06	-0.03	0.02	-0.07	-0.19	-0.38	0.15	0.31	0.02	0.29	0.76	0.70	-0.36	-0.35	-0.05	0.04	-0.18	-0.40	0.00
16	-0.47	-0.01	-0.28	-0.06	-0.13	0.02	-0.07	-0.07	-0.05	0.60	0.03	0.47	0.55	0.10	0.25	-0.41	0.04	-0.29	-0.15	0.04	-0.13
17	-0.41	-0.04	-0.01	0.05	0.01	0.05	-0.09	-0.28	-0.04	0.40	0.18	0.18	0.73	0.29	-0.02	-0.39	-0.02	-0.00	-0.29	-0.07	-0.16
18	-0.53	0.34	0.11	-0.18	-0.10	0.11	0.29	-0.13	-0.08	0.32	-0.08	-0.06	0.68	-0.20	-0.17	-0.35	0.18	0.10	-0.24	-0.01	-0.01
19	-0.40	0.25	-0.03	-0.23	-0.27	-0.36	0.21	-0.48	-0.51	0.34	0.12	0.98	0.66	-0.12	-0.64	-0.55	0.22	0.06	-0.04	0.27	0.50
20	-0.35	0.54	-0.18	-0.22	-0.38	0.14	0.29	-0.76	0.08	0.34	-0.44	1.15	0.64	0.51	-1.08	-0.37	0.24	-0.13	-0.34	0.28	0.02
21	-0.42	0.06	-0.03	-0.11	-0.12	-0.01	0.01	-0.20	0.00	0.42	0.03	0.13	0.76	0.23	-0.10	-0.47	0.09	-0.03	-0.19	-0.08	0.05
22	-0.25	0.80	0.29	0.02	-0.85	-0.13	0.25	-1.08	-0.62	-0.02	0.39	0.34	0.24	-0.15	-0.33	-0.25	0.81	0.29	0.01	0.08	0.16
23	-0.16	0.89	0.16	-0.03	-0.12	-0.13	0.33	-0.95	0.27	0.41	-0.37	0.71	0.08	0.11	-0.04	-0.39	0.70	-0.53	-0.24	-0.25	-0.43
24	-0.17	0.00	-0.08	-0.18	-0.05	-0.07	-0.01	-0.21	-0.02	0.57	0.19	0.30	0.60	0.29	0.25	-0.45	-0.14	-0.21	-0.36	-0.09	-0.18
25	-0.38	0.07	-0.03	-0.34	-0.03	0.04	0.27	-0.32	-0.11	0.42	-0.01	0.21	0.55	0.06	-0.04	-0.19	0.21	-0.10	-0.33	0.02	0.02
26	-0.40	0.05	0.00	-0.12	-0.05	-0.01	0.16	-0.09	-0.04	0.42	0.01	0.06	0.72	-0.01	-0.03	-0.41	0.03	0.00	-0.37	0.05	0.00
27	-0.38	-0.03	-0.05	-0.26	-0.07	-0.01	0.06	-0.07	0.03	0.49	0.05	0.08	0.72	0.14	0.05	-0.43	-0.06	-0.04	-0.19	0.04	-0.06
28	-0.35	0.18	-0.10	-0.43	0.10	-0.08	0.31	-0.26	0.13	0.47	-0.22	0.16	0.48	-0.02	0.01	-0.35	0.18	-0.10	-0.14	0.04	-0.04
29	-0.49	0.19	-0.14	0.21	-0.07	0.26	-0.01	-0.06	-0.08	0.39	-0.15	0.27	0.48	-0.14	0.09	-0.47	0.20	-0.25	-0.11	0.02	-0.16
30	-0.43	0.46	-0.02	-0.17	0.15	-0.03	0.46	-0.63	0.00	0.37	-0.33	0.15	0.39	-0.32	-0.02	-0.43	0.46	-0.02	-0.19	0.21	-0.07

31	-0.33	0.01	-0.12	-0.32	-0.09	0.01	0.11	-0.29	0.10	0.51	-0.05	0.28	0.68	0.40	-0.08	-0.33	0.11	-0.18	-0.33	-0.08	-0.01
32	-0.42	0.03	-0.19	-0.21	-0.03	0.02	0.17	-0.45	-0.28	0.48	0.10	0.42	0.68	0.07	-0.08	-0.39	0.18	0.08	-0.32	0.11	0.02
33	-0.25	0.17	-0.01	-0.33	0.16	0.01	0.44	-0.24	-0.01	0.38	-0.20	0.02	0.45	-0.24	-0.00	-0.38	0.20	-0.00	-0.32	0.16	-0.00
34	-0.33	0.13	0.20	-0.13	-0.12	-0.03	0.01	-0.28	-0.12	0.29	0.18	0.09	0.82	0.25	-0.33	-0.48	-0.08	0.13	-0.17	-0.07	0.05
35	-0.36	0.19	0.15	-0.18	-0.04	-0.07	0.32	-0.21	-0.06	0.21	-0.16	0.27	0.73	-0.03	-0.67	-0.40	0.18	0.18	-0.31	0.08	0.20
36	-0.33	-0.08	0.05	-0.12	-0.06	0.01	-0.01	-0.03	-0.04	0.33	0.12	0.02	0.82	0.26	-0.21	-0.42	-0.12	0.09	-0.28	-0.09	0.09

Factor summary

	PC1	PC2	PC3
Eigenvalue	0.12	0.03	0.03
% variance explained	32.6%	15.5%	14.8%

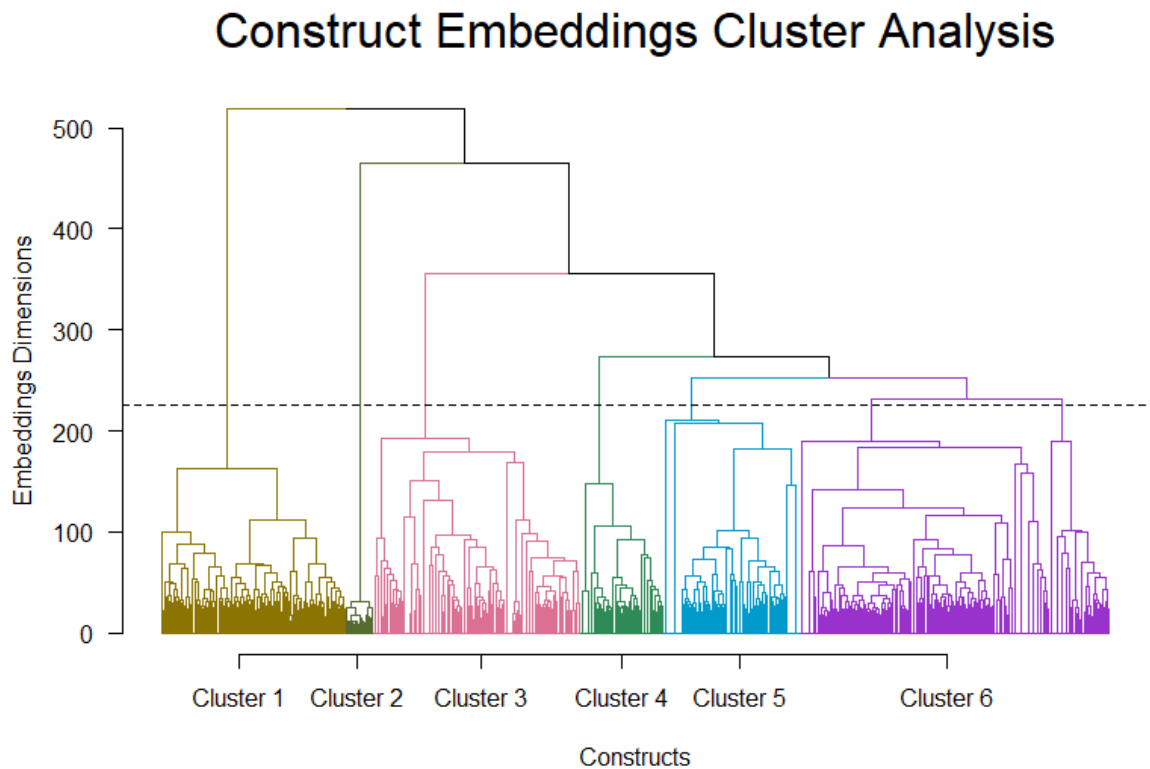
*P=Participant number

Note. PC = Principal Component. Factor loadings identified as defining principal components are indicated by bold typeface, indicating a loading of +/-0.50 or above. Factor loadings are rounded to two decimal places for reporting purposes, meaning that some loadings reported as 0.5 are not highlighted as defining.

6.3.2 Cluster analysis

Constructs from 36 completed grids were subjected to hierarchical cluster analysis using 512-dimension word embeddings to numerically represent semantic content, via the Hclust command within RStudio 1.0.143 (RStudio Team, 2016). Visual inspection of the resulting dendrogram (figure 6.2) suggested six clusters be retained (see appendix C), but further exploration of these clusters found that these were too few to provide coherence. However, reducing the data to seven or more clusters resulted in some consisting solely of single duplicated words. According to the criteria described by Everitt and colleagues (2001) and Han and colleagues (2011), ‘good’ clusters provide sufficient isolation and cohesion to be helpful in interpreting data, they offer clusters characterised by non-random content signified by sufficient homogeneity to enable meaningful interpretation. The results failed to meet these criteria.

Figure 6.2. *Dendrogram of cluster semantic content.*



6.4 Discussion

6.4.1 Principal component 1: violation of baseline expectations.

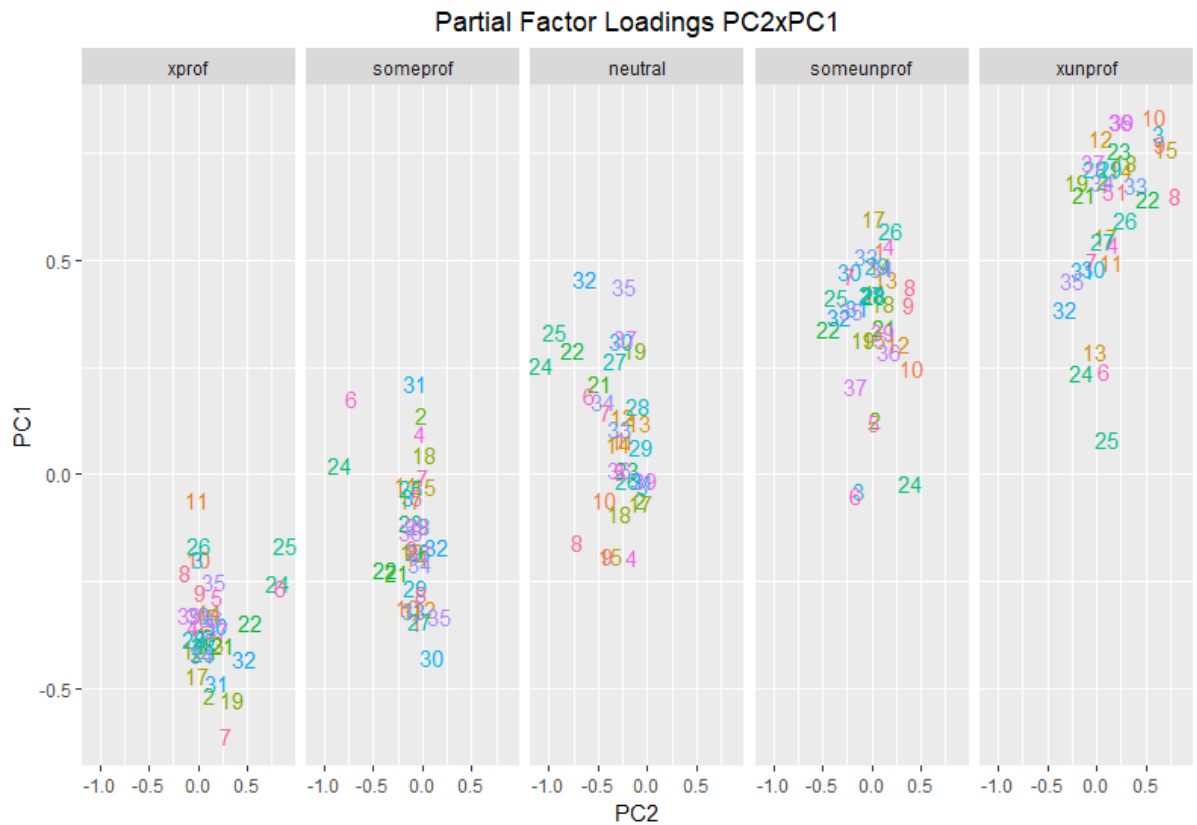
The content of constructs and comments loading at defining levels suggested that principal component one (PC1) related to the extent to which individuals violate the minimum standards of performance expected of all colleagues to meet the fundamental requirements of professionalism. Individuals deemed high on PC1 were described in such terms as being miserable or short-tempered, selfish and lacking team spirit, untrustworthy, threatening or bullying, impatient and unhelpful, unpleasant and discourteous, irresponsible, arrogant, ignorant, and generally behaving in ways inappropriate to the workplace. Individuals scoring low on PC1 were described as being efficient, organised, mature, dedicated, honest, supportive, non-manipulative, and flexible. The pattern of factor loadings suggested that all participants agreed on the direction of PC1, with low scores for *extremely professional* colleagues steadily increasing to high scores for *extremely unprofessional* colleagues, as shown in figure 6.3. The demographics of participants whose elements defined PC1 generally reflected those within the total participant sample.

The constructs generated in relation to PC1 were unsurprising, with those associated with a low score (professionalism) being frequently emulated within the literature (e.g. Anderson et al., 2014; Ben-David et al., 2004; Bonke, 2006; Carey & Ness, 2001; Carter et al., 2015; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Swick, 2000). However, it was somewhat surprising to observe them all clustered within a single component. Research suggests that individuals describe their understanding of or views towards professionalism in terms of distinct themes (Baernstein et al., 2009). However, this new data gathered taking a new approach grounded in phenomenological experience, suggests that although these themes may be reported by individuals as separate, their use in judging professionalism within everyday life may not be. The apparent disparity between what people say about professionalism and how they actually judge it may be understood using theories of action and values (Argyris & Schön, 1974). Argyris & Schön (1974) distinguish between the outward accounts individuals provide of their

values and actions (their 'espoused theories') and the actual values that govern their more fundamental decisions regarding how to behave or what judgement to make (their 'theories-in-use'). The present data tentatively suggests that the thoughts people articulate regarding professionalism and the actual judgements they make, may be similarly inequivalent. In order to understand ways to increase perceived professionalism, the contribution of theories-in-use ought not to be underestimated and warrants further study.

It was noted that the loadings defining PC1 were overwhelmingly drawn from the *somewhat unprofessional* and *extremely unprofessional* elements, with very few from the elements indicating higher professionalism. This suggests that PC1 may relate to how unprofessional colleagues are, rather than how professional they are. Put another way, scoring highly on PC1 is likely to result in a colleague being perceived as unprofessional by the participants of this study, but a low score is insufficient to ensure they are perceived as actively professional. This finding appears congruent with previous claims that professionalism may be more easily recognised by its absence than presence (Hammer, 2006; Wilkinson et al., 2009). Given that this component accounted for more than twice the variance of the other two, we should consider whether professionalism is primarily a positive attribute characterised by its presence, or alternatively the negative space left by an absence of unprofessionalism. The data suggests that according to the shared construing of this participant group, institutions may be unable to successfully select or assess candidates based on the presence of positive professionalism at all, and should therefore attend more closely to the absence of unprofessional characteristics in order to increase the perceived professionalism of its students or employees.

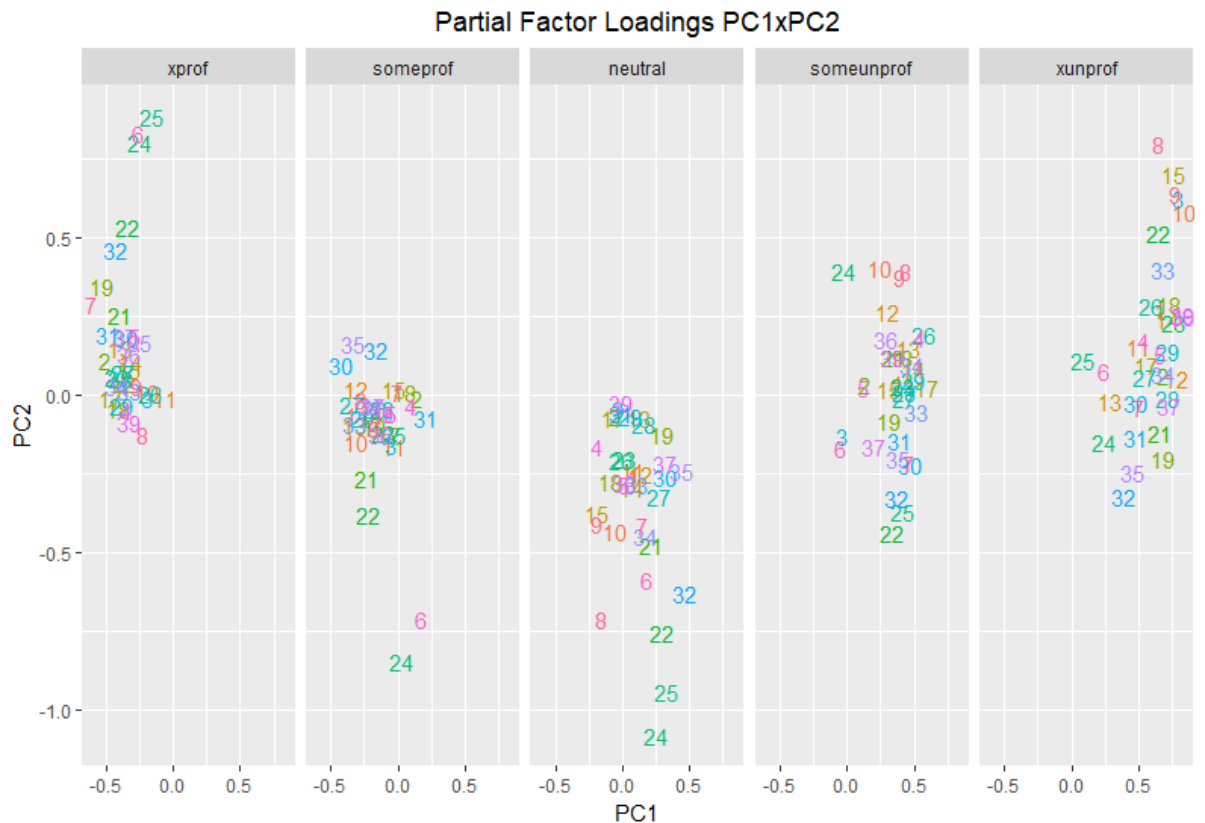
Figure 6.3. Biplot of PC1 (violation of baseline expectations) and PC2 (logical ruthlessness) partial factor loadings.



6.4.2 Principal component 2: logical ruthlessness.

The content of constructs and comments defining principal component two (PC2) related to successful colleagues, but whose success came at an interpersonal cost. Comments defining PC2 included themes such as efficiency and focus, being confident and capable, and getting the job done, but at the same time being rude, mean and irritable, ruthless, unfriendly, assertive, unemotional, and even aggressive. The pattern of factor loadings was inconsistent, with high scores being associated with either *extremely professional* or *extremely unprofessional* colleagues. Low scores were generally associated with colleagues who were more neutral in terms of professionalism, as shown in figure 6.4. This suggests that people exhibiting a high level of PC2 would split the participants of this study into extreme views whereby they are deemed both *extremely professional* and *extremely unprofessional* at the same time by different individuals. People exhibiting low levels of this component will not provoke opposing reactions, however, instead being viewed more consistently as neutral or middling in relation to professionalism.

Figure 6.4. *Biplot of PC2 (logical ruthlessness) and PC1 (violation of baseline expectations) partial factor loadings.*



The perception of constructs associated with PC2 was unexpected.

Contemporary understanding of professionalism within both professional and academic literature requires colleagues to be effective in their role while maintaining positive character (Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Swick, 2000). PC2 did require professionals to be effective and competent, but crucially combined this with conduct actively contradicting the concept of good character. This was unexpected as goodness of character is the single strongest theme within the academic professionalism literature (e.g. Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Swick, 2000). Behaviours congruent with this aspect of PC2 have generally been used previously to describe lapses in, or attributes opposing, professionalism (e.g. Adam et al., 2015; Arnold, 2002; Gillespie, Paik, Ark, Zabar & Kalet, 2009; Hendelman & Byszewski, 2014; Kulac, Sezik, Asci & Doguc, 2013; Papadakis et al., 2005; Stewart, Wyatt & Conway, 2011; Taylor & Grey, 2015). Consequently, it was interesting to find that

some of the study participants perceived individuals scoring highly on PC2 to be *extremely professional*. This suggests that current understanding of professionalism is at odds with the judgements made by participants in appraising the conduct of others. Further research would be required to explore whether PC2 is also observed within wider populations, but within the present sample, some insight into those perceiving PC2 as professional versus unprofessional was provided by the demographic data gathered.

Further scrutiny of the demographic characteristics of participants with elements defining PC2 revealed that all those participants who felt it was an indication of professionalism were male and all those participants who felt it was an indication of unprofessionalism were female. This suggests that male participants found the ruthless qualities associated with PC2 to be not only acceptable, but actively professional. This is particularly interesting because while the total sample gender balance was almost three quarters female (72%), the elements defining PC2 represented an equal gender split (F=11, M=11). This suggests that PC2 may be subject to an additional gender effect in that the men within our sample were more likely to use it as a basis for professionalism decisions than the women. Our data suggest that not only is there a gender effect in predicting performance in, and understanding of, professionalism (Karnieli-Miller et al., 2010; Papadakis et al., 2004), men and women also perceive professionalism in opposing ways within the present sample. Perhaps then gender differences in the performance of professionalism do not relate to their possessing differing levels of professionalism as an ability or attribute, but rather that men and women may be actively aiming to deliver opposing professional characters, according to their view of what professionalism ought to look like, characters that may then be perceived in opposing ways by male versus female assessors. The implications of findings relating to PC2 and gender are unclear at this point and out of the scope of this thesis. However, given ongoing issues regarding gender inequality in the workplace within contemporary society, the role of gender in the perception of professionalism may provide fruitful avenues for further study.

6.4.3 Principal component 3: personal likeability.

The content of constructs and comments defining principal component three (PC3) related to how interpersonally likeable individuals were. Comments defining PC3 included themes such as being warm and friendly, welcoming and helpful, talkative, flexible, and interested in the welfare of others. Likeability was high where colleagues undertook their role unselfishly and helpfully, while demonstrating warmth through talkative natures, with individuals demonstrating quieter or more introverted interpersonal approaches being more often perceived negatively.

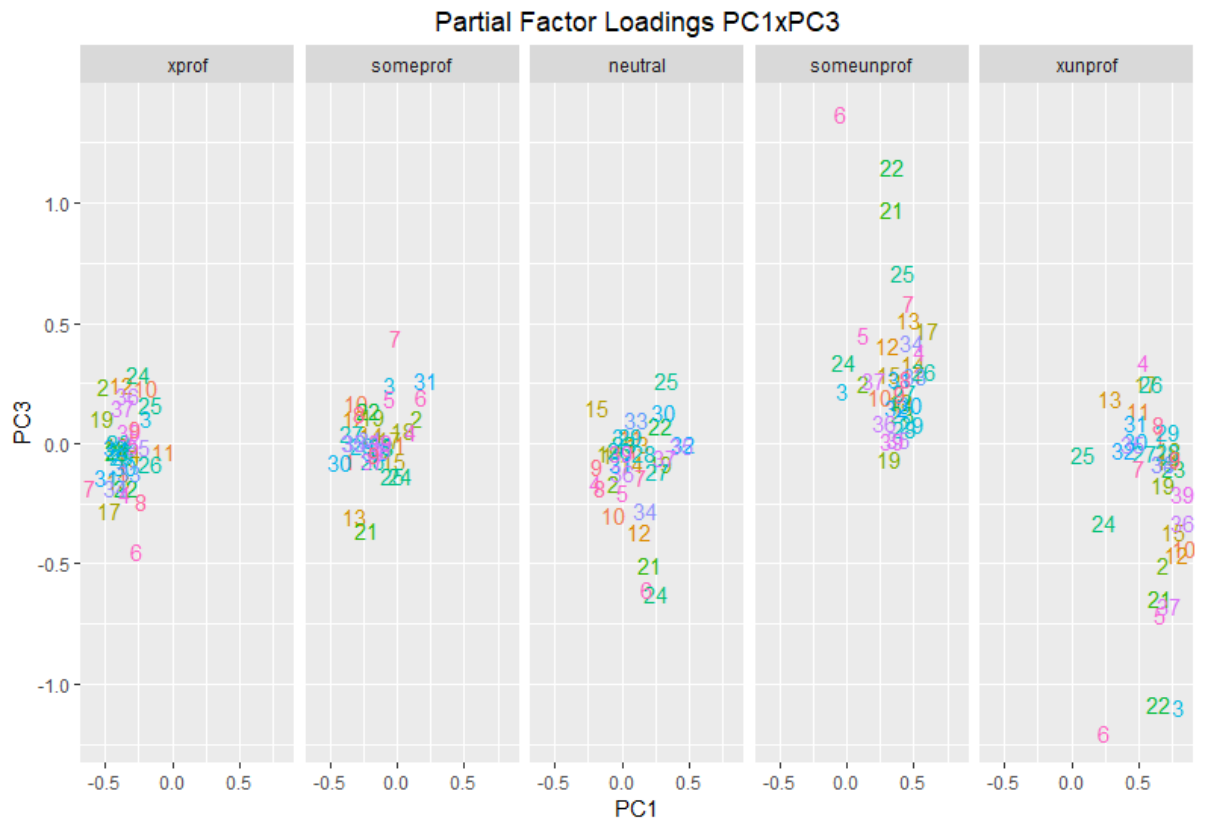
Another theme within the grids defining this component was that individuals were not deemed capable or effective in their professional roles, being perceived as struggling, nervous and insecure, limited, and needing help and support. These characteristics affected all elements, regardless of whether they were perceived as likeable or not. The pattern of defining factor loadings was clustered in the *neutral*, *somewhat unprofessional*, and *extremely unprofessional* elements. High scores for personal likeability were generally perceived as *somewhat unprofessional* amongst the present participant sample, and low scores as middling at best and extremely unprofessional at worst, as shown in figure 6.5. Once again, the gender balance of elements defining PC3 was more equal than in the total sample (F=9, M=8) suggesting that a disproportionately high number of men felt that PC3 was important when considering professionalism. Once more, the role of gender in perceptions of professionalism is supported here, suggesting again that additional study is required to explore this further.

The content of elements defining PC3 suggest that its relationship to professionalism is complex. Individuals described as highly likeable generally came from the *neutral* and *extremely unprofessional* elements. Individuals perceived as more professional (those from the *somewhat professional* and *extremely professional* categories) were generally judged as neutral regarding personal likeability, with loadings clustering more clearly around zero. This suggests that PC3 distinguishes professional and unprofessional colleagues by its relevance, rather than content. Defining comments associated with PC3 suggest that this component may only be relevant when judging those who are unprofessional, because those who are professional demonstrate impersonal and objective work approaches that result in

more effectively maintained professional boundaries. As such, personal relationships may be rendered of little relevance where individuals deliver professionalism. However, as the level of professionalism reduces due to lower perceived competence or effectiveness, how much one individual likes another may become more pertinent to their judgements.

Although the role of likeability in professionalism has not received much discussion, likeability bias has been demonstrated in person evaluations more generally. For example, evidence suggests that likeability correlates positively with peer assessment of performance (Aryadoust, 2017; Carmona, Iyer & Reckers, 2014; Sonnentag, 1998), meaning that likeable individuals will be rated as performing more highly, regardless of their actual performance. This is incongruous with the findings of the present study, which suggest that likeability is only relevant to individuals performing poorly in terms of professionalism. However, it is worth noting that during interviews, pilot participants repeatedly expressed a desire to separate judgements of professionalism from how much they liked individuals. Consequently, this component may have been subject to some level of active suppression by participants that could have disproportionately affected judgements regarding unprofessional individuals. This hypothesis is merely conjecture, however, and therefore requires further investigation.

Figure 6.5. *Biplot of PC3 (personal likeability) and PC1 (violation of baseline expectations) partial factor loadings.*



6.4.4 Cluster analysis: proof of concept

The cluster analysis approach used within this study was exploratory in nature, targeting potential for use rather than direct results. It is therefore important to note that it was applied according to maximum convenience rather than as a fully appraised analytical technique. The value of this was not in drawing firm conclusions regarding the semantic content of professionalism-related constructs, but rather in providing evidence of proof of concept for using word embeddings as a means of quantifying RGT data for cluster analysis. Within this study, the approach proved unhelpful, but the results did offer insight into potential reasons for this.

Decisions regarding the specifics of cluster analysis are often largely based on practical considerations rather than theoretical or statistical rationale, with the overriding aim being that results are useful or meaningful to the question under study (Han et al., 2011). The decisions made in applying cluster analysis within this study

were taken accordingly. Hierarchical clustering using the agglomerative Ward's method of linkage was used due to its popularity within the UK, its ready availability within common statistical software packages, and the easily accessible visualisation of results offered by the dendrogram (Everitt et al., 2001). However, making different decisions might be helpful in exploring the potential of word embeddings further. For example, the dataset used for this analysis was extremely large. A total of 844 construct poles were generated across 36 repertory grids. As the poles were used individually, 512 numerical vectors were associated with each of the 844 constructs resulting in 432,128 datapoints to be analysed. However, most clustering approaches are optimally effective with datapoints numbering in the hundreds, unless they are specifically scalable (Han et al., 2011). Hierarchical clustering is known to encounter difficulties when scaled to large datasets due to the high number of agglomerative or divisive decisions to be made (Han et al., 2011). This suggests that different results may be achieved by using a different, non-hierarchical clustering approach. An alternative option to address this issue might be to describe the dataset using fewer word embedding dimensions. It may be illuminating to reanalyse this dataset using a different, smaller set of dimensions to explore its impact on results. Applying 512 dimensions to a smaller dataset may also provide insight into the impact of the number of datapoints on results.

The method of linkage used within this study is known to be susceptible to 'noisy' data, such as that containing a high number of errors or outliers (Han et al., 2011). Within the data of this study, different participants self-generated constructs according to their own subjective perception of the descriptors chosen to identify them. The semantic content of the same words may therefore have been different for different participants, although the word embeddings used to describe them would be identical. This introduces noise or error into the dataset that would impact the clusters identified using Ward's method of linkage. In order to remedy this, future research might consider using repertory grids with supplied constructs, in order to minimise error resulting from participant-generated constructs.

Ward's method, and hierarchical clustering in general, is optimally effective for identifying equally sized clusters that are spherical when plotted graphically (Everitt et al., 2001; Han et al., 2011). The results of principal component analysis reported within this study suggest that professionalism does not consist of equally

sized components, with the first accounting for significantly more variance than the other two. Although principal components and clusters are statistically and conceptually different, it may be reasonable to expect that the clusters likely to be most helpful in understanding the construct of professionalism may not be of equal size, in light of these findings. Therefore, hierarchical clustering may not have provided the most helpful results. In such cases, it is recommended to use density-based methods of clustering instead, as these are able to detect clusters of varying size and shape (Han et al., 2011).

Overall, although the results of hierarchical clustering were unhelpful within this study, its potential may still be latent. Manipulating datasets in terms of constructs used or number of dimensions attached to them and considering alternative methods of clustering and linkage may still offer potential for future research.

6.4.5 Study evaluation.

The findings of the present study should be understood in relation to the limitations of both the methodology used and its application. The RGT provides insight into the construing of the participant sample only. The sample within the present study was over two thirds female (F=72%), reflected a majority of individuals relatively early within their careers (early career=58%), and educated to a higher level than that within the general UK population (educated to postgraduate level=61% versus 27% educated to undergraduate level in the general population; Office for National Statistics, 2011). This means that extrapolations made from this sample about the construing of wider populations or any other individuals is inappropriate (Easterby-Smith et al., 1996). Despite this, the findings may be used as a starting point from which to generate hypotheses regarding the construing of others that may then be tested using larger scale studies designed to provide generalisable findings.

Within this study, the researchers chose to subject data to quantitative analysis to uncover areas of shared construing (Brewerton & Millward, 2001; Catania & Randall, 2015; Cooper, 2010). This was a departure from the original

intended use of repertory grids in practice, a departure deemed unjustified by some due to the separation of data from its original purpose of understanding individual systems of construing (Brewerton & Millward, 2001; Easterby-Smith et al., 1996). RGT data aims to uncover individual differences in construing and so aggregating data across individuals may distort results (Easterby-Smith et al., 1996). However, quantitative analyses of RGT data are accepted within more contemporary literature (Catania & Randall, 2015) and are suggested to enable more precise expressions of the qualitative data contained by multiple grids (Jankowicz, 2004).

Using standardised elements was necessary within the present study to provide comparability of construing across multiple individuals and allow the use of MGFA quantitative analyses. This is an accepted approach in using the RGT for research purposes (Fransella et al., 2004; Jankowicz, 2004), as it enables the quantitative analysis of multiple grids and ensures that content relevant to the aims of the study is accessed (Jankowicz, 2004). However, this carries risks of reinforcing ‘correct’ answers amongst participants, as it constitutes increased researcher influence over the elicitation process (Brewerton & Millward, 2001; Catania & Randall, 2015; Fransella et al., 2004; Jankowicz, 2004; Tan & Hunter, 2002). To moderate this issue without losing sight of the quantitative aims of the study, the research team offered general element categories providing participants the freedom to attach self-generated, more specific elements. Although not entirely true to the original qualitative purposes of the RGT, this enabled maximum integrity to the research method without losing sight of the pragmatic study aim of exploring shared construing quantitatively.

This study also used a majority of online data collection. Although the RGT process is easily achieved within this medium, there may be some impact on the quality of data. Specifically, electronically completed repertory grids prevent researchers from clarifying participant construing. Consequently, constructs must be interpreted at face value, potentially resulting in the loss of meaningful data. In addition, where individuals have given little thought to a topic prior to completing a grid, as was remarked by a number of participants of the present study, they may be more likely to spontaneously generate constructs reflecting a perceived ‘correct’ answer. As such, where the RGT is used to target tacit construing as in the present

study, interview-based data collection may have elicited different results (Jankowicz, 2004).

As with all factor analysis, the onus in MGFA is on researchers to interpret the content of principal components extracted. The principal components described within this chapter therefore reflect the results of analysis as viewed through the lens of the research team's construals. Consequently, they may reflect the research team's construing of other people's construals of professionalism, rather than the latter themselves (Butler, 2009a). Also, in pursuing a solely factor analytical approach, the present study may have lost potentially meaningful data in the form of construct semantic content (Easterby-Smith et al., 1996). The data elicited comprised a large set of constructs within which there was a significant number of duplicates. Constructs appearing with higher frequency were interpreted more centrally in component descriptions. However, without further analysis, it is unclear whether different participants intended to convey identical semantic content when using the same terminology.

The RGT relies heavily upon the assumption that social aspects of individual differences become encoded implicitly in the language used to describe them, a view known as the lexical hypothesis (John et al., 1988; Mollaret, 2009; Poropat & Corr, 2015). However, a key criticism of the lexical hypothesis is that even where commonplace wording is used, there is ambiguity rendering it open to (mis)interpretation (Ashton & Lee, 2005; Jankowicz, 2004; John et al., 1988; Mollaret, 2009; Poropat & Corr, 2015). As a result, the component descriptions provided may be incomplete, with gaps essentially filled by the research team's construals of the semantic content of the participant sample's language (Butler, 2009a). This may be particularly problematic for the present study as it targets implicit or tacit knowledge about professionalism. PCT and the RGT methodology specifically warn against assuming equivalence between the subjective content of constructs and their objective meaning in natural language, particularly where constructs are implicit (Fransella et al., 2004). This issue may particularly affect the data of the present study as it attempts to access theories-in-use that may not yet be fully explored by individuals. Once again, further research is needed to explore the complexity of this problem and extricate the objective data from issues of researcher reflexivity.

Finally, within the present study, none of the extracted principal components demonstrated an eigenvalue exceeding one. The extraction only of components with an eigenvalue exceeding one is known as the Kaiser-Guttman criterion, which demonstrates that components are stable and account for sufficient variance to reflect at least a single whole variable (Auerswald & Moshagen, 2019). Eigenvalues of the magnitude demonstrated within the present study indicate that extracted principal components may be unstable, thus obscuring the correct factor structure (Auerswald & Moshagen, 2019; Girden & Kabacoff, 2011; Igundunasse, 2016). However, the Kaiser-Guttman criterion is effective in studies with large participant sample sizes (>250) and so may not actually be relevant to the present study. In addition, the Kaiser-Guttman criterion receives criticism for being inaccurate and resulting in arbitrary factor structures that lack meaningful coherence (Costello & Osborne, 2005; Furr, 2011; Furr & Bacharach, 2014; Girden & Kabacoff, 2011). Moreover, the RGT is designed to target individual systems of construing that, according to PCT, are organic and evolving and change as individuals gather data about their environment (Easterby-Smith et al., 1996). Therefore, it may actually have been anticipated that the extracted principal components would be unstable. Conversely, pre-planned, espoused theories of professionalism might be expected to be more stable and thus deliver higher eigenvalues. Crucially, the instability of the components extracted may therefore actually suggest that the present study methods have successfully targeted construing and by-passed espoused theories of professionalism as intended.

6.4.6 Future Directions.

The findings reported here provide a starting point from which to further explore professionalism as a subjectively construed construct. The principal components extracted suggest areas where construing is shared by participants, but the picture painted by these components is complex and requires further elaboration. Although our sample appears to share construing around what behaviours violate the fundamental core of professionalism (PC1), they do not always agree on what professionalism looks like versus unprofessionalism (PC2) and suggest that it may not be a discrete concept at all, instead being inexorably linked to more interpersonal

aspects of the social world (PC3). These findings suggest multiple potential future lines of enquiry.

PC1 suggests that theories-in-use used to judge professionalism may be significantly different to those espoused. Future research should consider methodological approaches that by-pass espoused theories of professionalism, instead targeting more implicit, experiential aspects of theories-in-use (Catania & Randall, 2015). PC2 suggests that although everyone may agree on the significance of ruthlessness to success, this characteristic may polarise opinion in terms of perceived professionalism. There is therefore merit in going beyond studying similarities in construing and exploring differences in perceived professionalism also. Particularly, it would be worthwhile exploring the role of gender in perceived professionalism. Although gender issues in professionalism have received attention relating to individuals *producing* professionalism through their conduct (Karnieli-Miller et al., 2010; Papadakis et al., 2004), it has not yet been studied from the perspective of how that professionalism is *received*. Finally, PC3 suggests that personal likeability is relevant to perceptions of professionalism but that its role is unclear. This warrants further research to establish whether PC3 reveals an aspect of perceived professionalism or whether it is uncovering a separate but related issue.

Future research should also seek to clarify the content of components of shared construing in terms of semantic content, and explore their stability with other participant populations, before generalisable conclusions may be drawn. In addition, in response to claims that the use of statistical methods that may have distorted results by aggregating idiosyncratic data (Easterby-Smith, 1996), attempts should be made to replicate the present findings using methodological approaches overcoming this issue, such as Q methodology (Block, 2008; Stephenson, 1935; van Exel & De Graaf, 2005; Watts & Stenner, 2007).

A final direction for future research relates to the proposed approach to cluster analysis. Although the dataset within this study may have been too large to provide informative clusters when constructs were described using 512 dimensions of word embeddings, this technique may still have utility with other RGT datasets and/or embeddings sets. Further methodological testing would be required to explore this.

6.5 Chapter Summary and Conclusions

This chapter describes an empirical study of professionalism as a subjectively construed concept using the RGT. Although the findings described have perhaps raised more questions than they have answered, they are helpful in beginning to understand the nature of professionalism as a perceived rather than produced concept. The data provide multiple avenues through which this understanding may be further progressed and enhanced. Although the findings of this study do not enable generalised conclusions, the data suggest that current understanding of professionalism may be at odds with real-world judgements made by individuals in daily life. As such, a focus on theories-in-use associated with professionalism is recommended. The data also suggests that perceived professionalism may be a negative characteristic marked by the absence of unprofessionalism, with further study recommended to focus on the potential implications of this for measuring professionalism. Findings further suggest a need to explore the role of gender in perceived professionalism and how it relates to issues of ruthlessness and likeability. The findings of this study recommend that further attempts to explore perceived professionalism may benefit from paying particular attention to researcher reflexivity when interpreting findings to ensure transparency of conclusions.

Chapter 7: Shared Subjectivity in Understanding Professionalism: A Q Methodological Study

Note to reader: this chapter is written to maximise the accessibility of Q methodology for a general audience, as recommended by Watts and Stenner (2012), including using long-form narrative factor descriptions. The findings of this study are of relevance to a broad and multidisciplinary audience and so its reporting targets non-specialist users. Taking into account claims of pervasive and widespread misunderstanding of the theoretical underpinnings and applications of Q methodology (Watts & Stenner, 2012), this chapter provides detailed elaboration on its use in this study for the benefit of all readers. The chapter is therefore longer than might be expected relative to its scientific contribution.

Abstract

Professionalism is a major focus for employers today, resulting in pressure on educators to assure the professionalism of graduates. Efforts to provide such assurances are, however, hindered by a lack of consensus as to how professionalism should be defined. This study sought to understand professionalism from the subjective viewpoint of individuals receiving or observing it using Q methodology, a method particularly suited to exploring such tacit person-evaluation judgements. The aim was to explore shared perspectives regarding professionalism amongst a group of adults to provide new insight into how it is viewed today. Thirty-four participants were asked to sort 42 statements relating to the topic of professionalism according to the extent to which they agreed or disagreed with them. The statements were derived from an academic literature review and media search, and the content of a previous, interview-based study. By-person factor analysis was used to identify shared viewpoints, resulting in a four-factor solution accounting for 42% of the variance observed. The four extracted factors were interpreted holistically under the headings of *intellectual professionalism*, *collaborative professionalism*, *personal professionalism*, and *professional professionalism*. Findings suggest divergent views regarding the nature of professionalism, meaning that attempts to define it in norm-referenced ways may be of limited relevance for practical applications. The data also suggest removing the traditional occupational boundaries of professionalism to view it as a more general concept underlying role-specific behaviour. Recommendations for future research include focusing on professionalism as a subjectively perceived construct, rather than one that may be objectively defined.

7.1 Introduction

Professionalism is a major focus for employers and educators today (Altirkawi, 2014; Arndt et al., 2019; Arnold, 2002; Baernstein et al., 2009; Black et al., 2019; Blake & Gutierrez, 2011; Buck et al., 2015; Carter et al., 2015; Chisholm et al., 2006; Evans, 2008; Evetts, 2014; de Mendonça et al., 2016; Goldie, 2013; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004). High profile cases documenting lapses in professionalism have led stakeholders to question professional services and their staff (de Mendonça et al., 2016; Hammer, 2006; Lynch et al., 2004; Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013; Monrouxe & Rees, 2012), with indirect pressure reaching training and education institutions in their efforts to assure higher standards of professionalism amongst their graduates (Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013).

In order to make these assurances, high quality and valid measurement of professionalism is necessary, but this is lacking (Goldie, 2013; van Mook et al., 2009). Recent reviews of attempts to measure professionalism have raised concerns regarding the validity of extant measures (de Mendonça et al., 2016; Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; Veloski et al., 2005; see also the findings of chapter 5), with the need for a theoretically grounded and empirically testable construct of professionalism reportedly impeding progress (Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; see also the findings of chapter 5). In the absence of a theoretically grounded construct of professionalism, previous attempts to define it have used a top-down approach that involved asking sector experts to describe their thoughts regarding professionalism, as a product of both their personal expertise and experience (Birden et al., 2014). Such descriptions are widely available but vary extensively, leading to significant heterogeneity. Rather than attempt to reconcile these divergent views empirically, general research practice has relied upon implied rather than explicitly stated definitions of professionalism according to the author's personal viewpoint, as a means of avoiding difficulties in finding consensus around a fully elaborated definition (see the findings of chapter 5).

Although a historical shift beginning in the 1960's moved understanding away from professionalism as something regulated internally by professions and towards recognising the importance of how it is perceived by stakeholders, the

boundaries between sector-specific definitions of professionalism remain in place (e.g. Carter et al., 2015; Evans, 2008; Evetts, 2003; Finn et al., 2010; General Dental Council, 2013; General Medical Council, 2013, 2016). Approaching professionalism from the perspective of stakeholders has, however, provided opportunities for individuals to describe in detail the espoused theories held regarding professionalism.

Espoused theories are the pre-planned accounts of behaviour or values given by individuals to others (Argyris & Schön, 1974). Such theories commonly constitute the most socially acceptable or coherent description available of the motivations for human behaviour, reflecting the individual's prior learning and experience but not necessarily the true reasons for their actions (Argyris & Schön, 1974). For example, during cognitive processing, espoused theories relating to an individual's own behaviour are unconsciously transformed to maximise congruence with external criteria, such as their employer's requirements. As a result, espoused theories for one's own behaviour may be altered to meet company policy, rather than reflect their true aetiology. This enables individuals to view themselves as positive and productive members of the workforce, as they deliver on the expectations of their supervisors and peer colleagues (Hinojosa, Gardner, Walker, Cogliser & Gullifor, 2017). This does not mean that espoused theories constitute attempts to mislead other people, they are simply a by-product of cognitive processing; the true reasons for our actions are often beyond cognitive awareness, unless a special effort is made to access them (Argyris & Schön, 1974). Consequently, theories elicited by asking stakeholders to describe their accounts of professionalism in an unstructured way may not fully reflect the ways in which they apply themselves to viewing it in the real world. Empirical findings that support this argument are described in chapter 6.

In contrast to espoused theories, theories-in-use represent less-consciously accessible reasoning that forms the true basis of behaviour (Argyris & Schön, 1974). As an example, consider an individual applying to take on additional duties within their work role. While the individual may describe their desire to contribute to new initiatives or increase performance within their team during a selection process, they may inform their closest colleagues that they are truly seeking a change of direction. Both of these accounts reflect espoused theories tailored to their audience, as the

individual's true motivations lie within a complex and less coherent interaction of factors including a desire to shake up their working life, ambitions to set themselves up for future promotion, wishing to work longer hours to avoid a difficult home-life situation, or simply a fundamental desire to please others and received additional recognition. These latter factors of which the individual may have limited or no awareness are their theories-in-use.

The factors underlying behaviour are dynamic and changing in ways that make it difficult for individuals to pinpoint and articulate them coherently, and so they generate espoused theories to help reduce the cognitive demands associated with attempting to negotiate their theories-in-use. Where personal motivations and viewpoints represent complex individual-society interactions, such as the person evaluations associated with perceiving professionalism, individuals may be particularly prone to delivering espoused theories in response to questioning (Block, 2008). The speed and somewhat automatic nature of person-judgements mean that they may be significantly different to those espoused following pre-planning and forethought (Block, 2008). As a result, where one seeks to improve the quality of professionalism as perceived by others, theories-in-use may offer more fruitful avenues for exploration. Q methodology is particularly advantageous in this regard. It offers a research approach targeting subjective knowledge using highly structured methods that are effective in supporting participants to articulate tacit viewpoints or reasoning, such as theories-in-use (Baker et al., 2010). As a result, Q methodology is particularly suited to the task of accessing theories-in-use regarding appraisals of professionalism.

The present study systematically explores theories-in-use by understanding professionalism through the subjective lens. It targets the ways that professionalism is viewed by stakeholders and explores aspects of those viewpoints shared by individuals. This study examines shared subjectivity in perceptions of professionalism amongst a group of adults from a range of occupational sectors and backgrounds using Q methodology and discusses the resultant data with respect to its implications for theoretical accounts of professionalism.

7.2 Methods

7.2.1 Materials.

7.2.1.1 *The Q sort.*

Q methodology is a mixed methods research methodology developed to provide a systematic and robust approach for exploring individual subjectivity (Baker et al., 2010; Brown, 1972, 1980, 1993, 1996; Mason et al., 2018; McKeown & Thomas, 2013; Stainton Rogers, 1995; Stenner & Stainton Rogers, 2004; Stephenson, 1935, 1953; van Exel & De Graaf, 2005; Watts & Stenner, 2012). Q methodology employs an adapted form of factor analysis that summarises the structure of the data where multiple participants are assessed across multiple variables, as fewer underlying variables or factors (Block, 2008; Brown, 1980; McKeown & Thomas, 2013; Watts & Stenner, 2012). Q methodology factor analysis is used where data represents multiple individual viewpoints rather than variables, correlating them to reveal fewer viewpoints that summarise the shared perspectives amongst participants (Stephenson, 1935, 1953; van Exel & De Graaf, 2005; Watts & Stenner, 2007). These shared perspectives are known as shared subjectivity (Baker et al., 2010). Traditional factor analysis correlates variables across multiple people to reveal the underlying variables that account for variance in the data, while Q factor analysis correlates multiple people across multiple viewpoints to reveal the shared subjectivity that accounts for common variance in the data. Traditional factor analysis seeks to describe variables objectively, but the crucial, unique characteristic of the Q approach is that it ensures that factors remain relevant to the individual viewpoints from which they are derived, safeguarding the subjectivity lost in traditional approaches seeking objectivity (Baker et al., 2010; Brown, 1980, 1993; Stephenson, 1935, 1977; van Exel & De Graaf, 2005; Watts & Stenner, 2012).

Q methodology is associated with a unique method for data collection known as the Q sort (Brown, 1980; Watts & Stenner, 2012). The Q sort provides a bespoke unit of measurement for exploring subjectivity and is the only research method to meet the assumptions and requirements of Q factor analysis (Brown, 1980; Watts & Stenner, 2012). The Q sort requires participants to undertake a ranking task to sort a series of stimuli according to their personal psychological significance (Brown,

1980; van Exel & De Graaf, 2005; Watts & Stenner, 2012). These stimuli most commonly take the form of a series of statements about a topic that are developed by the research team prior to data collection using a method deemed appropriate to the specific aims of the study and with reference to concourse theory (Brown, 1993; McKeown & Thomas, 2013; Stephenson, 1952, 1986a; van Exel & De Graaf, 2005; Watts & Stenner, 2012). Participants are asked to sort the stimuli according to a guiding question or statement known as the condition of instruction, such as how strongly they agree or disagree with each statement, or how like or unlike themselves they perceive it to be (Brown, 1993; McKeown & Thomas, 2013; van Exel & De Graaf, 2005; Watts & Stenner, 2012).

Once completed, participants' final ranking configurations constitute holistic representations of each individual's viewpoint regarding the subject matter, and where different individuals share similarity in their sorting patterns, this is interpreted as indicating similarity of viewpoint (Watts & Stenner, 2012). Following the sorting task, final sorting configurations are numbered to record the positioning of items and enable them to be analysed quantitatively using Q factor analysis (Watts & Stenner, 2012). Although subjected to quantitative analysis, Q research findings are not generalisable, instead relating only to the sample currently under study (Brown, 1980; Watts & Stenner, 2012). Q sort studies are able to deliver the rich and detailed data associated with qualitative research, while retaining the sound statistical basis more typical of quantitative enquiry (Brown, 1996; Watts & Stenner, 2007, 2012).

7.2.1.2 The current study.

The current study takes a Q methodological approach to exploring major viewpoints relating to professionalism. The research materials were developed prior to data collection by the research team and took the form of statements that may be made about the topic.

7.2.1.2.1 Identifying the concourse.

According to concourse theory, all topics are subject to a concourse (Brown, 1980, 1993; McKeown & Thomas, 2013; Stephenson, 1986a; van Exel & De Graaf, 2005; Watts & Stenner, 2012). The concourse may be best understood as an analogy with traditional research approaches that sample individuals representative of the total population of interest. As Q methodology samples viewpoints rather than people, the concourse represents the total population of viewpoints of interest, in this case, those pertaining to professionalism (Brown, 1980; Stephenson, 1986a). The concourse within the present study was all possible statements that may be made to express views regarding professionalism. The research team identified a large number of concourse statements relating to professionalism through academic literature and media searches, and a previous research study involving repertory grid technique interviews to explore how individuals construe professionalism (for full details, see chapter 6). These statements were used to develop a representative sample for use in the Q sort.

7.2.1.2.2 Developing the Q set.

The items to be ranked during the Q sort are known as the Q set (Brown, 1993; McKeown & Thomas, 2013; Stainton Rogers, 1995; van Exel & De Graaf, 2005; Watts & Stenner, 2012). The initial sample of concourse statements totalled 612, reduced to 434 by removing semantic duplicates. These 434 statements were subjected to thematic analysis undertaken by the lead researcher, which identified 19 themes (see appendix D). The themes and statements were reviewed and refined by the lead and two other researchers with expertise in psychological measurement to reduce their number while maximising content validity by ensuring they represented the full breadth of the concourse. This was achieved through an iterative process of deletion, merging, and re-wording, before the final Q set was reviewed by a researcher with expertise in professionalism to maximise clarity and accessibility.

The final Q set was developed with the overarching aim of ensuring balanced coverage of the topic and by questioning whether individual statements would be useful in revealing similarity or disparity of viewpoint amongst participants. The

criteria used required that each statement make a single point to avoid double-meaning; be free from occupation or profession-specific language, or technical jargon, to ensure relevance for a general population sample; and follow on from a prefix to enable ease of comprehension and prevent the inclusion of double negative statements. The prefix was “For a person to be professional, they must...”. Each iteration of the Q set was trial sorted by a member of the research team to ensure its size was manageable when making fine distinctions between statements. The condition of instruction was to sort the Q set according to level of agreement with each statement, ranging from most agree to most disagree.

Best practice guidance suggests that Q sets of between 10 and 100 items are optimal but should be piloted prior to finalisation to assess balance, coverage, accessibility, and clarity (Brown, 1980; Newman & Ramlo, 2010; Stainton Rogers, 1995; van Exel & De Graaf, 2005; Watts & Stenner, 2012). The final Q set in this study numbered 42 items and was piloted with a sample of five participants who undertook face-to-face Q sort interviews supplemented with procedural questions. Participants were asked to think aloud while sorting and answered questions probing the balance and coverage of the Q set, and the number of statements and their ability to summarise a holistic representation of participant viewpoints. Following piloting, minor adjustments were made to the instructions provided to participants, but the Q set remained unchanged (see table 7.1).

Table 7.1*Themes identified within the professionalism concourse.*

Theme	Statement number	Statement “For a person to be professional, they must...”
Accountability	[38]	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers
	[42]	Welcome a sense that they are accountable to others
Being a good/bad worker	[35]	Demonstrate a strict adherence to policy and procedural guidance at all times
Boundaries	[23]	Keep their personal lives separate from their professional - personal issues or experiences should not influence actions or attitudes in the workplace
Communication	[2]	Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang
	[4]	Be able to express themselves clearly, and in a focused and articulate way
Competency/ effectiveness and professional development/ education	[30]	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training
	[31]	Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills
	[33]	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done
	[34]	Be aware of the limits of their own competence, inform others about them, and remain within them
Congruence	[9]	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only
	[29]	Behave responsibly outside of the workplace, including being a positive role model for society
Ethics and morality	[32]	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved
	[36]	Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly
Humanistic	[6]	Not come across as uncaring or callous
	[8]	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone
Image management	[3]	Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others
	[7]	Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)
Independence and initiative	[26]	Be able to work independently and confidently according to their own initiative
Interpersonal skills	[1]	Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful
Leadership/being led	[14]	Be appropriately assertive and able to influence decisions
	[15]	Not be resentful of work-related scrutiny and challenge
Personal qualities and characteristics	[21]	Embody characteristics such as honour, duty, justice, and courage
	[39]	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity
	[40]	Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance
	[41]	Be able to manage stressful situations with objectivity rather than emotion
Reflection and self-awareness	[20]	Acknowledge, accept, and learn from their own mistakes
	[24]	Be perceived as demonstrating the required professional standards by those around them

	[25]	Not be defensive or easily upset in response to workplace events or change, or when being challenged
Relationships with colleagues	[5]	Work collaboratively with colleagues, rather than competitively
	[10]	Recognise that having positive relationships with colleagues is part of being good at your job
	[11]	Not be dishonest, but it is OK if there are times when they do not tell the whole story
	[12]	Be someone that demonstrates loyalty, someone that a colleague can rely on
	[13]	Have a keen awareness of where humour can become inappropriate
Relationships with customers	[27]	Build rapport with customers, in a way that encourages mutual respect and partnership
	[28]	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice
Social orientation	[16]	View themselves and other professionals as appointed leaders in society
	[17]	Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs
Sociological view and how professionalism has changed	[18]	Be free to exercise their expertise and judgement, without answering to a hierarchy of managers
	[19]	Take pride in their profession, and defend it to others
	[22]	Be defined by their quality or character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it
Wellbeing	[37]	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work

7.2.2 Participants.

Within Q methodology studies, the participant sample or P set (McKeown & Thomas, 2013, van Exel & De Graaf, 2005) should be sufficient to provide data saturation. Guidelines are tentative in nature, but P sets of between 40 and 60 participants are recommended, as are ensuring that the P set remains smaller than the Q set while also allowing for four to five significantly loading Q sorts onto each extracted factor (Brown, 1980; Brown, 1993; van Exel & De Graaf, 2005; Watts & Stenner, 2007, 2012). The P set in this study numbered 34 against a Q set of 42 supporting a four-factor solution and was therefore deemed appropriate in size.

Convenience sampling was employed to recruit participants aged 18 years or over using the social media platforms Facebook, LinkedIn, and Twitter for an electronic Q sort survey. In addition, strategic sampling was used to undertake Q sort interviews targeting participants representing a range occupations and professions. Participants received no incentive for taking part and following data collection, only participants returning fully completed Q sorts and demographic data were retained for analysis (n=34). The mean age of participants was 37.26 years (SD=13.07), ranging from 20 to 64 years. Participants were mainly female, although the gender balance was almost equal (F=16, M=15, declined=3), and were drawn from a range of occupational backgrounds. Participants reported an average number of years in the

workplace in total of 14.76 (SD=11.12), ranging from 1 to 40 years. For a full summary of the participant sample characteristics, see table 7.2.

Table 7.2*Summary characteristics of participant sample (n=34) and distribution of defining Q sorts.*

Characteristic		N	%	Factor defining Q sorts*			
				F1 (n=6)	F2 (n=11)	F3 (n=4)	F4 (n=3)
Age	18-30 years	10	29%	1	5	1	0
	31-45 years	13	38%	4	2	2	1
	46-60 years	8	24%	1	3	0	2
	61+ years	2	6%	0	0	1	0
	Declined	1	3%	0	1	0	0
Gender	Female	16	47%	0	7	2	1
	Male	15	44%	5	3	1	2
	Declined	3	9%	1	1	1	0
Sector of current/most recent occupation	Administration	6	18%	0	3	0	0
	Dental profession	5	15%	1	0	0	1
	Education profession	5	15%	2	1	1	0
	Health care profession	1	3%	0	0	0	0
	Management/ leadership	5	15%	1	2	0	1
	Medical profession	1	3%	0	0	0	0
	Sales profession	4	12%	1	3	0	0
	Security services	1	3%	0	0	1	0
	Other	5	15%	1	1	2	1
	Declined	1	3%	0	1	0	0
Time spent in workplace total	Early career (0-10 years)	14	41%	3	6	0	0
	Developing career (11-20 years)	9	26%	1	2	2	1
	Established career (21-30 years)	5	15%	1	0	1	2
	Late career (31+ years)	4	12%	1	2	1	0
	Declined	2	6%	0	1	0	0
Level of education completed	Levels 1-2 (compulsory high school e.g. GCSEs)	1	3%	1	0	0	0
	Levels 3-6 (further and undergraduate higher education)	15	44%	0	7	2	2
	Levels 7+ (postgraduate higher education)	15	44%	4	3	2	1
	Declined	3	9%	1	1	0	0

*Factor defining Q sorts were identified by a significant loading onto one factor and non-significant relationships with all other factors (for further details, see section 8.2.4).

7.2.3 Procedure.

7.2.3.1 Q sort interviews.

Prior to commencing interviews, participants were provided with information about the study and a consent form, with consenting participants then also completing a demographic questionnaire. Participants were briefly introduced to the topic of professionalism and asked to focus throughout the interview on their own views regarding it, rather than those held by society, their employer, or anyone else. They were advised that they would be presented with a series of statements that they may agree/disagree with or feel undecided/neutral towards and were introduced to the statement prefix. Understanding of how to use the prefix was checked verbally.

Participants were handed a shuffled set of cards, each printed with a single statement and its corresponding number and were informed that the number was an arbitrary shorthand for identifying each statement. They were asked to familiarise themselves with the Q set by reading each card and separating them into three piles: those statements they tended to agree with, those they tended to disagree with, and those towards which they were undecided or had no strong feelings. Participants were advised that they could undertake this task quietly or think aloud as they made their decisions depending on what they found most helpful and that notes would be taken regarding any comments they made as the process continued.

Following the initial sort, instructions were given to re-sort the cards aiming for a final distribution presented to them as a blank Q sort template (figure 7.1). The Q sort template was shaped to reflect a quasi-normal distribution in line with best practice guidance (Stainton Rogers, 1995; Watts & Stenner, 2012) with twelve columns ranging from -6 (most disagree) to +6 (most agree). Participants were informed that the positioning of statements within columns was meaningless and that the + and – symbols at the top of each column did not correspond to positivity or negativity respectively. They were provided with a written reminder of the statement prefix and presented with their ‘agree’ pile.

Participants were asked to move the cards that they more strongly agreed with towards the right of the table, and those that they least strongly agreed with towards the left. Following this, they were first asked to select the two statements

that they most strongly agreed with and to place them in the column labelled '+5'. Participants were then asked to select the three statements that they most strongly agreed with from the remainder and to place them in the column labelled '+4', and so on until all cards from this pile were sorted. Following this, the procedure was repeated for the 'disagree' pile using the opposite end of the template, and finally for the 'undecided/neutral' pile to fill in all remaining boxes within the template.

Figure 7.1. *Quasi-normally distributed Q sort template.*

For a person to be professional, they must...											
Most disagree			Neutral/undecided						Most agree		
-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6

NB/ where a statement includes the word 'customer', substitute this with a word to describe any individual requiring a service from the professional (customer/client/patient/service user).

Once all cards had been sorted, participants were asked to review their configuration and make any changes that they would like to. An example final Q sort is included in figure 7.2. Once participants confirmed satisfaction with their final configuration, it was recorded by hand on the template using statement numbers before a post sort interview was undertaken. The interview was semi-structured, using the following prompt questions:

- Regarding the statements placed in column +6, why did you place them there? What made you feel so strongly about them?

- Regarding the statements placed in column -6, why did you place them there? What made you feel so strongly about them?
- Did any statement appear out of place to you?
- Were any statements difficult to understand?
- Having given it some thought, what does the topic mean to you overall?
- Do you have any other comments about the statements or the issues in general that the sorting process made you think about?

The interviewer took notes throughout the sorting session and post-sort interview. These notes were used to inform the interpretation of results by identifying areas of difference between researcher and participant viewpoints, to minimise researcher bias and maximise shared understanding.

Figure 7.2. Example completed Q sort configuration.



7.2.3.2 Electronic Q sort.

The study was advertised via a post to the social media platforms Facebook, LinkedIn, and Twitter. The post included a link to complete the study and an invitation to share the post amongst readers' own networks. The online survey was created using version 1.0.4 of the HTMLQ software (see appendix E; Aproxima Gesellschaft für Markt- und Sozialforschung Weimar, 2015) and could be completed using participants' own devices. It was recommended that the survey be completed in a single sitting, but partially completed surveys could also be closed and returned to later. Upon clicking the link, information introducing the study and topic were displayed, along with a consent form before proceeding to the Q sort.

Participants were presented with each statement in turn on a 'card' that could be dragged and dropped to move it around. They were asked to read each card and drag it into one of three boxes according to their level of agreement with it. The first box was for items that the participant tended to disagree with, the second for those they tended to agree with, and the third for remaining cards that did not fit clearly into either of the first two boxes. Participants were asked to sort statements according to their own personal viewpoint rather than those that a colleague, employer, or anyone else might hold. They also received instructions for accessing the survey 'help' section and were informed of the prefix to be used for each statement.

Following initial sorting, participants were presented with the distribution template and asked to re-sort the statements into it one by one. They were also presented with the three boxes containing the pre-sorted statements. Participants were asked to scroll through their agree box and choose the two statements that they most agreed with, before dragging these into the column labelled '+6'. They were then asked to scroll through their disagree box and choose the two statements they most disagreed with, before dragging these into the column labelled '-6'. This process was repeated for each column until all agree/disagree statements had been sorted, before participants were asked to fill the remaining boxes with statements from the neutral box. Once all statements had been sorted into the template, they were asked to review their configuration and make any changes they would like before confirming it.

Participants were then presented with post-sort interview questions with free-text response boxes. They were presented with the statements that they most agreed and disagreed with and asked to explain why they had placed them as such. Finally, there were presented with a demographic questionnaire for completion before submitting their data.

7.2.4 Analyses.

Completed grids were correlated and subjected to by-person factor analysis using a combination of hand and varimax rotation, using the specialist Q

methodology software package PQMethod (Schmolck, 2014). For a full discussion of by-person factor analysis including the analytical approach, factor extraction, rotation, arrays, and interpretation, see chapter 4, section 4.2.3.5.1. Factor rotation allows researchers to view factor analysis results from different viewpoints (Brown, 1980). Hand rotation is recommended by traditional Q methodologists as it allows researchers to manipulate data manually to discover perspectives that might be missed by statistical rotation (McKeown & Thomas, 2013). Varimax rotation is generally favoured today, however, as it rotates data automatically to account for maximum possible variance in statistical terms (Stainton Rogers, 1995; van Exel & de Graaf, 2005). The former is recommended for in-depth analysis of a small number of Q sorts while the latter is well-suited to larger studies seeking to account for maximum shared subjectivity. A combination of both enables a balance of employing the methodology in line with its traditional aims and underlying theory, while meeting the ends of exploring shared subjectivity (Block, 2008; Watts & Stenner, 2012).

Best practice in Q methodology prioritises meaningful contribution to qualitative understanding of the topic under study, and factor extraction is therefore recommended to be undertaken on a primarily qualitative basis with reference to existing theory and qualitative comment data gathered from the P set. Within the present study, in the absence of existing theory regarding professionalism, qualitative consideration was given to the level of coherency provided by factors based on their factor arrays and comments associated with defining Q sorts. In other words, factor extraction was based on the degree to which a factor ‘made sense’ in the context of participant comments and the construct of professionalism. However, as the aim of this study was to explore shared subjectivity, quantitative criteria were also considered. These criteria included extracting factors with an eigenvalue exceeding one and those with four or more loading/defining Q sorts, Humphrey’s rule that a factor be deemed significant where multiplying its two highest loadings results in a value higher than twice the standard error, visual inspection of scree plots, and a factor solution resulting in an acceptable amount of variance being accounted for, with the minimum level of acceptability set at 35% (Brown, 1980; Kline, 1994; Watts & Stenner, 2012). Both qualitative and quantitative criteria were considered

during factor extraction with the overriding aim being to account for maximum variance without compromising the coherency and usefulness of the factor solution.

Each extracted factor constituted a major viewpoint shared by the P set. Significantly loading Q sorts were identified by statistically significant similarity between its sorting pattern and that associated with the factor. Statistical significance was calculated using a standardised formula representing best practice in Q methodology (Brown, 1980; Watts & Stenner, 2012), which multiplies the standard error by 2.58, rounded to two decimal places. In this study, a loading of 0.40 or above was calculated to be statistically significant ($p < 0.01$). Where a Q sort loaded significantly onto one factor and had a non-significant relationship with all other factors, it was identified as a factor defining Q sort.

Following factor extraction, results were used to create factor arrays (McKeown & Thomas, 2013; Watts & Stenner, 2007, 2012). Factor arrays are completed Q sorts representing the configuration expected from a perfect correlation with that factor. Factor arrays may also therefore be thought of as merged statistical approximations of all the defining Q sorts associated with each factor (Watts & Stenner, 2012). Factor arrays were created using a weighted average of defining Q sorts, meaning that sorts with higher factor loadings were weighted more heavily and therefore had a greater impact on resulting factor arrays. Each statement within factor arrays received a factor score, providing an indication of the strength of feeling towards that statement within the given factor; higher factor scores indicated stronger agreement and lower scores indicated unimportance or rejection of that statement. In order to enable cross-factor comparisons, factor scores were standardised before factor arrays were created; Z scores were calculated and ranked to indicate the position of each statement within the factor array (Watts & Stenner, 2012).

Completed factor arrays formed the basis of factor interpretation, supplemented by qualitative data gathered during post-sort interviews. Factor arrays were subjected to detailed scrutiny, with the aim of revealing and explaining the viewpoint captured therein (see appendix F for full details). Scrutiny focused on statements generating particular strength of feeling within factor arrays, namely those placed at the most agree/disagree positions ('+6' and '-6') and distinguishing

statements. Distinguishing statements were identified as those occupying locations within one factor array that were statistically significantly different to their location in all other factor arrays. Factor interpretation considered qualitative data from both post-sort interviews and observations or comments made spontaneously during sorting within interviews. Final factor interpretations took the form of holistic, narrative descriptions of the viewpoint captured by each factor (Block, 2008; Watts & Stenner, 2012).

7.2.5 Intersubjectivity.

As discussed within chapter 4, section 4.2.3.5.2, intersubjectivity requires the use of reflexivity and/or triangulation exercises to enhance transparency and acknowledge potential researcher bias in interpreting factors (Creswell & Creswell, 2018; Denzin & Lincoln, 2008; Guba & Lincoln, 2008; Johnson et al., 2007; Teddlie & Tashakkori, 2009). To this end, factor interpretation was undertaken by the lead researcher with a second member of the research team periodically evaluating samples to provide scrutiny and challenge regarding accuracy and appropriateness. The lead researcher also undertook an exercise using an amended version of the approach described by Gauttier (2017). In response to claims of interpretation bias in phenomenological research, Gauttier (2017) sought to increase transparency in factor interpretation by using Q methodology to collect participant viewpoints, but also explore researcher interpretation of those viewpoints. This was achieved by completing the same Q sort as participants, from what the researcher perceived to be their perspectives. This approach was used in the present study by the lead researcher who completed Q sorts from the perspective of each factor, based on the current iteration of the factor interpretation and its description. These Q sorts were entered into the same analyses as the participant Q sorts to ascertain their level of similarity to the factors extracted. Where a factor had been understood and described accurately, it was hypothesised that the researcher would create a Q sort from that factor's perspective that would load significantly onto the correct factor alone. Where reflexive Q sorts failed to load significantly onto the target factor, or cross-loaded onto multiple factors, they were further refined and scrutinised to improve accuracy. This process was repeated until the lead researcher had achieved

sufficiently accurate understanding of the factor to enable repeated significant loadings.

7.3 Results

Thirty-four completed grids were correlated and subjected to by-person factor analysis using a combination of hand and varimax rotation, undertaken using PQ Method (Schmolck, 2014). Four factors were extracted, accounting together for 42% of the variance observed amongst the P set. Table 7.3 displays the variance accounted for by each factor and rotated factor loadings, and identifies each factor's defining Q sorts. Six Q sorts cross-loaded significantly onto more than one factor and were therefore not used to inform factor interpretation. Factor interpretations take the form of narrative descriptions written from the perspective of a hypothetical individual with a perfect correlation with that factor.

Table 7.3

Rotated Q sort factor loadings rounded to two decimal places and sector of current or most recent employment.

Q sort	Employment sector	Factor 1	Factor 2	Factor 3	Factor 4
1	Sales	0.0402	0.4022X	0.3696	0.2531
2	Other (student)	0.3370	0.2917	0.5731X	0.1509
3	Other (project management)	0.0374	0.5631X	0.2984	0.1108
4	Administration	0.4193	0.3753	0.4031	0.0586
5	Other (research)	0.7274X	0.1283	0.2015	0.0787
6	Other (fire and rescue services)	0.1715	0.0790	0.5153X	-0.1665
7	Education	0.0610	0.3684	0.4016X	0.3568
8	Administration	0.3546	0.4236X	0.1614	0.0574
9	Administration	0.0389	0.7111X	-0.0609	-0.1674
10	Management/leadership	0.6372X	0.0555	0.0485	0.0151
11	Medical profession	0.4995	0.1324	0.6241	0.0935
12	Sales	0.2494	0.5606X	0.0622	-0.1030
13	Education	0.6582X	0.2457	0.1758	0.0703
14	Administration	-0.0320	0.1071	0.3886	0.2034
15	Security services	0.2739	0.0503	0.6894X	0.2100
16	Administration	0.1004	0.4223X	0.2955	0.1081
17	Education	0.5673X	0.1481	0.2086	-0.1561
18	Healthcare	0.2871	0.1038	0.0143	0.1307
19	Management/leadership	0.1813	0.5557X	0.2050	0.3845
20	Declined	0.2735	0.4657X	0.2078	0.0743
21	Education	0.1302	0.6605X	-0.0912	0.0066
22	Sales	0.2444	0.4485X	0.2108	0.0200
23	Management/leadership	0.1377	0.4136	0.0965	0.4819
24	Dental profession	0.0975	-0.1359	0.1787	0.7056X
25	Management/leadership	0.3378	0.1564	-0.0781	0.4267X
26	Dental profession	0.0974	-0.0615	0.2350	0.3594
27	Dental profession	0.5540	-0.0722	0.1605	0.6085
28	Dental profession	0.4451	0.0328	0.4511	0.4273
29	Management/leadership	0.1032	0.6022X	-0.3133	0.1037
30	Administration	-0.0123	-0.0989	0.1466	0.0633
31	Sales	0.3952X	0.2601	-0.0259	0.2917
32	Dental profession	0.5840X	0.1880	0.0912	0.2398
33	Education	0.2194	0.5063	0.4018	0.2452
34	Other (financial services)	-0.0542	0.0993	0.0508	0.5493X
% variance explained		12%	13%	9%	8%

Note. Significant loadings ($p < 0.01$) identified as defining Q sorts are indicated by bold typeface and X, according to the criteria described earlier.

7.3.1 Factor descriptions.

7.3.1.1 Factor 1: intellectual professionalism.

The factor array for factor one is presented in figure 7.3. Factor one has an eigenvalue of 8.40 and explains 12% of the study variance. Six participants were significantly associated with this factor. Five were male (one participant declined to provide their gender), and they had an average age of 39 years (ranging from 28 to 55 years). Three participants reported working in education, with the remainder working in leadership/management, sales, and dentistry. Participants had been in the

workplace in total for an average of 16 years (ranging from 5 to 35 years). One participant had completed doctorate level education, three had completed a Masters degree or equivalent, and one had completed GCSE level education (one participant declined to answer). The researcher's interpretation of *intellectual professionalism* is described in detail in table 7.4.

Table 7.4

Description of factor 1: intellectual professionalism.

Intellectual professionalism

What professionalism should look like is more important than what it shouldn't. Characteristics that are less important tend to have a context-dependent nature and may be subject to more significantly caveated agreement rather than being actively disagreed with. For example, physical appearance is generally unimportant to professionalism (3: 2; 7: 1)*, but there may be circumstances where it is worthwhile considering its impact, even though it does not specifically reflect upon professionalism. For example, one must acknowledge that the perceptions of others may be swayed by cultural norms or vary widely due to the ambiguity of acceptable standards, and therefore separate such perceptions from the appraisal of professionalism. Similarly, although an individual's conduct outside of the workplace is irrelevant to professionalism, it may be an issue for those in occupations where regulators pay attention to this (29: 2).

Professionals are identified by their superior knowledge and so discussions with customers should not be led by them or their perceptions of the professional, but instead focus on what the professional believes is most important (14: 7). Although not a bad thing, collaborative customer relationships are relatively unimportant to professionalism (27: 6), as professional practice is about acting in line with the evidence rather than the way you feel about each other (8: 6). As professionalism relies upon an analytical approach rather than the way individuals relate to one another, there is no need for special attention to particular groups of individuals (17: 5), as decisions are based on the best available evidence and will therefore always ensure an appropriate course of action is taken. Ultimately, professionalism is not a matter of the subjective perceptions of others (3: 2; 24: 4), it is more closely related to the individual's ability to use their expertise to be maximally effective (33: 10).

Professionalism requires that a person generally gets along with colleagues, but in a way that enables them to perform well in their role (10: 8). For example, being collaborative rather than competitive (5: 10) and enjoying workplace humour (13: 5) are important because competition and discord may have a detrimental effect on work performance and results. However, it is important to recognise that there may be times where colleague relationships breakdown; this does not make an individual unprofessional as long as their performance is not affected (10: 8). In addition, although there should be loyalty amongst colleagues to an extent, this should not be blind in ways that would impact performance (12: 6). Where colleagues exhibit objectively sub-optimal performance, although it is important that they can recognise and rectify this themselves, a professional recognises that it may sometimes be appropriate to bring relevant evidence to their attention (38: 7). Generally speaking, colleague relationships are relevant to professionalism but only in as much as they impact upon work performance (10: 8).

One of the key themes of professionalism is that the individual must be effective in their role and, in particular, independent enough to perform well without close supervision (26: 12). This independence is supported by having a complete skill set (31: 10) and being able to deploy that skill-set in a way that ensures an effective, evidence-based approach (33: 10); using objective knowledge and reasoning ensures that colleagues can be trusted to

work effectively on their own initiative. It also equips them with the flexibility to solve problems that are less than clear-cut (39: 9), due to their grasp of the objective facts. Being led by the information available means that at times, procedural guidance may not provide the best course of action. At such times, the systematic and evidence-based practice of a professional means that they can challenge and improve this guidance without blindly following it (35: 3).

Physical and mental wellbeing are fairly important to professionalism, because when a colleague is unwell, it may have a detrimental impact upon their performance (37: 9). Performing well and improving are unemotional matters; feedback and challenge offer opportunities for improvement, not emotional reaction (25: 11; 15: 9). Unless a person does their job well, they cannot be considered professional as it is not about character alone (22: 4), and if they do perform their job well, individual self-improvement and ongoing development are rendered less important as the skills and expertise required are already in place (20: 8; 30: 8).

An important aspect of being effective in a role is being objective. Taking an objective, evidence-based approach ensures that the right course of action is decided upon (14: 7) and that when things go wrong or are questioned, emotional reactions are not required (15: 9; 25: 11; 41: 11). Having a firm grasp on the objective facts enables professionals to be resolute in their assertions, without concern for causing offence, as it is based on logic and not personal or emotional reasoning (14: 7).

The most clearly defined aspect of professionalism is that an individual must be a fundamentally good person. Although an individual may not be professional without being effective in their role (22: 4), this incorporates aspects that are subject to context but those characteristics that make one a good person are of unwavering and necessary importance. This does not mean however that the person must embody hero-like qualities (21: 7), but rather that they exhibit congruence, meaning that what you see outwardly matches what they think inwardly; they are as upstanding in their character as they are effective in their role (9: 12; 26: 12). Their actions are underpinned by a strong but appropriate moral and ethical compass (32: 11), which has a positive influence on their actions in the workplace (23: 4) and enables all situations to be regarded dispassionately to ensure fair and equal treatment for all. Being guided by a good character, a professional objectively respects their colleagues (5: 10; 10: 8), which may include having uncomfortable conversations with them on occasion (38: 7). The professional is always honest and has no need to lie (11: 1) because their decisions and actions are always supported by the objective facts, and their strong moral compass means that they can be trusted to work independently and on their own initiative in making evidence-based and objective decisions (26: 12). Overall, a professional is very effective in their role, but achieves this through information and reasoning, and not by compromising their personal values.

Note. This table provides a long-form narrative description of the viewpoint associated with factor 1, according to the best practice procedure described by Watts and Stenner (2012). This factor represents a composite amalgamation of all those participant viewpoints defining it and so the description is written from the perspective of a hypothetical participant with a perfect loading to factor 1.

*Numbers in brackets denote the statement number and position within the factor array that different aspects of the factor description relate to, with columns numbered 1-12 from left to right. As such, (3: 2) indicates that statement number 3 was within column number 2, as shown within figure 7.3.

Figure 7.3. Factor array for intellectual professionalism.

For a person to be professional, they must...

Most disagree			Neutral/undecided						Most agree		
1	2	3	4	5	6	7	8	9	10	11	12
11 Not be dishonest, but it is OK if there are times when they do not tell the whole story	29 Behave responsibly outside of the workplace, including being a positive role model for society	18 Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	24 Be perceived as demonstrating the required professional standards by those around them	13 Have a keen awareness of where humour can become inappropriate	12 Be someone that demonstrates loyalty, someone that a colleague can rely on	38 Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	1 Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	37 Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	31 Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills	32 Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	26 Be able to work independently and confidently according to their own initiative
7 Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	16 View themselves and other professionals as appointed leaders in society	35 Demonstrate a strict adherence to policy and procedural guidance at all times	22 Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	17 Demonstrate a commitment to prioritising the underserved and those with diverse or unmet needs	36 Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	4 Be able to express themselves clearly, and in a focused and articulate way	10 Recognise that having positive relationships with colleagues is part of being good at your job	39 Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	33 Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	41 Be able to manage stressful situations with objectivity rather than emotion	9 Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only
3 Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	28 Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	19 Take pride in their profession, and defend it to others	40 Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	2 Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang	21 Embody characteristics such as honour, duty, justice, and courage	20 Acknowledge, accept, and learn from their own mistakes	34 Be aware of the limits of their own competence, inform others about them, and remain within them	5 Work collaboratively with colleagues, rather than competitively	25 Not be defensive or easily upset in response to workplace events or change, or when being challenged		
		23 Keep their personal and professional lives – personal issues or experiences should not influence actions or attitudes in the workplace	6 Not come across as uncaring or callous	27 Build rapport with customers, in a way that encourages mutual respect and partnership	42 Welcome a sense that they are accountable to others	30 Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	15 Not be resentful of work-related scrutiny and challenge				
				8 Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	14 Be appropriately assertive and able to influence decisions						

Note. For further detail regarding the preparation and use of factor arrays, please see chapter 4, section 4.2.3.5.1.

7.3.1.2 Factor 2: collaborative professionalism.

The factor array for factor two is presented in figure 7.4. Factor two has an eigenvalue of 2.35 and explains 13% of the study variance. Eleven participants were significantly associated with this factor, although the following details are based largely on ten as one participant declined to complete the demographic questionnaire. Participants were seven females and three males with an average age of 36 years (ranging from 20 to 57 years). Three participants reported working in administration and three in sales, two in leadership/management, and one in education and project management respectively. Participants had been in the workplace in total for an average of 15 years (ranging from 1 to 40 years). Three participants had completed a Masters degree or equivalent, four had completed a first degree or equivalent, and three had completed A levels or equivalent. A detailed description of the researcher's interpretation of *collaborative professionalism* is included in table 7.5.

Table 7.5

Long-form narrative description of factor 2: collaborative professionalism.

Collaborative professionalism

A number of key themes are important to professionalism, but it is not just about the way colleagues conduct themselves, it is also important to clarify what should and should not be encompassed by it. For example, it is extremely important that judgements regarding professionalism are limited to conduct in the working environment and during working hours, as behaviour outside of these contexts is irrelevant (29: 1)*. Individuals should be free to behave as they wish outside of work, with no reflection upon their professional status or reputation. Although professionals must demonstrate congruence in the workplace, it remains attached to the professional role and stops at the boundary of the working day. As a result, it is not something that individuals must embody with their whole person, just their professional self (9: 2). Personal appearance should be treated similarly; personal preferences in style and presentation are irrelevant to professionalism (7: 2). Although the judgements of others should remain within these limits, part of the responsibility for ensuring the separation of personal and professional lies with the professionals themselves; they must ensure that they leave their personal lives and personalities at home (23: 10). This issue fundamentally relates to ensuring fairness in appraising professionalism; it is simply unfair to include certain aspects of personal lives and personalities (29: 1). Failure to appropriately separate these issues leads to unfair expectations such as requiring unreasonable sacrifices for the workplace role (16: 1; 28: 2; 29: 1).

Having said this, fairness is broader still than the criteria used by others to judge professionalism. It also pertains to the status of professionals themselves, and specifically that this should remain equal to all others (16: 1; 27: 9; 36: 3). Professional expertise does not provide an individual with any moral or implied right to take a position of superior status or power. Professionals must respect all others and their contributions in the same way that all others should equally respect theirs; mutual respect is very important. Similarly, professionals ought to be treated fairly by their employers or regulators. It is unfair for example to expect professionals, simply by virtue of their conduct, to be able to recognise any issues relating to their mental wellbeing when by definition, one may be unable to do so (37: 3); even the highest level of professionalism cannot change this. Similarly, it is important to acknowledge that all self-knowledge is finite, so a professional may at times require support to recognise the limits of their own competence and where additional training may be required. As such, automatic self-awareness of such issues should not be expected as part of professionalism (30: 6; 34: 5). There should be clear rules for all colleagues within a given context, and it is the employer's responsibility to ensure that these are clearly articulated and implemented (4: 11).

When in the workplace, it is important to clearly define where professionalism ends. Some characteristics are generally positive and to be encouraged, but not actually necessary for a person to be considered professional. For example, while it is potentially advantageous to employ an individual who takes pride in their profession and defends it accordingly to others, this is not actually necessary for a person to embody professionalism (19: 4). Similarly, a colleague guided by sound moral and ethical principles may also be desirable, but where this is lacking, a colleague can still deliver professionalism (32: 5; 11:

5). Likewise, although one may wish to get along with all colleagues in positive and productive relationships, one can still be professional at times where this is not the case (10: 8). Finally, unrealistic expectations of a noble, hero-like character are also unnecessary to professionalism; although potentially desirable in any individual, this is not a necessary characteristic of a professional (8: 3; 21: 4).

Moving on to what professionalism is and should be, some characteristics although necessary for professionalism, are less important as they constitute a minimum baseline from which true professionalism may be built. Such characteristics are akin to just being a satisfactorily mature adult in the workplace, such as having a broadly effective and balanced skill set (31: 7), being able to present oneself in attire appropriate to the task to be undertaken (3: 7), not being dishonest (11: 5), and being respectful of others (1: 7) by not being callous or uncaring towards their welfare (6: 7; 17: 6). It also includes recognition that no colleague is beyond accountability (18: 4; 42: 7), and the ability to manage one's emotions in a way that enables them to be valuable rather than problematic (41: 9). For example, although an individual may disagree with or even resent workplace challenge (15: 6), they express this appropriately and by making logical rather than emotional arguments to support their viewpoint (25: 11). This ability relates to the earlier point of leaving personal lives at home; emotional feelings and reactions are a personal characteristic and are therefore not appropriate in a professional context (23: 10). Ultimately, professionalism requires the ability to express oneself clearly, and to be able to articulate their thoughts in ways that do not constitute emotional reactivity (4: 11).

Higher levels of professionalism are characterised by an individual's recognition that their own intentions are only half the story; a person must also be deemed to be professional by those around them. However, it is important to acknowledge that the perceptions of others may not be the same as the standards required by your profession or employer (24: 8), meaning that their perceptions may not necessarily be based on professional standards but rather the less tangible and more qualitative aspects of conduct (24: 8). For example, in cases where the perception of others contravenes professional guidelines, an effective professional does not deny this feedback, but instead treats such instances as areas for development and seeks to learn from them (20: 12) in ways that will ensure they can balance professional standards and interpersonal relationships more effectively in future. They are able to communicate clearly and effectively to build rapport with customers, enabling mutual understanding and collaboration even where personal views do not coincide with professional requirements (4: 11; 5: 10; 24: 8; 27: 9). As a result, a true professional will learn to effectively manage the way that they come across to others using their communication skills rather than manipulative interpersonal strategies (31: 7), ensuring that customers fully understand and are satisfied with the service they have received (4: 11; 7: 9; 27: 10).

A professional also has positive relationships with their colleagues (10: 8), as achieved through their actions and communications, rather than their ability to manage their own image through interpersonal manipulations (31: 7). They are perceived as demonstrating a high level of reliability, loyalty, and commitment to the work of the team (12: 10), as they do not take a competitive approach (5: 10). Their proven track record ensures that a professional can be trusted to deliver on their objectives consistently and independently, without a need for constant input from others (12: 10; 26: 9). This supports positive relationships with colleagues (10: 8) but also involves taking an appropriately critical approach to their own work as a means of ongoing improvement (38: 8). Professionalism is not about getting along for the sake of harmony; it is about preserving

productive and effective relationships (10: 8; 33: 11; 38: 8). This may require a different set of skills to those required to meet professional or regulator standards (24: 8). As a result, in order to remain perceived as professional, an individual may need to learn to shift their focus from technical competency toward the communication skills required to preserve collaboration, even when disagreement occurs (5: 10; 20: 12; 31: 7). Clear communication ensures that in situations where there are differences of opinion, a way forward is found together through collaboration rather than competition (5: 10). For example, while one colleague may take a liberal approach to defining the appropriateness of workplace humour, another may feel that lines are crossed or take offence; part of being professional is recognising that how humour is perceived defines its appropriateness, engaging in open discussion to understand the nature of differing perceptions, and responding accordingly (13: 9). Clear and honest communication supports this, by enabling both parties to understand the perspective of the other (4: 11). How others perceive your professionalism is just as important as how you intend it to be, and so being challenged by others, although not always welcome emotionally (15: 6), should be expected and utilised as a means of improvement behaviourally (15: 6; 25: 11).

Overall however, the most important characteristic of a professional is the way that they go about ensuring they do their job to the highest standards. Professionalism is less about having all the technical skills and experience a job may require, and more about having the humility to be honest about your own performance and times when you haven't gotten something quite right (20: 12; 31: 7). A professional takes a critical stance towards their own performance that allows them to acknowledge their mistakes and take steps to ensure they become a learning experience (20: 12; 23: 10; 25: 11), and this means that they are never defensive or emotionally reactive when challenged by others (25: 11). Critical self-awareness also enables a professional to be flexible in working with the demands of their role. This supports their ability to manage any stress constructively and with objectivity (39: 8; 41: 9), so that they may clearly articulate their emotions rather than display them directly (4: 11). The professional also then has the ability to ensure their attitudes and behaviours remain appropriate at all times, regardless of the situation. This ensures that less formal interactions remain suitable for the professional context (13: 9), and that more formal interactions remain respectful (27: 9). It ensures that a colleague remains a collaborative partner and not an enemy in workplace competition (5: 10). Overall, professionalism is about the character of the person, rather than their job title or any technical qualification they may possess (22: 12).

Note. This table provides a long-form narrative description of the viewpoint associated with factor 2, according to the best practice procedure described by Watts and Stenner (2012). This factor represents a composite amalgamation of all those participant viewpoints defining it and so the description is written from the perspective of a hypothetical participant with a perfect loading to factor 2.

*Numbers in brackets denote the statement number and position within the factor array that different aspects of the factor description relate to, with columns numbered 1-12 from left to right. As such, (29: 1) indicates that statement number 29 was within column number 1, as shown within figure 7.4.

Figure 7.4. Factor array for collaborative professionalism.

For a person to be professional, they must...

Most disagree			Neutral/undecided						Most agree		
1	2	3	4	5	6	7	8	9	10	11	12
29 Behave responsibly outside of the workplace, including being a positive role model for society	9 Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	36 Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	40 Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	35 Demonstrate a strict adherence to policy and procedural guidance at all times	2 Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang	1 Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	10 Recognise that having positive relationships with colleagues is part of being good at your job	26 Be able to work independently and confidently according to their own initiative	5 Work collaboratively with colleagues, rather than competitively	4 Be able to express themselves clearly, and in a focused and articulate way	22 Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it
16 View themselves and other professionals as appointed leaders in society	28 Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	8 Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	19 Take pride in their profession, and defend it to others	32 Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	14 Be appropriately assertive and able to influence decisions	31 Have a well-rounded set of skills – they must demonstrate both technical expertise and interpersonal skills	38 Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	13 Have a keen awareness of where humour can become inappropriate	23 Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	33 Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	20 Acknowledge, accept, and learn from their own mistakes
	7 Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	37 Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	21 Embody characteristics such as honour, duty, justice, and courage	34 Be aware of the limits of their own competence, inform others about them, and remain within them	30 Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	6 Not come across as uncaring or callous	39 Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	41 Be able to manage stressful situations with objectivity rather than emotion	12 Be someone that demonstrates loyalty, someone that a colleague can rely on	25 Not be defensive or easily upset in response to workplace events or change, or when being challenged	
			18 Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	11 Not be dishonest, but it is OK if there are times when they do not tell the whole story	17 Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs	42 Welcome a sense that they are accountable to others	24 Be perceived as demonstrating the required professional standards by those around them	27 Build rapport with customers, in a way that encourages mutual respect and partnership			
					15 Not be resentful of work-related scrutiny and challenge	3 Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others					

Note. For further detail regarding the preparation and use of factor arrays, please see chapter 4, section 4.2.3.5.1.

7.3.1.3 Factor 3: *personal professionalism*.

The factor array for factor three is presented in figure 7.5. Factor three has an eigenvalue of 1.75 and explains 9% of the study variance. Four participants were significantly associated with this factor. They were two females and one male (one participant declined to provide their gender) and had an average age of 42 years (ranging from 28 to 61 years). Two participants worked in education, and the others in the fire and rescue service and security services. Participants had been in the workplace in total for an average of 22 years (ranging from 12 to 35 years). Two participants had completed a Masters degree or equivalent, one had completed a first degree or equivalent, and one had completed a Diploma of Higher Education or equivalent. A detailed description of the researcher's interpretation of *personal professionalism* is included in table 7.6.

Table 7.6

Long-form narrative description of factor 3: personal professionalism.

Personal professionalism

Professionalism is defined more by what you should be than what you shouldn't. Where an individual embodies key characteristics of professionalism, everything else will follow naturally and as a result become less important. For example, in order to be professional, one must be good at their job, which primarily requires them to be a strong communicator (2: 9; 4: 9)* who can use humour appropriately (13: 8) and build rapport effectively with stakeholders (27: 7). They also have the flexibility to adapt to challenges (39: 8) and are perceived by others to be a reliable and loyal colleague (12: 8; 24: 8). They are not overly emotionally reactive (25: 7; 41: 6) and get on fairly well with their colleagues (10: 6). However, this collection of skills and abilities, although important, are not extremely so, as they are all underpinned by something more general. The above professional skills and abilities are actually based on much more personal aspects of the individual's overall quality of character.

Accountability is important to professionalism, but it must crucially be exercised in the right way (42: 7). More specifically, although accountability requires an individual to recognise the limits of their role and to act in accordance with them (18: 4; 34: 7), it is not about blind compliance or acceptance; it is important for a professional to be willing and able to challenge colleagues, their employer, or the profession. For example, while policy and procedural guidelines serve a purpose, blind compliance with them is not professional and so they should not be followed with too strict an adherence (35: 2). Instead, where there is a valid reason, a professional should challenge them in order to make improvements. This does not mean that professionals are above accountability however, as working too independently and without regard for colleagues or guidance is also problematic (26: 3). As a result, being held accountable for their behaviour both within work and personal contexts (29: 7) is generally something unavoidable that a professional must accept (25: 7; 42: 7), although they are entitled on a personal level to resent it (15: 5). In addition, being professional does not require someone to accept accountability for the conduct, performance, or behaviour of their colleagues (38: 4), or indeed the wider profession (19: 3), as these are role-specific requirements that are transcended by professionalism.

In order to perform well, the difference between relationships with customers and colleagues must be acknowledged. Having positive relationships with customers and other stakeholders is part of doing a job well and is therefore important (27: 10), but when it comes to colleagues, it is more important that relationships are productive than positive in an interpersonal sense (10: 6). As a result, colleague relationships should be less focused on interpersonal skills (31: 6) and instead prioritise improvement. As long as colleague relationships remain respectful (1: 12), it is possible for a professional to have characteristics that not everyone may like on a personal level and to even work with such characteristics in a positive way. This means that that attributes such as assertiveness, a competitive streak, or a direct nature (5: 4; 6: 4; 40: 6) can be used constructively to enhance performance. Fundamentally, this speaks to the ability to see the value of individual colleague personalities and stories. For example, although personal style choices are often discussed as potentially problematic in relation to professionalism, diversity of

workforce is actually an asset, so far from being problematic, individuality of expression should be encouraged (7: 1). Although technically there may be times when clothing is relevant to the workplace (3: 5), imposing false boundaries that require individuals to sacrifice their personal selves for separate and different professional identities precludes opportunities to capitalise on the benefits of diversity and should therefore be discouraged (23: 2; 28: 1). Similarly, although some may feel that personalities high in competitiveness or assertiveness may cause issues in the workplace, a professional appreciates that understanding these characteristics in their colleagues and themselves mean that they can be harnessed to positive effect and should therefore be respected for their potential value (40: 6). By learning about the more personal aspects of others, a professional develops the ability to take their perspectives and can then harness their individuality to improve their conduct and performance (1: 12). Overall, professionalism has a much closer relationship to the personal than might otherwise be thought.

All of the above characteristics are part of a larger and more important attribute. Specifically, professionalism is based on a person's character rather than a role they choose to adopt at certain times (9: 10), meaning that they will generally already be equipped with these characteristics. Professionalism requires a person to be of good character personally, striving for ideological characteristics such as honesty and loyalty, and a sense of duty and justice, as guided by their sound moral principles (11: 3; 12: 8; 21: 10; 32: 9). It is important that they live the experience of professionalism, rather than perceiving it as something to be switched on and off in different situations (9: 10; 22: 11; 29: 7). This all supports one of the most important aspects of professionalism, namely that a professional uses their insight into the perspectives of others to always ensure that their conduct is respectful, regardless of the situation or person being spoken to (1: 12).

Of equal importance to this is that a professional recognises that their ability to perform well in all of their work-related duties may always be further improved (30: 12). Professionals are characterised by the way that they apply themselves to a role (33: 11), rather than their existing skills (31: 6) or their job title (22: 11). When a mistake is made, a professional uses this to inform their ongoing development (20: 9). As a result, a professional need not engage in excessive self-sacrifice to perform well (28: 1), which in turn supports professionals to take care of their own wellbeing (37: 11). This cycle of ongoing learning and development is vitally important to professionalism (30: 12).

Note. This table provides a long-form narrative description of the viewpoint associated with factor 3, according to the best practice procedure described by Watts and Stenner (2012). This factor represents a composite amalgamation of all those participant viewpoints defining it and so the description is written from the perspective of a hypothetical participant with a perfect loading to factor 3.

*Numbers in brackets denote the statement number and position within the factor array that different aspects of the factor description relate to, with columns numbered 1-12 from left to right. As such, (2: 9) indicates that statement number 2 was within column number 9, as shown within figure 7.5.

Figure 7.5. Factor array for personal professionalism.

For a person to be professional, they must...

Most disagree			Neutral/undecided						Most agree		
1	2	3	4	5	6	7	8	9	10	11	12
7 Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	35 Demonstrate a strict adherence to policy and procedural guidance at all times	26 Be able to work independently and confidently according to their own initiative	38 Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	14 Be appropriately assertive and able to influence decisions	41 Be able to manage stressful situations with objectivity rather than emotion	42 Welcome a sense that they are accountable to others	24 Be perceived as demonstrating the required professional standards by those around them	4 Be able to express themselves clearly, and in a focused and articulate way	9 Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	33 Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	30 Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training
28 Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	23 Keep their personal and professional lives – personal issues or experiences should not influence actions or attitudes in the workplace	19 Take pride in their profession, and defend it to others	5 Work collaboratively with colleagues, rather than competitively	3 Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	36 Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	8 Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	39 Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	20 Acknowledge, accept, and learn from their own mistakes	27 Build rapport with customers, in a way that encourages mutual respect and partnership	37 Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	1 Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful
	16 View themselves and other professionals as appointed leaders in society	11 Not be dishonest, but it is OK if there are times when they do not tell the whole story	6 Not come across as uncaring or callous	17 Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs	10 Recognise that having positive relationships with colleagues is part of being good at your job	34 Be aware of the limits of their own competence, inform others about them, and remain within them	12 Be someone that demonstrates loyalty, someone that a colleague can rely on	2 Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang	21 Embody characteristics such as honour, duty, justice, and courage	22 Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	
			18 Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	15 Not be resentful of work-related scrutiny and challenge	31 Have a well-rounded set of skills – they must demonstrate both technical expertise and interpersonal skills	29 Behave responsibly outside of the workplace, including being a positive role model for society	13 Have a keen awareness of where humour can become inappropriate	32 Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved			
					40 Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	25 Not be defensive or easily upset in response to workplace events or change, or when being challenged					

Note. For further detail regarding the preparation and use of factor arrays, please see chapter 4, section 4.2.3.5.1.

7.3.1.4 Factor 4: *professional professionalism*.

The factor array for factor four is presented in figure 7.6. Factor four has an eigenvalue of 1.64 and explains 8% of the study variance. Three participants were significantly associated with this factor. Although this was lower than the quantitative guidance of four to five significant loadings, the content of the factor was examined qualitatively and deemed of sufficient interest to the topic to warrant extraction. The significantly associated participants were one female and two males and had an average age of 43 years (ranging from 33 to 49 years). One worked in financial services, one in dentistry, and one in leadership/management. Participants had been in the workplace in total for an average of 23 years (ranging from 12 to 30 years). One had completed doctoral level education, one had completed a first degree or equivalent, and one had completed a Certificate of Higher Education or equivalent. A detailed description of the researcher's interpretation of *professional professionalism* is included in table 7.7.

Table 7.7

Long-form narrative description of factor 4: professional professionalism.

Professional professionalism

Professionalism relates largely to a collection of outcomes that must be portrayed both personally and by the larger profession. The ways in which these outcomes are achieved are less important as long as the end result remains the same, which ultimately ensures compliance with legal requirements and that those external to the profession view it positively.

There are two crucial outcomes to be delivered through professionalism, the most important being that the professional is seen to deliver congruence in their conduct. What an observer sees outwardly should also appear to be what is on the inside, as actions are governed by a clear set of moral values (9: 12; 32: 12)* that mean they are fully honest at all times (11: 2). The second most important aspect of professionalism is that an individual pays close attention to their competence, taking an open and honest stance towards their own mistakes (20: 11) and activities that lie outside their current level of skill (34: 10). Professionalism requires a rounded and balanced skill set (31: 11) and developing such competence is an ongoing process throughout a career (30: 11).

Following this, the means of achieving the above outcomes start to become relevant. For example, it is important that professionals operate according to a strict code of respect for others, which is somewhat derivative of the most important characteristic of congruence. Living by the values of dignity and respect for all (8: 10) will ensure that professionals can place their personal feelings aside (23: 10) to build mutually respectful relationships with customers, colleagues, and stakeholders (1: 9; 27: 9). This ensures that all parties are seen to be equal in transactions, and that professions cannot therefore be criticised for granting any implied right of superiority to its members (16: 1).

Professionals are also subject to a system of strict accountability, which relates to their continuing professional development. Professionals operate within a management hierarchy (18: 1), and their actions are governed by strict policy and procedural guidance at all times (35: 9). As a result, characteristics required of occupations that are less closely governed, such as applying knowledge (33: 2) creatively to generate ideas and then sell them to others (14: 2) are unimportant. Similarly, flexibility (39: 4) and the ability to independently take the initiative (26: 5) are also rendered relatively less important to professionalism. Professionals must accept that they will be held accountable for their workplace actions (42: 8) and to a lesser extent, the character displayed within their personal lives also (22: 9; 29: 8). However, it is perfectly acceptable for them to be unhappy about this (15: 3; 25: 3). Accountability is unavoidable (42: 8), but as professionals embody the characteristics of congruence and an ongoing approach to learning, they should be able to defensively refute the challenge it brings effectively (25: 3).

Generally speaking, colleague relationships are not crucial to professionalism either way. Overall, it is good for colleagues to be collaborative in their endeavours (5: 8), although competition is not a bad thing should it arise (40: 4). There is also some value in demonstrating loyalty to colleagues (12: 6) and supporting their development (38: 7), although neither of these things are particularly defining of professionalism. Similarly, it

might be considered good by some for individuals to care for their own wellbeing (37: 6), and by others to sacrifice this for the needs of the customer (28: 7), but these are actually relatively neutral issues in comparison with those more important characteristics of professionalism listed above.

Professionalism is a collection of characteristics that does not always form a coherent whole as context plays a huge role. As such, there are some behaviours that are acceptable only under specific circumstances. Some of these behaviours may be perceived as bordering on unacceptable if externally observed by those without an understanding of the context however, so it is important that professionals learn to appreciate what is appropriate when. For example, professionals are required to deliver the utmost respect for customers and their wishes (8: 10), but there may be times when customer needs contravene their wishes. In such circumstances, it may be more acceptable for professionals to disregard the customer's wishes and act instead in their best interests, although this should not be taken as a general rule (36: 7). In addition, professionalism generally requires that individuals separate their personal lives and feelings from the professional context (23: 10). However, there are some circumstances where an individual can express their feelings towards a situation without becoming unprofessional, such as where they are being challenged or their actions scrutinised (15: 3; 25: 3). Although technically objectivity is preferred within the workplace, emotional expressions demonstrate the commitment and passion a professional holds for their work so this is not strictly necessary for professionalism (41: 5). Similarly, although it is important that professionals are seen to be respectful to all (1: 9; 8: 10; 27: 9), it is acknowledged that there are contexts when certain types of humour may be more accepted, such as at less formal or social events (13: 4). In addition, although colleagues must provide a united and collaborative front outwardly (5: 8), a competitive streak inwardly is not problematic as long as it does not adversely affect competence (40: 4), even if it results in less than positive relationships between colleagues (10: 3). Finally, although broadly speaking a code of dress and appearance should be followed (7: 4), there may be certain very limited circumstances that make casual attire more acceptable (3: 5). Ultimately, although not the most important aspect of individual professionalism, the ways that a professional is perceived by those around them is important in so far as it forms part of the accountability previously discussed (24: 8).

Overall professionalism is a very complex issue, borne out of the changing nature of professionalism. Professionals are required to serve the dual masters of what their profession requires of them today, versus the established working culture that has been stable and productive throughout the history of the profession, which may not always amount to the same thing. As a result, many behaviours can be acceptable or unacceptable depending on their expression and context. Although the key priorities mentioned at the outset are non-negotiable as they ensure that the profession and its members remain in good standing with regulators and the general public, it must be acknowledged that they are at times extremely difficult to deliver. It must therefore be acceptable for individuals to act in ways that are less than perfectly in line with the expectations of regulators or the general public without this undermining their professionalism.

Note. This table provides a long-form narrative description of the viewpoint associated with factor 4, according to the best practice procedure described by Watts and Stenner (2012). This factor represents a composite amalgamation of all those participant viewpoints defining it and so the description is written from the perspective of a hypothetical participant with a perfect loading to factor 4.

*Numbers in brackets denote the statement number and position within the factor array that different aspects of the factor description relate to, with columns numbered 1-12 from left to right. As such, (9: 12) indicates that statement number 9 was within column number 12, as shown within figure 7.6.

Figure 7.6. Factor array for professional professionalism.

For a person to be professional, they must...

Most disagree			Neutral/undecided						Most agree		
1	2	3	4	5	6	7	8	9	10	11	12
18 Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	33 Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	10 Recognise that having positive relationships with colleagues is part of being good at your job	40 Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	4 Be able to express themselves clearly, and in a focused and articulate way	37 Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	28 Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	24 Be perceived as demonstrating the required professional standards by those around them	27 Build rapport with customers, in a way that encourages mutual respect and partnership	23 Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	20 Acknowledge, accept, and learn from their own mistakes	32 Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved
16 View themselves and other professionals as appointed leaders in society	11 Not be dishonest, but it is OK if there are times when they do not tell the whole story	15 Not be resentful of work-related scrutiny and challenge	7 Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	41 Be able to manage stressful situations with objectivity rather than emotion	19 Take pride in their profession, and defend it to others	38 Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	29 Behave responsibly outside of the workplace, including being a positive role model for society	22 Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	8 Live by values such as dignity and respect, compassion, and unconditional positive regard for everyone	31 Have a well-rounded set of skills – they must demonstrate both technical expertise and interpersonal skills	9 Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only
	14 Be appropriately assertive and able to influence decisions	25 Not be defensive or easily upset in response to workplace events or change, or when being challenged	13 Have a keen awareness of where humour can become inappropriate	3 Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	6 Not come across as uncaring or callous	21 Embody characteristics such as honour, duty, justice, and courage	42 Welcome a sense that they are accountable to others	1 Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	34 Be aware of the limits of their own competence, inform others about them, and remain within them	30 Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	
			39 Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	26 Be able to work independently and confidently according to their own initiative	12 Be someone that demonstrates loyalty, someone that a colleague can rely on	36 Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	5 Work collaboratively with colleagues, rather than competitively	35 Demonstrate a strict adherence to policy and procedural guidance at all times			
					2 Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang	17 Demonstrate a commitment to prioritising the underserved and those with diverse or unmet needs					

Note. For further detail regarding the preparation and use of factor arrays, please see chapter 4, section 4.2.3.5.1.

7.4 Discussion

Defining professionalism is a challenge generating much controversy but despite a proliferation of interest, no consensus has yet been identified, hampering research efforts to support the effective development and measurement of professionalism (Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; see also the findings of chapter 5). The present study took a novel approach to understanding professionalism from subjective viewpoints using Q methodology to identify shared subjectivity that may be used to form the basis of a generalised definition of the construct. By-person factor analysis of 34 completed Q sorts suggested a four-factor solution, identifying four major viewpoints amongst participants as to important characteristics of professionalism. This evidence suggests that professionalism may be defined in a pluralist way by the participants of this study, which has implications for future research when taken in the context of major themes from the extant literature pertaining to the professionalism debate.

7.4.1 Comparing results with themes from the existing literature.

7.4.1.1 *Being a good person.*

The data from this study suggest that professionalism requires an individual to be a good person, although there was significant divergence as to what this meant in real terms. All four factors made reference to individual character, but the most desirable aspects of this differed. *Intellectual* professionals described this characteristic as involving congruence between internal thoughts and feelings and external manifestations thereof, with morality and fairness being strong guiding principles. *Collaborative* professionals felt that a collaborative nature was required, based on keen self-awareness and understanding of the social consequences of behaviour. *Personal* professionals also sought congruence, although related specifically to a collection of somewhat ideological qualities, including a sense of duty, honour, and justice. *Professional* professionals also required that individuals demonstrate good character but introduced the caveat that the acceptability of outward manifestations of this were defined by the context and audience involved.

Previous attempts to define professionalism align with all of the factors to some extent, with concepts such as ethics, respect, empathy, having a service orientation, altruism, integrity, and a caring nature all featuring frequently in existing literature and forming the strongest and most consistent theme throughout (e.g. Ben-David et al., 2004; Bertolami, 2004; Bonke, 2006; Buck et al., 2015; Carey & Ness, 2001; Frohna, 2006; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Kim et al., 2004; Monrouxe & Rees, 2012; Swick, 2000; Thomas et al., 2007; Treviño et al., 2006). However, the interpretation of *professional professionalism* moves away from this conception somewhat. Previous literature does generally recognise the role of context in professionalism, but this tends to be in relation to behavioural expressions of the attributes described above (Carter et al., 2015; Finn et al., 2010; Fochtman, 2006; Goldie, 2013; Kuczewski, 2006). For example, context may alter how one communicates their empathy, but not the fundamental adherence to its importance as a professional quality. Participants who subscribed to *professional professionalism* instead alluded to the guiding principle being the potential audience of behaviour rather than a stable set of positive internal attributes, and that behaviour could therefore be altered according to the audience present. Displaying such changeable conduct may cause difficulties when a *professional* professional interacts with colleagues or stakeholders of a different viewpoint, particularly those who expect more consistent commitment to the aforementioned attributes.

7.4.1.2 Accountability.

Accountability was another strong theme within the data, which is also echoed in previous literature. Accountability is cited as of central importance to professionalism within the academic and professional regulatory literature alike, with growing emphasis in the latter (Carey & Ness, 2001; Evetts, 2003; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski, 2014; Swick, 2000), but the present study found that this concept may require further unpicking.

The factors identified in this study suggest that individuals may differ in their views of accountability and specifically how it relates to balancing compliance and independence. For example, both *intellectual professionalism* and *personal professionalism* express the view that accountability does not speak to blind compliance, but rather that individuals should exercise their independence to challenge policy and procedural guidance where appropriate. Conversely, *professional professionalism* expresses the opposing view that compliance is of supreme importance in accountability and that independent action is therefore rendered unimportant. Finally, *collaborative* professionals took a view more similar to the previous literature in that accountability formed part of the baseline expectations of any mature adult in the workplace and was therefore a necessary but not sufficient factor in professionalism.

These different and sometimes opposing views of professionalism could lead to individuals being perceived as more or less professional by stakeholders depending on the observer's viewpoint, despite their adhering faithfully to what they feel is the most professional behaviour relating to accountability. For example, professional guidelines from the General Dental Council published in 2013 require that with regards to patient safety, dental professionals are fully compliant with policy and procedure at all times. Although this is compatible with the *professional* professional's concept of professionalism, it may actually conflict with the behaviour of *intellectual* and *personal* professionals who interpret accountability as less related to compliance and more about the right circumstances to take responsibility for challenging guidance.

7.4.1.3 Effectiveness.

The requirement to demonstrate effectiveness within a role was a theme common to *intellectual*, *collaborative*, and *personal professionalism*, with only those participants identifying significantly with *professional professionalism* failing to endorse it as a major pre-requisite for professionalism. This is congruent with the existing literature, as the ability to communicate effectively and deliver excellence feature prominently in previous attempts to define professionalism (Carter et al., 2015; Gleeson, 2007; Hafferty, 2006; Hammer, 2006; Hendelman & Byszewski,

2014; McNeill, 2001; Swick, 2000). However, the absence of effectiveness in *professional professionalism* may also be unsurprising when taking the debate in a historical context.

The traditional concept of professionalism states that assessing the effectiveness or competence of a professional requires such specialist knowledge that individuals external to a profession would be unable to do so (Evans, 2008; Inui et al., 2006). Given that the viewpoint of *professional* professionals suggests that professional conduct is something contextualised by the audience viewing it, it is perhaps unsurprising that effectiveness does not feature, given that the audience may not be assumed to have the level of knowledge required to make such judgements. Indeed, some commentators whose assertions align with the more traditional view of professionalism, including authors published at a time when this conceptualisation was dominant and students of medicine whose training remains firmly adherent to the traditional view as the dominant paradigm, do show tendencies to omit effectiveness or excellence in their conceptualisation of professionalism (e.g. Carey & Ness, 2001; du Toit, 1995; Finn et al., 2010; Monrouxe & Rees, 2012; Shirley & Padgett, 2006; Stern, 2000). Similarly, within regulatory guidelines associated with the older and more traditional professions such as medicine and dentistry, effectiveness is generally referred to as a matter of competence, but this competence is left unquestioned where formal training or a qualification is already in place (General Dental Council, 2013; General Medical Council, 2013, 2016). Overall, the new understanding of professionalism provided by the results of the present study suggests that despite its apparent omission within *professional professionalism*, effectiveness may not simply be deemed unimportant, but rather relevant only to issues of regulation and qualification, rather than professionalism.

7.4.1.4 Self-management and continuing professional development.

Within the academic literature, attempts to define professionalism commonly include some element of self-management or development, encompassing the concept of self-improvement through ongoing professional learning and development (Ben-David et al., 2004; Doukas, 2006; Gleeson, 2007; General Dental Council, 2013; General Medical Council, 2013, 2016; Hafferty, 2006; Hammer,

2006; Stern, 2000; Swick, 2000). Although a strong theme within the literature and particularly within guidance issued by professional regulatory bodies (e.g. General Dental Council, 2013; General Medical Council, 2013, 2016), this theme was less pervasive within the data of the present study.

While *collaborative* and *personal professionalism* felt that ongoing learning was a key feature of professionalism, it was less important within *intellectual professionalism* and not mentioned at all with any significance in *professional professionalism*. This latter point is interesting because, as discussed above, *professional professionalism* holds some key features in common with the traditional concept of professionalism, the influence of which remains evident in the guidelines of regulatory bodies of older professions such as medicine and dentistry even today. Consequently, it is interesting that the two conceptualisations part company over this key issue. One reason for this divergence may be that times have changed in professional practice (Evans, 2008; Evetts, 2003, 2014), as suggested by comments associated with *professional professionalism's* defining Q sorts, and so individuals subscribing to this view may be more sharply aware of the intricacies of the issue, specifically in its relationship to self-regulation.

The traditional view of professionalism is that professions are self-regulating, setting their own standards for conduct and performance, entry and disciplinary action (Evans, 2008; Lewis, 2006). However, the concept of self-regulating professions has since become obsolete as a result of increased governmental oversight of public spending and services, but also in response to criticism claiming that self-regulation was nothing more than a means of professionals securing and maintaining power and social privilege (Castellani & Hafferty, 2006; Evans, 2008; Evans, 2008; Shirley & Padgett, 2006). Although continuing professional development remains prevalent in professional guidelines today at the individual level, its apparent relative unimportance in *professional professionalism* may reflect an awareness of the social undesirability of the broader, profession-level concept of self-regulation and policing. In order to project a more socially desirable concept of professionalism, *professional* professionals may have actively suppressed the importance of this issue within their viewpoint to prevent potential interpretations that it relates to profession-level self-regulation rather than individual-level self-management.

7.4.2 Points of divergence from the existing literature.

7.4.2.1 Colleague relationships.

Despite the significant overlap described above, some key themes in the data of the present study are not echoed within the literature or take a much less central role. For example, colleague relationships featured heavily within all four factors amongst this participant group. While teamwork does feature in previous attempts to define professionalism (e.g. Ben-David et al., 2004; General Dental Council, 2013; General Medical Council, 2013, 2016), it is with lower frequency than it arose within the comments of participants and a general acceptance that it is a straightforward concept. The different factors in the present study conceived of colleague relationships in a more complex way. While *collaborative* professionals felt that positive colleague relationships delivered productivity, *personal* professionals felt that positive relationships were not necessary for productivity. *Intellectual* professionals felt that colleague relationships should be viewed objectively, and so positive or negative valence was of little importance, while *professional* professionals felt that although positive colleague relationships were generally accepted as a good thing, they did not form part of professionalism and so were not relevant to the debate.

Colleague relationships were a key theme amongst the participants of this study even though the importance and meaning they attached to them differed. This suggests that this aspect of perceived professionalism may require further elaboration within both the academic literature and professional regulatory guidance, especially in light of recommendations to enhance team working based on the Mid Staffordshire NHS Foundation Trust Public Inquiry findings (2013).

7.4.2.2 The remit of professionalism.

A second point of contention between the findings of this study and previous attempts to define professionalism lies in reference to its boundaries, and what professionalism should and should not include. This P set expressed firm and strongly held views regarding the relevance of individual behaviour outside of the workplace. *Collaborative professionalism*, which was the factor accounting for the

largest proportion of variance within this P set, asserted that conduct outside of the workplace was irrelevant to professionalism, with its inclusion provoking strong reactions and being deemed to contravene basic fairness and the right to a personal life. *Professional* and *personal professionalism*, however, felt that this issue was unavoidable and had to be accepted but with a caveat that no individual should be required to accept this happily, and so allowing individuals to react negatively towards it without any loss of professionalism.

The debate regarding the limits of professionalism does not feature commonly within the existing literature but where it does, it generally relates to the viewpoints of students undergoing vocational training in areas where professional regulation is required, such as medicine or dentistry. It should be noted, however, that such discussions are increasing as social media becomes a more prominent issue for professionals, educators, and regulators (Finn et al., 2010; Lie, Trial, Schaff, Wallace & Elliott, 2013; Williams, Field & James, 2011). This is interesting as the idea of taking responsibility for non-work behaviour does feature in professional regulatory guidelines (e.g. General Medical Council, 2013, 2016), despite both stakeholders and students appearing to agree that it is less than desirable. This would suggest that further exploration is required to understand why there is such disparity between professional, academic, and stakeholder viewpoints regarding this issue.

7.4.2.3 The role of sector of employment.

It was observed that there was no strong pattern between factor loadings and the employment sector of participants in the present study. As can be seen in table 7.3, amongst Q sorts loading significantly onto the same factor, there was no strong representation from any single profession or group of similar professions. This is somewhat surprising because, as noted earlier, much of the extant literature relating to professionalism is based on a historical belief that it may only be understood by members of the same profession (Evans, 2008; Inui et al., 2006). Subsequent research has propagated this, exploring professionalism largely within defined occupational areas based on the rationale that it can only be judged meaningfully within the context of, and by members of, its own profession (e.g. Jha et al., 2007; Li et al., 2017; Veloski et al., 2005). It may therefore have been reasonable to expect

that the viewpoints of members of the same profession would have clustered in terms of similarity. The fact that this was not found within the present study challenges the assumption made within the literature to date, suggesting that viewpoints as to what professionalism ought to look like and the personal attributes underpinning it supersede professional boundaries.

Having said this, the two views that professionalism may be profession or non-profession specific are not irreconcilable and may actually be complementary when interpreted in the context of existing psychological theory. The role of situational context is much discussed relating to professionalism as a means of explaining why the professional conduct of a single individual may vary across situations (Evans, 2008; Finn et al., 2010; Fochtmann, 2006; Goldie, 2013; Monrouxe et al., 2011). The cognitive affective personality system theory suggests that such behavioural variation reflects not only the external, situational context of behaviour, but variation in variables specific to the individual also (Mischel, 1969, 1973, 1977, 1979; Mischel & Shoda, 1995; Shoda et al., 1994). Placing professionalism in this context means that although individuals may subscribe to a relatively stable view of what is important to it, behavioural variation may result from the interaction of this viewpoint with factors including the context provided by the individual's profession and professional training. In this sense, although an individual's core concept of professionalism may not be profession-specific, their interpretation of it in practice will be influenced by the context provided by their professional training and occupation, and the culture within their employing organisation.

This is evident in research collecting student views regarding professionalism during medical training, as themes in their interpretation of the construct often relate closely to the clinical/medical contexts within which they spend their time (e.g. Finn et al., 2010; Karnieli-Miller et al., 2010; Monrouxe et al., 2011; Monrouxe & Rees, 2012). While individuals may subscribe to non-specific views of professionalism in the abstract, the ways that they would expect their professional values to be expressed practically in day to day life may be more profession specific. As a result, viewpoints may not actually be expected to cluster by participant employment sector, although the behavioural manifestations of them may well do so.

7.4.3 Study evaluation.

Methodologically speaking, Q studies rely on the degree to which the Q set represents the full breadth of the discourse related to the chosen topic (Brown, 1993; McKeown & Thomas, 2013; van Exel & De Graaf, 2005). Professionalism is an extremely broad issue with relevance across a wide range of arenas and stakeholders (Altirkawi, 2014; Arndt et al., 2019; Black et al., 2019; Buck et al., 2015; Chisholm et al., 2006; Evans, 2008; Mendonça et al., 2016; Goldie, 2013). However, documented discussion relating to professionalism has largely been contained within a limited number of fields, the most prominent of which being medical education. The present study sought to balance representativeness by supplementing statements drawn from this field with comments from participants of a separate, general population study and an online media search, as recommended in the use of Q methodology (Brown, 1996). The previous study from which statements were drawn used the repertory grid technique, which explores individual construing with minimal researcher influence and is therefore recommended for use as a basis for studies exploring subjective viewpoints in more detail (Jankowicz, 2004). However, it was likely that the medical literature would still have had a disproportionate impact upon the final Q set, simply due to the comparative availability of related statements. Indeed, evidence from a recent methodological review of measures of professionalism across all occupational sectors (for full details, see chapter 5) found that almost 70% of the articles reviewed referred solely to medical contexts, with the majority of the remainder relating to pharmacy. Professionalism within medical education has a distinctly clinical emphasis and so some statements may be of lower relevance to professionalism in other sectors. For example, statements relating to having a service orientation that prioritises the underserved within society will be of higher intuitive relevance to a medic in clinical practice than a salesperson. The potential impact of a disproportionately high influence from medical literature on the statements sorted is that rather than expressing viewpoints regarding professionalism as a general topic, participants may actually have provided views on a topic more closely resembling professionalism within medical education specifically. Further study may seek to replicate the findings of this study using a different Q set not based on medical education literature, to assess the similarity of the two factor structures.

Another issue pertaining to the Q set lies in the decision to focus on positive or desirable attributes of professionalism (e.g. “welcome a sense that they are accountable to others.”), rather than negative or undesirable attributes associated with unprofessionalism (e.g. “refuses to accept responsibility for their own actions”). Although not exclusively so, many participants reported that they tended to agree with the majority statements presented to them, with the discriminating factor instead being how much they agreed with them. It may therefore be prudent to consider the potential impact of undertaking a Q sort using negatively worded statements targeting unprofessionalism rather than professionalism. There is evidence to suggest that when individuals evaluate the professionalism of others in the real world, the criteria used may be more concerned with a lack of unprofessional behaviour than the presence of professional conduct (for full details, see chapter 6). Consequently, a different factor structure may be elicited where viewpoints as to unprofessionalism are targeted and this structure may actually have greater relevance for appraisals of professionalism. Such findings could hold additional insight as to the potential improvement of professionalism through teaching and learning.

Q methodology rests upon the assumption that the P set is purposively targeted to represent individuals whose viewpoints have particular relevance to the topic being explored. The purpose of the present study was to explore the views of a range of individuals from the general population, so as to explore shared subjectivity regarding the most general conception of professionalism. However, the P set contributing to this study was limited in several areas. More than half of participants were aged 45 years or under, meaning that there was a particularly high representation from early to mid-career phases. A P set with greater representation from older workers who had spent longer in the workplace may result in a different factor structure. Similarly, a number of occupational sectors were more heavily represented among the P set. Participant recruitment for interviews took place at a UK university within both the schools of psychology and dentistry and as a result, these sectors were well represented. An implication of this was that the P set was relatively well-educated. The most recent UK census reports the proportion of individuals educated to higher education level in England and Wales at around 27% (Office for National Statistics, 2011). However, in the present study, more than 80% of the P set was educated beyond compulsory education or equivalent, with 44%

educated to post graduate level. Individuals working in administrative and management/leadership roles were also well-represented. There is significant potential for overlap between the responsibilities of roles in these areas and so it may be prudent to consider their influence in combination. When combined, the P set becomes heavily weighted toward management and administration, with a third of participants self-categorising as working in this area. When considered in the context of the specific aim of this study to explore professionalism as a general construct, these limitations may have influenced the results observed.

The final factor structure within this study included one factor with only three defining Q sorts. It is possible that adding additional Q sorts may have altered the content of the four-factor structure or highlighted further potential factors that have been missed as a result of insufficient representation amongst the P set. Indeed, fourteen of the thirty-four Q sorts did not load significantly onto any factors and so it is possible that there are additional viewpoints unrepresented within the present study. Taking this and the above limitations in combination, future studies might remedy this this by using larger Q and P sets to increase the representativeness of both, and permit exploration of more complex factor structures.

A number of other considerations concern the nature of Q methodology itself and are therefore not necessarily limitations of the present study, but qualifications placed upon how its findings may be appropriately interpreted. Firstly, Q methodology does not seek to provide generalisable findings or indicate the prevalence of viewpoints within a more general population (Stainton Rogers, 1995; Stephenson, 1952; Thomas & Baas, 1992; van Exel & De Graaf, 2005). The results of this study speak to nothing more than the major viewpoints held by the P set engaged. These viewpoints may also be held by a proportion of individuals within a more general population, but this inference may not be made from the current data. Further study employing an R methodology approach would be required before such generalisations could be considered.

A further consideration concerns the interpreted meaning of the viewpoints identified. As with all factor analytic approaches, the present study relied upon researcher interpretation to assign meaning to extracted factors. It is therefore possible that other research teams would interpret the same data in different ways.

The factors described within this report should be considered not as pure accounts of the subjectivity of participants, but as that subjectivity described in a way shaped by the subjectivity of observers.

7.4.4 Implications and future directions.

Attempts to define professionalism to date have generally consulted a range of individuals, whether relatively few in individual research studies or many in regulatory body published guideline development, and aggregated the views of those consulted into a norm-referenced concept of professionalism. The findings of the present study suggest that this approach may be inappropriate in the case of professionalism, as the subjective viewpoints of different individuals may actively contradict one another. This means that in meeting the requirements of one person's viewpoint, another person's requirements may be automatically contravened. Taking the example of the General Dental Council's (2013) professional guidelines, the first requirement of the first principle therein requires that dentists listen to their patients and work alongside them collaboratively to agree a course of treatment. Although this would be acceptable to *collaborative*, *personal*, and *professional* professionals, a patient subscribing to *intellectual professionalism* would disagree, instead preferring that dentists lead the conversation based on their superior knowledge in the area. A shift in emphasis may therefore be indicated that moves away from professionalism as a detached construct contextualised by situation, and instead seeks to understand it as a dynamic social process contextualised by individuals and interpersonal exchanges rather than the objective features of the scenario.

In terms of future directions, if practical use is to be made of these findings in professionalism related teaching and learning, additional research is required. Although generalisations cannot be made from the findings discussed in this chapter, they may be used as a basis to develop an interpersonal theory of professionalism from which to derive testable hypotheses as to the structure of perceived professionalism within the general population. Larger scale R methodology studies to test these hypotheses could then be undertaken to provide generalisable findings. This would enable broader understanding of professionalism amongst stakeholders,

and thus support educators and employers to design interventions to enhance perceived professionalism accordingly.

7.5 Chapter Summary and Conclusions

This chapter describes an exploration of professionalism using Q methodology. The data presented are the first to explore professionalism as a matter of subjective personal opinion and reveal areas where such subjectivity may be shared amongst multiple individuals. The reported results suggest that the changing face of professionalism should be defined in a dynamic way that emphasises the role of perception. Findings suggest that different individuals hold divergent views as to the nature of professionalism and that in some areas these views actively contradict one another. Consequently, extant attempts to define professionalism in norm-referenced or aggregated ways may have limited utility in attempting to measure and improve professionalism. Instead, future endeavours should seek to retain the multiplicity of viewpoints when defining professionalism, as a means of creating interventions and measures that reflect the interpersonal nature of the construct. The results of the current study further suggest that individuals may have a core concept of professionalism that transcends occupational boundaries, although this concept may be permeable enough to allow the flexible application of that core concept in different contexts. This suggests that traditional thinking that confines professionalism to single occupations or professional roles may be out of step with the perceptions of individuals in the real world. As such, employers, educators, and regulators may benefit from shifting the concept of professionalism away from one applying it only one profession at a time, and towards the possibility of a more general, underlying professionalism from which role-specific performance is subsequently derived.

Chapter 8: A Stakeholder Consultation Regarding the Model of Interpersonal Professionalism using the Nominal Group Technique

Abstract

The exploration and measurement of professionalism are growing priorities for employers and educators alike, but research efforts have been limited by a reported lack of theory defining the construct and its underlying mechanisms that has impeded progress in measuring professionalism as a psychological attribute. This study introduced a proposed theoretical model of interpersonal professionalism and invited relevant stakeholders to provide feedback regarding its content. Trainee and experienced occupational psychologists were invited to take part in this initial stakeholder consultation as likely future users of the theory in applied settings and individuals with expertise in occupational psychometry. Four groups of occupational psychologists were consulted using the nominal group technique. Consensus was established by summing votes provided by the participants of each group regarding the priorities perceived as most crucial to the development of the model. The content of all feedback ideas was also subjected to thematic analysis using the constant comparative technique. Data suggest that consensus was low among the stakeholders consulted, such that no clear priorities were identified. Thematic analysis suggested that feedback clustered around five key themes: factors important to professionalism, what is professionalism?, evaluation of the model, professionalism development, and measurement. Thematic analysis also provided insights regarding the use of the nominal group technique in the context of professionalism. Findings are discussed in terms of implications for the future development of the model of interpersonal professionalism and with regards to relevant literature.

8.1 Introduction

Interest has been growing in professionalism across an increasing number of fields for some years (Alcolta et al., 2016; Blake & Gutierrez, 2011; Carter et al., 2015; Evans, 2008). Once limited to a small number of older professions, professionalism has become a focus for newer professions including pharmacy, teaching, and nursing (Chisholm et al., 2006; Evans, 2008; de Mendonça et al., 2016). This growth in attention stems partially from several high-profile cases citing failures in professionalism and has resulted in increased pressure on educators to assure the quality of graduates in this respect (Evetts, 2006; Goldie, 2013; Marei et al., 2018; Mid Staffordshire NHS Foundation Trust Public Enquiry, 2013; Papadakis et al., 2004; Trathen & Gallagher, 2009; Zijlstra-Shaw et al., 2012).

Responding to this pressure requires that educators can effectively measure professionalism in ways that enable benchmarking to support decisions regarding qualification and progression (Ben-David et al., 2004; Bonke, 2006; Lynch et al., 2004; van Mook et al., 2009). The evidence discussed in chapter 5 suggests that this has not yet been achieved, fundamentally because of a lack of a theoretically grounded definition of the construct to be measured. Before work may begin to develop a psychometrically sound measure of professionalism, a theory of professionalism is required. This chapter proposes a theoretical model of professionalism based on existing psychological theory and empirical data reported in chapters 6 and 7 and reports an initial stakeholder consultation to support the theory-building process.

8.1.1 The model of interpersonal professionalism.

The data reported in chapters 6 and 7 provide a framework for understanding the expectations individuals have of professionals and the criteria on which they base decisions regarding the level of professionalism displayed by others. This data was explored in relation to existing psychological theory explaining the processing of incoming information and resultant production of behaviour within the context of interpersonal exchanges. The model of interpersonal professionalism suggests that professionalism is a dynamic, interpersonal phenomenon that should be

contextualised in ways that accounts for intra-individual behavioural variation across similar circumstances and intra-individual behavioural stability resulting in different outcomes. Specifically, it ought to explain both why individuals might be perceived as demonstrating excellence in professionalism during one interaction but fall short thereof during similar sessions at other times, and also why individuals previously perceived as demonstrating professionalism appear unable to adapt their behaviour appropriately for success in different situations. A model is presented below that meets both of these requirements by combining the data reported in chapters 6 and 7, the cognitive affective personality system theory, and theories of emotional intelligence.

8.1.1.1 The cognitive affective personality system theory and theories of emotional intelligence.

For a full discussion of the cognitive affective personality system (CAPS) theory and theories of emotional intelligence, see sections 1.3.3.2 and 1.3.3.3 of chapter 1. In summary, the CAPS theory accounts for behaviour within a theory of personality (Mischel & Shoda, 1999). Drawing on Bandura's social learning and cognitive theories of behaviour (Bandura, 1991, 1999a, 1999b; Mischel, 1979; Mischel & Shoda, 1998), trait theories of personality such as the big five (Mischel & Shoda, 1998 Goldberg, 1992), and constructionist theories of individual construing (Kelly, 1955; Walker & Winter, 2007), the CAPS theory describes the perception of a stable and enduring personality as the product of change-resistant cognitive processing biases associated with idiographic cognitive-affective units (CAUs). These CAUs are encoding, competencies and self-regulation, expectancies, values/goals/motivations, and affect (Mischel, 1973; Mischel & Shoda, 1999; Shoda et al., 1994). Emotional intelligence relates to individual abilities associated with processing emotional information and using it to inform reasoning (Brackett et al., 2011; Mayer & Salovey, 1993, 1995; Mayer et al., 2016; Mayer et al., 1990; Mayer et al., 2004b, 2008; Mayer et al., 2011; Salovey & Grewal, 2005). In the context of the CAPS theory, emotional intelligence is considered to constitute a specific set of competencies. Individual CAUs, including emotional intelligence competencies, dynamically interact with one another to guide the processing of incoming

information about situations in the external world, activating or inhibiting behaviours based on the characteristics of the situation in a continuous bidirectional individual-environment interaction (Mischel, 1973; Mischel & Shoda, 1995, 1998, 1999). This complex, ongoing, multi-factor interaction accounts for the intuitive sense of self akin to personality that remains stable across different situations, whilst also allowing for behavioural variation even where situations appear superficially similar.

This theory complements a number of themes within the professionalism literature. Many authors have emphasised the role of context in professionalism (Birden et al., 2014; Carter et al., 2015; Evans, 2008; Fochtmann, 2006; Kelley et al., 2011; Troman, 1996; van Mook et al., 2009; Verkerk et al., 2007), suggesting that no two situations may be judged according to the same criteria. The CAPS theory explains the role of context and the mechanisms by which it results in behavioural variation, also aligning well with assertions that professionalism should be measured at multiple times using multiple metrics to provide an average assessment of performance (Buck et al., 2015; Goldie, 2013; Wilkinson et al., 2009). An account of professionalism grounded in the CAPS theory, however, overcomes the resourcing and logistical implications of such assessment, suggesting that it is the mechanisms underlying behavioural variation that are of interest, rather than constantly sampling their manifestations. Finally, the CAPS theory also accounts for the enduring and stable nature of perceived professionalism. Although behaviour may vary across situations in ways that increase or reduce perceived professionalism in individual cases, the new professionalism literature acknowledges that it is generally perceived to be something more stable or individually intrinsic (Bertolami, 2004; Carter et al., 2015). The CAPS theory explains this as characteristic patterns in individual processing and competencies providing sufficient consistency for others to perceive it as stability of professionalism.

Where situations rely on individual abilities in processing incoming emotional information and using this to self-regulate behaviour in ways that achieve desired interpersonal outcomes, competencies associated with emotional intelligence and their interaction with the other CAUs are particularly pertinent. It is asserted that this is the case in professionalism. The data described in chapters 6 and 7 suggest that different individuals have different views as to what professionalism is and how it might be manifested. Where two opposing views clash, individuals may rely more

upon their emotional intelligence to identify and manage situations to successful conclusions and less on the specific content of their individual concept of professionalism. In this light, the potential implications of emotional intelligence abilities suggest that they would be important inclusions to a theory of interpersonal professionalism based on the CAPS theory.

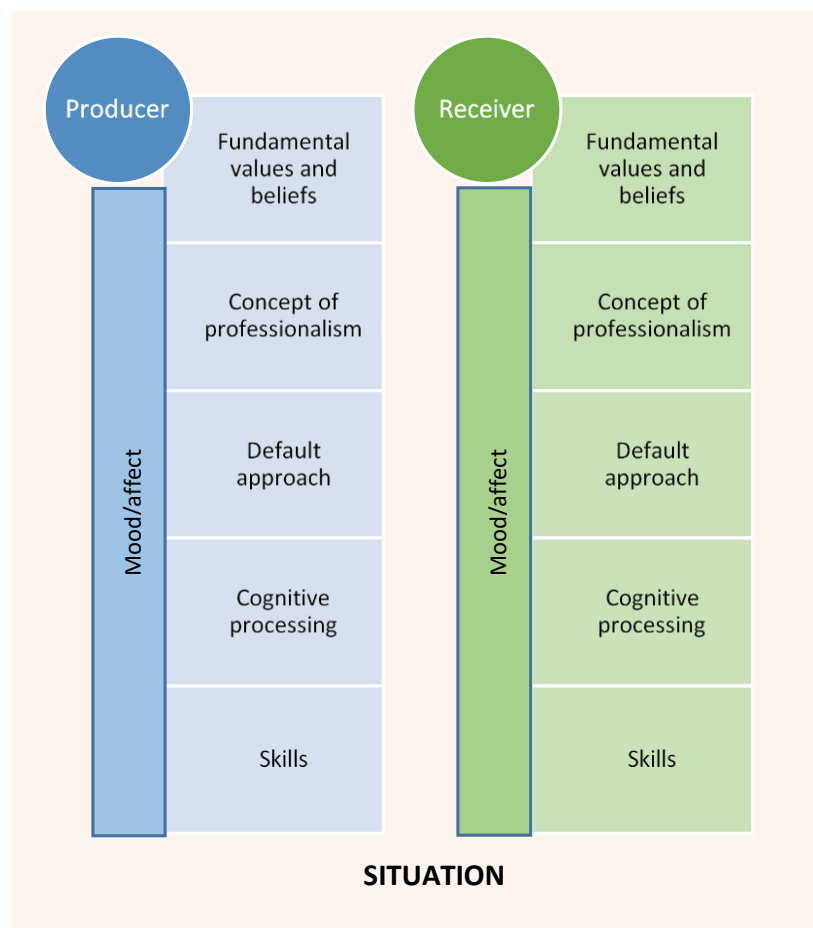
8.1.1.2 An integrated model of interpersonal professionalism.

The proposed model describes professionalism as a perceived construct arising out of a complex interplay between cognitive, behavioural, and affective processes of all individuals involved. In the case of a two-person interaction where the professionalism of one party is of interest, the CAPS and emotional intelligence of both individuals will affect the appraisals made and overall success of the interaction in terms of perceived professionalism. Where both individuals perceive the professionalism of the other, the picture becomes more complicated, and where more than two individuals are present, more complex still. To maximise clarity, the example of a two-person interaction will be used where the professionalism of only one individual is of interest. The two parties will be known as the *producer* of professionalism, the person whose conduct is being evaluated; and the *receiver* of professionalism, who is undertaking that appraisal. This interaction may be most clearly thought of as occurring between a professional such as a doctor or dentist (the producer) and a new patient (the receiver).

Upon entering the interaction, both parties have values and beliefs governing the way that they view the anticipated exchange, their role within it, and their desired outcomes. Each individual has beliefs regarding professionalism and what it should look like; the producer has a concept that they wish to embody, and the receiver has a concept against which they evaluate the producer's conduct. These concepts might be the same or different, as suggested by the data described in chapter 7. Upon entering the exchange, the producer will have little contextual information to guide their approach and so is likely to stick closely to their concept of professionalism as a default behaviour. As the interaction progresses, however, the producer can use emotional intelligence competencies to identify and apply new information portrayed by the receiver. This information may be portrayed verbally or non-verbally and can

be used by the producer to decide whether they are being perceived positively or negatively and as a basis to self-regulate and adapt their behaviour accordingly. Consequently, the producer also relies in part on the emotional intelligence of the receiver, who must be able to convey their emotions and reactions effectively. However, they also rely on the capacity provided by their own cognitive tendencies to adapt flexibly to new information and their emotional reasoning skills to use this information effectively. Taken together, these factors determine whether the producer can manage their perceived professionalism using the abilities to learn from new information and apply it rapidly and effectively to plan, execute, and self-regulate their ongoing behaviour.

Figure 8.1. *Factors affecting the producer and receiver’s interpersonal approaches, in the context of individual affect and the external situation.*



In accordance with the CAPS theory, all of the factors described above operate within additional internal and external context. Internal context relates to the affect or mood of each party, which can vary widely across the course of a single meeting. External context relates to environmental factors, such as the potential consequences of the interaction, how familiar the process is to each individual, and whether anyone else is observing their communications. As these factors interact, both the producer and receiver constantly monitor the interpersonal interaction and behave according to their cognitive processing of it. A visual representation of this is provided in figure 8.1.

The implications of this model for developing, assessing, and measuring professionalism can be considered under a number of stages using the two-person interaction described above, as shown in figure 8.2 with the stages of interaction described in table 8.1.

Figure 8.2. *The stages of a two-person professional encounter between a producer and a receiver.*

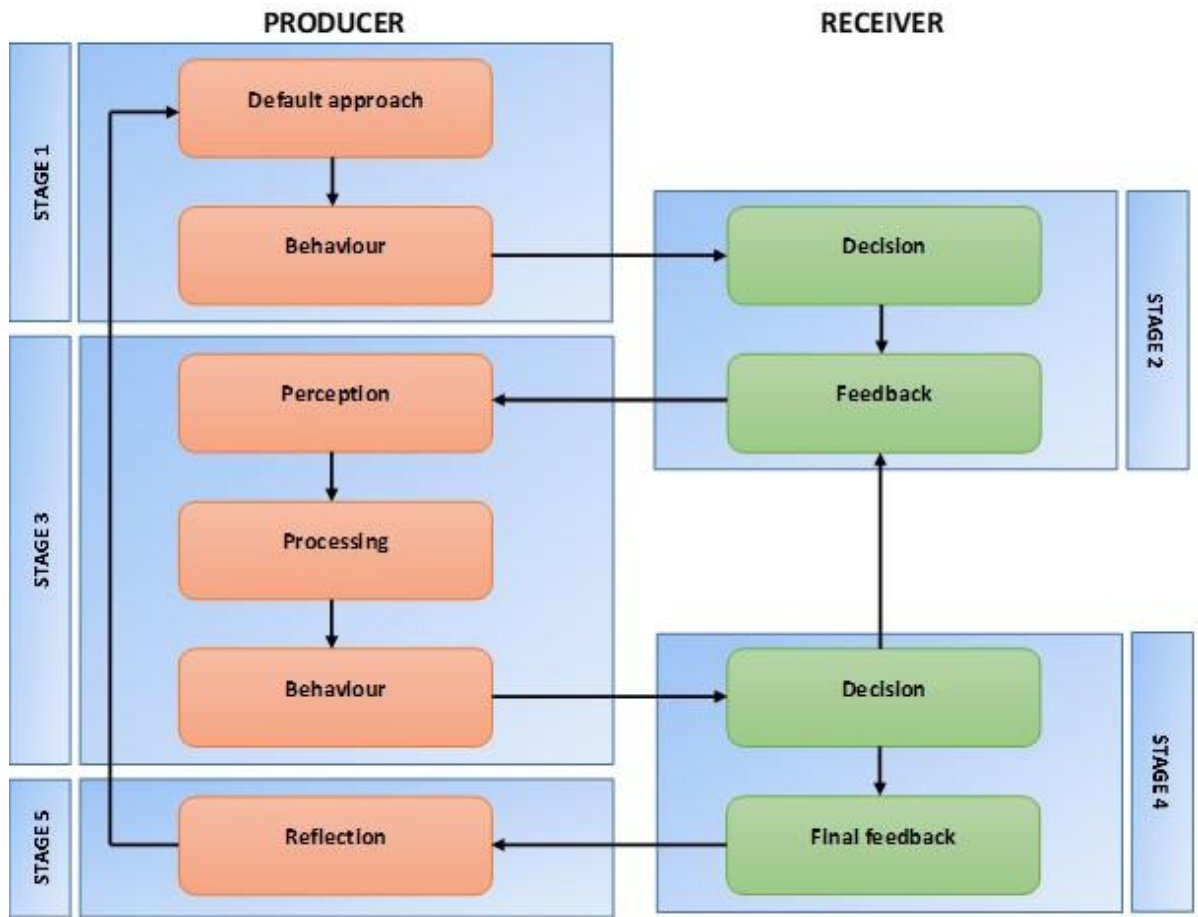


Table 8.1*Description of the stages of a two-person professional encounter between a producer and a receiver.*

Stage	Description	Contextual factors
1. Default approach	<p>The producer enters an unfamiliar situation with the receiver.</p> <p>In the absence of information about the receiver's internal context, the producer sticks closely to their default approach.</p> <p>What that default looks like and how closely the producer sticks to it is affected by the interaction of internal and external contextual factors.</p>	<p>What are the producer's motivations? What outcome holds the most subjective value?</p> <p>How rigid is the producer's concept of professionalism? Do they struggle to stray from it, even if it has been unsuccessful in the past?</p> <p>What are the producer's cognitive tendencies? Are they likely to recall positive outcomes from using their default approach more than negative, and therefore persist with it?</p> <p>Does the producer have the skills to alter their default approach? Do they have the skills to learn from previous experience and regulate future behaviour to make changes?</p> <p>What is the producer's mood like today?</p> <p>How familiar is the situation? Is the encounter something the producer undertakes on a daily basis, even though the other person is unknown to them?</p>
2. Initial decision	<p>The receiver uses their default approach upon entering the interaction, but in different ways to the producer.</p> <p>The initial decision the receiver makes as to whether the producer is professional is a product of interacting contextual factors.</p>	<p>What are the receiver's motivations? What outcome holds the most subjective value?</p> <p>How rigid is the receiver's concept of professionalism? Do they struggle to look favourably on behaviours that contravene it?</p> <p>What are the receiver's cognitive tendencies? What are their expectations based on previous experiences?</p> <p>Does the receiver have the skills to alter their behaviour if they need to, to tell the producer that they are unhappy? Is it important to them to provide this feedback?</p> <p>What is the receiver's mood like today?</p> <p>Is the situation familiar to the receiver? Is the encounter something familiar and comforting, or likely to provoke anxiety?</p>

3. Perception of feedback information	<p>As the encounter progresses, the producer uses their skills to manipulate their behaviour to achieve the desired outcome.</p> <p>The producer's ability and desire to do this is affected by internal and external contextual factors.</p>	<p>How skilled is the receiver in expressing their initial feedback?</p> <p>Does the producer have the skills to accurately identify and use this feedback? Can they rapidly plan and regulate their behaviour successfully based on this new information?</p> <p>How does the producer tend to process feedback cognitively? Do they tend to learn from it or ignore the more negative aspects?</p> <p>Have the producer's motivations changed since entering the encounter? Are they still prioritising the same values or has something about the situation or the receiver changed them?</p> <p>Has the producer's mood changed?</p>
4. Final decision	<p>The receiver continues to make judgements about the producer's professionalism and provide feedback regarding those judgements on an ongoing basis as the encounter continues.</p>	<p>How skilled is the receiver in expressing their feedback? Do they express themselves clearly and are they motivated to do so?</p> <p>Have the receiver's motivations changed since entering the encounter? Are they prioritising the same values or has something about the situation or the producer changed them?</p> <p>Has the receiver's mood changed? If the encounter is not going well, is a shift in their mood interfering with their ability to deliver feedback?</p>
5. Reflection	<p>Upon conclusion of the encounter, the producer has the opportunity to reflect on and learn from it, but the likelihood of their doing so, and the effects of this, are subject to contextual factors, both internal and external.</p>	<p>How expressive and clear was the receiver in their feedback?</p> <p>Does the producer have the skills to accurately identify that feedback on a post hoc basis? Can they learn from it in a way that might change their default approach?</p> <p>What biases might be affecting the producer's post hoc processing? Are the recalled details altered in ways that allow them to view their own behaviour as in line with their underlying values, for example?</p> <p>Now that the encounter has ended, have the producer's motivations changed? For example, is their goal now to justify their own behaviour, learn from the experience and improve, or create a professionally acceptable version of the encounter for formal records or in case of challenge?</p> <p>Has the producer's mood changed since the encounter concluded and is this affecting their ability to accurately recall the details of the encounter?</p>

8.1.2 Research aims.

The present study introduced the proposed model of interpersonal professionalism for preliminary stakeholder consultation using the nominal group technique. This consultation targeted professionals deemed to constitute key potential stakeholders to the model in its future implementation, namely occupational psychologists. Participants were not sought with any significant or specific training in professionalism. The primary reason for this related to the longer-term vision to assess the potential to create a new, psychometrically sound tool to measure professionalism. The findings reported in chapter 7 suggest that professionalism may be perceived similarly by individuals regardless of their sector of employment. The implication of this was that whether individuals receive specific professionalism training, such as in medical or dental education, does not appear to influence their views regarding the construct of professionalism. The aim of the present study was to consult stakeholders with relevant expertise regarding a theoretically defined construct of professionalism within the psychological context. As such, while the views of individuals with specific professionalism training would be relevant during later stages of creating a measure of professionalism, namely to assess its appropriateness and utility for the specific requirements of their profession, it was not deemed to be a crucial factor during this study. However, training in occupational psychology was deemed to be relevant expertise because, as described in chapter 2, the best practice validation approach recommends that experts in psychometric theory are involved at even the earliest stages of developing a psychometric tool.

Consultation was undertaken based on the requirement within pragmatic research to ensure shared understanding of the implications of research through intersubjectivity prior to the wider communication of findings. It was intended to explore understanding and acceptability of the model and its feasibility for future development as a basis of a psychometric tool to measure professionalism. The aims were to gather feedback to inform a decision as to whether a full stakeholder analysis and consultation was warranted at this point, and to inform the materials to be used to undertake a such a consultation should it be suggested, by providing an indication of consensus regarding the priorities to be addressed.

8.2 Methods

8.2.1 Materials.

Participants were presented with a summary form of the model developed by the research team (see appendix G). The research team comprised expertise in psychology and psychometric theory to maximise the integrity of the interpretation and application of existing psychological theory and the empirical data described in chapters 6 and 7. The team also included a representative from a professional practice area with interest in professionalism (dentistry), to assess the model summary for accessibility and clarity.

8.2.2 Participants.

Participants aged 18 years or over with expertise in occupational psychology and psychometry were purposively recruited using the existing professional networks of the research team. Participants received no incentive for taking part in the study. As this study was preliminary in nature, the minimum number of groups likely to reach data saturation was set at three, based on the findings of Guest and colleagues (2016) that this would be sufficient to identify ninety percent of the extant feedback themes and all of the most prevalent ones. A minimum of five participants per group was targeted, based on best practice recommendations (Delbecq et al., 1975; Humphrey-Murto et al., 2017; McMillan et al., 2016), but where this number was not met for logistical reasons, groups still went ahead. This decision was based on the pragmatic aim to gain as much feedback as possible and understanding that the unit of number of groups was more important than number of individuals therein (Carlson & Glenton, 2011).

8.2.3 Procedure.

The study was advertised by email to the existing professional networks of the research team. Participants were invited to take part in focus groups to explore a proposed new theory of professionalism with the opportunity to contribute to its

development by providing comments and feedback, and raising questions and challenge. Before the focus group began, participants were provided with information about the study and asked to provide informed consent to their participation.

Focus groups were facilitated by one of two members of the research team using the nominal group technique. The nominal group technique and the ways that it differs from other focus group approaches is discussed in detail in section 4.2.4.2.4 of chapter 4, but table 8.2 outlines the major phases involved in using the technique. Following brief introductions and statements of purpose, process, and logistics, the facilitator invited participants to silently read the summary of the proposed model while working alone to record all comments and points of feedback occurring to them during or after their reading. Participants were asked not to sanitise their feedback, but to record it as it occurred to them. It was also acknowledged that comments generated may not constitute participants’ personal views and may reflect feedback anticipated from a colleague, client, or other stakeholder.

Table 8.2

Core phases of the nominal group technique.

Phase	Description
Silent generation of ideas	Participants work silently and independently to generate and write down as many ideas regarding the topic as possible.
Round robin recording of ideas	Each participant in turn reads out one of their ideas until no new ideas are available. There is no discussion. The group facilitator records the ideas for all to see.
Clarification discussion	Participants discuss each idea in turn with minimal input from the group facilitator. They are invited to ask questions regarding its intention, meaning, and underlying logic. Participants are asked to agree shared meaning, but reminded that there is no need to find consensus as to its importance.
Anonymous voting	Each participant works silently and independently to vote or rank the ideas that they feel are most important or of highest priority. These votes are aggregated to provide an overall indication of the level of consensus amongst the group.

*As described by Carney et al., 1996; de Ruyter, 1996; Delbecq et al., 1975; Harvey & Holmes, 2012; Hutchings et al., 2013; Lloyd, 2011; McMillan et al., 2016; Rankin et al., 2016; Reimer et al., 2019; Van de Ven & Delbecq, 1971.

Following silent generation, participants were asked to each read out one of their comments in turn in a round robin fashion until no new comments were available. The facilitator recorded these contributions for all participants to see using a flipchart. Participants were asked not to discuss their contributions and invited to continue adding ideas to their lists as the process continued if new comments occurred to them. Once all ideas had been recorded, a clarification discussion took place. Each idea was taken in turn and discussed by participants to clarify its meaning and identify duplicates that may be removed from the list. The facilitator took a non-directive role in this discussion, intervening only to keep the discussion moving to time or remind participants that they were seeking clarity rather than consensus.

Following clarification, participants were provided with voting sheets to complete privately and silently regarding the perceived importance of the ideas elicited. They were asked to write down the five ideas they deemed most important and assign points to them in the following format: most important received five points, second most important received four points, third most important received three points, fourth most important received two points, and the least important received one point. This scoring system was displayed visually for participants.

8.2.4 Analyses.

The data outputs were the list of ideas and the aggregated voting scores provided by each group. Votes cast by each group were counted and totalled to provide an indication of the level of consensus within and across groups. The content of all ideas was also subjected to thematic analysis using an adapted form of the constant comparative technique described by Boeije (2002). Upon completion of the first group, each idea was considered in turn and a code assigned to it. Once a code had been assigned, all other ideas were compared against it for inclusion with that code. This process was repeated for all ideas from the first group. This process was then repeated for each group by first comparing all ideas against the codes already identified and creating new ones as appropriate. Thematic analysis was undertaken following each group to assess progress towards data saturation. Data saturation was deemed to have been achieved where minimal new themes were identified following

an additional group. The themes and subthemes identified were then discussed in narrative form in relation to existing literature and their implications for the future of the model of interpersonal professionalism.

8.3 Results

Four nominal group sessions were carried out with a total of 17 participants. A summary of the demographic characteristics of participants is provided in table 8.3. Two groups were conducted with occupational psychologists in training and two with qualified occupational psychologists. The number of participants per group ranged from two to six. A summary of group characteristics is provided in table 8.4. The number of ideas generated ranged from 18 to 37. All generated ideas are provided in full using the wording agreed by participants in table 8.5.

Table 8.3*Summary of participant sample characteristics.*

Characteristic		Group 1		Group 2		Group 3		Group 4	
		N	%	N	%	N	%	N	%
Age	18-30 years	5	83%	1	25%	3	60%	0	0%
	31-45 years	1	17%	2	50%	0	0%	2	100%
	46-60 years	0	0%	1	25%	2	40%	0	0%
	61+ years	0	0%	0	0%	0	0%	0	0%
	Declined	0	0%	0	0%	0	0%	0	0%
Gender	Female	5	83%	2	50%	2	40%	1	50%
	Male	1	17%	2	50%	3	60%	1	50%
	Declined	0	0%	0	0%	0	0%	0	0%
Sector of current/most recent occupation	Administration	1	17%	1	25%	0	0%	0	0%
	Education	0	0%	0	0%	0	0%	2	100%
	Healthcare	2	33%	0	0%	0	0%	0	0%
	Management/leadership	0	0%	2	50%	3	60%	0	0%
	Sales	2	33%	1	25%	0	0%	0	0%
	Other - research	0	0%	0	0%	1	20%	0	0%
	Declined	1	17%	0	0%	1	20%	0	0%
Time spent in workplace	Early career (0-10 years)	5	83%	1	25%	3	60%	1	50%
	Developing career (11-20 years)	1	17%	1	25%	0	0%	1	50%
	Established career (21-30 years)	0	0%	2	50%	1	20%	0	0%
	Late career (31+ years)	0	0%	0	0%	1	20%	0	0%
	Declined	0	0%	0	0%	0	0%	0	0%
Level of education completed	Levels 1-2 (compulsory high school e.g. GCSEs)	0	0%	0	0%	0	0%	0	0%
	Levels 3-6 (further and undergraduate higher education)	4	67%	4	100%	0	0%	0	0%
	Level 7+ (postgraduate higher education)	1	17%	0	0%	5	100%	2	100%
	Declined	1	17%	0	0%	0	0%	0	0%

Table 8.4*Summary of group characteristics.*

Characteristic	Group 1	Group 2	Group 3	Group 4
Target participant sample	Occupational psychology practitioners in training	Occupational psychology practitioners in training	Occupational psychologists in professional practice context	Occupational psychologists in academic context
Number of participants	6	4	5	2
Total number of ideas generated	19	37	26	18

Table 8.5*Ideas generated by each group.*

Group	Number	Idea
Group one	1	Mood and own experiences can have a big impact.
	2	It is important to have boundaries, e.g. touch preferences, between a client and professional.
	3	Situational context is important – important for everything.
	4	Effective communication and listening are important among one another.
	5	It is worth considering shared assumed moral values as a result of the communities the people live in. Each society has ways of operating that are assumed. Contraventions e.g. sexual harassment mean that we cannot assume that the values are shared.
	6	What is professional to one individual may not be to someone else.
	7	It is important to have understanding between you and your client as to what you are both trying to achieve.
	8	The way somebody behaves and conducts themselves is important.
	9	Don't inflict your own personal (behaviour) opinions/beliefs onto others - behaviour and beliefs are not the same thing.
	10	It is worth establishing what professionalism is for, or there is no way for me to understand if I'm doing it.
	11	It is important to reflect on good/bad situations as a producer so you know what you're doing right/wrong.
	12	The model focuses on a two-person interaction and most interactions in team situations involve a third+ person.
	13	It is important to take other people's values and concepts into consideration to improve workplace relationships.
	14	It is worth considering that people tend to disregard their individual differences for the sake of professionalism. Most people aren't exactly themselves when in work.
	15	People's behaviour can be a direct result of the nature of the work they're involved in.
	16	It took a while to get my head around the way the interaction is described (producer/receiver).
	17	We need to set standards of professionalism to mitigate cases where it means different things to different people.
	18	It is worth having separate models for client versus colleague interactions.
	19	Can you train to be professional?
Group two	1	The model sets out values (i.e. terms or a framework) rather than interpreting professionalism. It provides a set of values and identifies them as professionalism.
	2	The model could be called 'interpersonal respect' rather than professionalism - it sets out how to respect other people's opinions rather than what professional behaviour is.
	3	The model incorporates emotional intelligence into interpersonal respect.
	4	Perception - how different people perceive ideas is open to different "translations" and is subjective.
	5	It depends on skills and abilities. Some aspects of professionalism cannot be learned/taught.
	6	It is mood-dependent.
	7	Individual differences in perceived concepts are not clear.
	8	I agree that professionalism is more than just the intentions of the producer.

9 I disagree that people rely on their professionalism in unfamiliar situations. Gut instinct may conflict with professionalism e.g. what professionalism says I should do vs what is right. You might leave a job in an unprofessional way because the work is unethical. You don't always follow the professionalism rules, you trust your gut instinct instead.

10 The model reflects a set of company values that both parties consider professional. It reflects something set by the company.

11 It depends on individual differences too e.g. the producer needs to be flexible with their thoughts and processes - you might not be able to teach this as it is ingrained within them; you either have it or you don't.

12 The model reads like it's based on consultancy. There is an idea that professionalism is based on the receiver's perception when you have to please them, but realistically in retail you just want to do the bare minimum.

13 The model does not reflect manager and worker interactions.

14 Professional = chartered status so you get to use the title because you have attained a level and gone up the scale first. The model is therefore about values, not professionalism as it doesn't matter how good or bad I am because I've attained professionalism.

15 The model states that it depends on the producer's and receiver's mood at the time but professionalism shouldn't be mood-dependent. Professionalism = regulating moods so personal feelings don't come into it. Professionalism should mute mood issues.

16 Regarding the final feedback and reflection stages, is the outcome of the whole interaction considered? You will get negative feedback if I didn't get what I wanted.

17 I liked the inclusion of social interaction e.g. if you're in the 'in-group' with managers or clients, your life is easier. If you're sociable and can get managers to like you, its easier because they think you're a hard worker.

18 The producer has a position of power - the model assumes a power dynamic.

19 It suggests that the producer's motivation might change but if you're professional, it shouldn't. You should have one way of working professionally.

20 I like that it considers previous experiences - you can look back to find things to help your current approach.

21 I agree that a good producer needs to be able to understand and use new information in future - when relying on 'base tendencies', you have to use new information. You can't just copy and paste.

22 The model is just a framework for courteous behaviour and respect.

23 It doesn't mention cultural differences at all - different cultures will understand different phrasing in different ways.

24 It doesn't consider different management skills and levels.

25 It doesn't consider ignorance - it automatically assumes that the producer looks at feedback. They might not listen to it.

26 It doesn't include lack of ability to use feedback even if you're trying.

27 There is no definition of professionalism in the model - it doesn't say what the interpretation of professionalism is in whatever role it is set in.

28 Context is so important - professionalism is context-dependent.

29 Professionalism is about consistency, so it shouldn't vary based on the situation.

30 Employers have written standards to guide professional behaviour e.g. codes of conduct.

31 Codes of conduct can be misunderstood due to language skills or cultural backgrounds.

32 "Courteous" means different things to different people e.g. banter vs bullying.

33 The model has no examples to provide a definition and help understand the model.

34 You can't expect training on every aspect of professionalism - some things you just have to pick up for yourselves i.e. how to interpret a definition/examples in practice.

35 The model needs to be subjective - it needs to take into account that codes of conduct and policies don't answer every question.

	36	The model assumes that the producer approaches the receiver so it doesn't account for the receiver starting the interaction.
	37	The model is ignorant - it assumes that an interaction can only happen this one way. Its needs to be more flexible.
Group three	1	The model reflects opportunities to develop/measure skill sets.
	2	I like figure 4 a lot - it blends situation, mood, and other factors. It is plausible.
	3	Emotional intelligence for the producer is crucial.
	4	I like that it is a two-way process.
	5	I like that it starts to take multiple factors into account and complex systems.
	6	However, successful outcome measures might be better to avoid oversimplification.
	7	Page 2 - there is no space in the model for experience. How people apply their knowledge/skills changes throughout their careers.
	8	The model does not mention the producer's ability to be open to and use feedback.
	9	Fundamental values and beliefs is very broad.
	10	Re. mood - it is interesting that it is considered but it might require further exploration.
	11	Figure 5 - I wonder how well this would be reflected in reality.
	12	Shannon & Weaver's model of communication concept of noise might be relevant.
	13	I think personality has an impact as part of fundamental values and beliefs or default approaches.
	14	Struggling to see the outcome of the model. What is the goal of the model?
	15	Competency? Where does it fit in?
	16	I think that the producer's ability to 'read' and act on the receiver's feedback will be a big determinant of how professional they are perceived to be.
	17	The relationship between the producer/receiver would affect how professional you can be/how much effort/skill it would take to be professional.
	18	Learned behaviours - can you learn to demonstrate professionalism?
	19	Other possible factors - physical environment is tidy/neat, waiting times, etc.
	20	If the producer has more maladaptive personality traits, it would detrimentally affect their professionalism e.g. overplayed/extremes of otherwise adaptive traits.
	21	Neuroticism/risk reaction/reward reaction - high or low levels may impact professionalism negatively.
	22	Being responsive rather than reactive would increase perceptions of professionalism.
	23	Space between stimulus and response - what you use this for may affect professionalism.
	24	Figure 4 - there may be a hierarchy among the factors.
	25	Satisfaction/dissatisfaction/outcome determines your view of the other person's professionalism.
	26	Outcome measures might be oversimplifications in themselves.
Group four	1	Professionals are held up to a historic and potentially fictional standard in the modern world.
	2	Concept of fundamental values and beliefs is important as it guides the behaviour of the producer.
	3	Fundamental values and beliefs of the producer might not be perceived/might be perceived differently by the receiver.
	4	Do producers and receivers have the same source in terms of their concept of professionalism?
	5	Halo and horns effects - anything that the producer portrays could be rapidly by-passed by these effects.
	6	Fig 1 mentions verbal and non verbal feedback - a neurodiverse producer, e.g. doctor on the autistic spectrum - the dialogue/interaction would be affected.

- 7 Producer and receiver state will completely change their perceptions of professionalism.
 - 8 Personal distance is not always professional.
 - 9 Fig 3 - does mood affect all of the contextual factors? E.g. fundamental values and beliefs/skills and abilities.
 - 10 Could use an equation approach to define the situation and behaviour, such as that of situational leadership theory (managerial grid).
 - 11 Stage model - stage 1 makes sense.
 - 12 Stage 2 role of familiarity and anxiety/pressure - hugely important for both parties in determining how the interaction goes.
 - 13 Emotional regulation strategies of both parties will play a role.
 - 14 Regarding the reflexive/iterative nature of the model - what factors would influence whether the feedback received actually influenced the default approach?
 - 15 Re measurement - a lot needs defining in order to measure professionalism and a simple questionnaire just won't do it!
 - 16 Model makes sense in my experience as you do change your approach based on receiver cues.
 - 17 Misinterpreting receiver cues could have serious consequences.
 - 18 Stage 1/default approach is similar to personality - there are elements of personality but it is not explicit.
-

Note. Ideas are recorded verbatim according to the wording agreed by each group. Where ideas refer to figures, stages, or pages, please see the model of interpersonal professionalism within appendix G.

8.3.1 Votes and scores.

The votes cast by each group were totalled, along with the number of individual votes received by each idea. The total scores were used to place the ideas prioritised by each group into rank order. One participant vote from group one and all participant votes from group two were excluded from the voting totals due to incorrect use of the voting procedure. The exclusion within group one was due to participant error but the exclusion of group two was identified as resulting from a facilitator training issue, which was rectified before any further groups took place. Ideas that received votes are displayed in rank order in table 8.6 along with their total vote scores and number of votes received. Voting results suggested little consensus amongst groups one and three. Group four demonstrated high consensus, but this was likely to be an artefact of the small participant number, as consensus is more likely among smaller numbers of individuals (Alberti, 2014; Hare, 1952; Manners, 1975).

Table 8.6*Group consensus regarding the prioritisation of ideas.*

Group	Ranking	Idea number	Idea	Number of votes	Total score
Group one	1	4	Effective communication and listening are important among one another.	3	14
	2	13	It is important to take other people's values and concepts into consideration to improve workplace relationships.	3	9
	2=	17	We need to set standards of professionalism to mitigate cases where it means different things to different people.	3	9
	4	2	It is important to have boundaries e.g. touch preferences, between a client and professional.	3	8
	5=	8	The way somebody behaves and conducts themselves is important.	2	7
	5=	15	People's behaviour can be a direct result of the nature of the work they're involved in.	2	7
	7=	10	It is worth establishing what professionalism is for, or there is no way for me to understand if I'm doing it.	1	5
	7=	11	It is important to reflect on good/bad situations as a producer so you know what you're doing right/wrong.	2	5
	9	7	It is important to have understanding between you and your client as to what you are both trying to achieve.	1	4
	10=	14	It is worth considering that people tend to disregard their individual differences for the sake of professionalism. Most people aren't exactly themselves when in work.	2	3
	10=	18	It is worth having separate models for client versus colleague interactions.	2	3
	12	1	Mood and own experiences can have a big impact.	1	1
Group three	1	18	Learned behaviours - can you learn to demonstrate professionalism?	2	8
	2	3	Emotional intelligence for the producer is crucial.	2	7
	3	20	If the producer has more maladaptive personality traits, it would detrimentally affect their professionalism e.g. overplayed/extremes of otherwise adaptive traits.	2	6
	4=	4	I like that it is a two-way process.	1	5
	4=	6	However, successful outcome measures might be better to avoid oversimplification.	2	5
	4=	12	Shannon & Weaver's model of communication concept of noise might be relevant.	1	5
	4=	16	I think that the producer's ability to 'read' and act on the receiver's feedback will be a big determinant of how professional they are perceived to be.	2	5
	4=	22	Being responsive rather than reactive would increase perceptions of professionalism.	1	5
	9=	1	The model reflects opportunities to develop/measure skill sets.	1	4
	9=	2	I like figure 4 a lot - it blends situation, mood, and other factors. It is plausible.	1	4
	9=	9	Fundamental values and beliefs is very broad.	1	4

	9=	13	I think personality has an impact as part of fundamental values and beliefs or default approaches.	1	4
	13=	5	I like that it starts to take multiple factors into account and complex systems.	1	3
	13=	24	Figure 4 - there may be a hierarchy among the factors.	2	3
	15=	7	Page 2 - there is no space in the model for experience. How people apply their knowledge/skills changes throughout their careers.	2	2
	15=	19	Other possible factors - physical environment is tidy/neat, waiting times, etc.	1	2
	17=	11	Figure 5 - I wonder how well this would be reflected in reality.	1	1
	17=	25	Satisfaction/dissatisfaction/outcome determines your view of the other person's professionalism.	1	1
	17=	26	Outcome measures might be oversimplifications in themselves.	1	1
Group four	1	15	Re measurement - a lot needs defining in order to measure professionalism and a simple questionnaire just won't do it!	2	10
	2	13	Emotional regulation strategies of both parties will play a role.	2	6
	3	4	Do producers and receivers have the same source in terms of their concept of professionalism?	2	5
	4	9	Fig 3 - does mood affect all of the contextual factors? E.g. fundamental values and beliefs/skills and abilities.	1	4
	5	17	Misinterpreting receiver cues could have serious consequences.	2	3
	6	3	Fundamental values and beliefs of the producer might not be perceived/might be perceived differently by the receiver.	1	2

Note. Ideas receiving votes are listed in rank order according to scores received from each group. Ideas generated but not listed in this table received no votes. Group one scores and votes based on data from five of the six participants due to incorrect use of voting procedure by one participant. Group two scores and votes not included due to use of incorrect voting procedure by all participants.

8.3.2 Thematic analysis.

The ideas from each group were analysed prior to the next group taking place to compare the data elicited following each group. Following the fourth group, only one new theme was identified, which did not appear in the previous three groups. This was interpreted as indicating data saturation. The themes and subthemes identified are displayed in table 8.7, along with their frequency and extensiveness (Krueger & Casey, 2015). Frequency reflects the total number of ideas relating to each theme or subtheme. Extensiveness reflects the total number of groups mentioning each theme or subtheme. The most prominent theme in terms of frequency was the one receiving the highest count and the most prominent subthemes were identified as those with frequency counts of 20 or above. The most prominent themes and subthemes in terms of extensiveness were those receiving counts of four.

Table 8.7*Themes and subthemes identified.*

	Theme	Frequency	Extensiveness		Subtheme	Frequency	Extensiveness
1	Factors important to professionalism	111*	4*	1a	The role of affect	7	4*
				1b	The role of context	13	2
				1c	The role of interpersonal context	25*	4*
				1d	The role of skills in the moment	11	2
				1e	The importance of collaboration	3	1
				1f	Duality of thought vs action	3	2
				1g	Individual differences don't matter	3	2
				1h	Individual differences do matter	15	3
				1i	The role of personality	2	2
				1j	The role of intention vs perception	4	2
				1k	The role of previous experiences	3	3
				1l	Agency/choice	9	2
				1m	Multi-factor model	6	1
				1n	The role of emotional intelligence	4	2
				1o	The role of pressure	1	1
2	What is professionalism?	49	3	2a	Defining professionalism subjectively	9	2
				2b	Defining professionalism objectively	9	2
				2c	Rules of professionalism and what is should/ought to be	21*	3
				2d	Status	3	2
				2e	Subjective basis of the concept	6	3
3	Evaluation of the model	88	4*	3a	Purpose of professionalism and the model	10	4*
				3b	Oversimplifications and omissions	31*	4*
				3c	Clarity of description	7	4*
				3d	Interpretations and perceived assumptions of the model	12	3
				3e	Perceptions of power within the model	2	1
				3f	The model vs reality	6	2
				3g	Aspects of the model that make sense/I like/I think are important	20*	4*
4	Professionalism development	12	3	4a	Teachability/learnability and professionalism	8	3
				4b	Opportunities for professional development	4	2
5	Measurement	4	2	None identified		4	2

Note. Concepts of frequency and extensiveness adapted from Krueger and Casey (2015). Frequency reflects the number of ideas associated with each theme/subtheme. Extensiveness reflects the number of groups mentioning each theme/subtheme. *indicates the most prominent themes in terms of frequency and extensiveness, defined as those appearing in 20 or more ideas and being mentioned by all four groups respectively.

8.3.2.1 Most prominent themes.

1. *Factors important to professionalism* (frequency=110) was the most prominent theme in terms of frequency. The related subtheme of 1c. *The role of interpersonal context* (frequency=24) was also identified as prominent according to frequency. Although no other themes were identified as prominent according to frequency, a number of subthemes were. These included 2c. *Rules of professionalism and what it should/ought to be* (frequency=21) and 3g. *Aspects of the model that make sense/I like/I think are important* (frequency=20). The most prominent subtheme in terms of frequency was 3b. *Oversimplifications and omissions* (frequency=31).

The most prominent themes in terms of extensiveness were 1. *Factors important to professionalism* and 3. *Evaluation of the model*. Within the former, the subthemes 1a. *The role of affect* and 1c. *The role of interpersonal context* were identified as most prominent in terms of extensiveness. Within the latter theme, the subthemes 3a. *Purpose of professionalism and the model*, 3b. *Oversimplifications and omissions*, 2c. *Clarity of description*, and 3g. *Aspects of the model that make sense/I like/I think are important* were identified as most prominent in terms of extensiveness.

8.4 Discussion

This study consulted relevant stakeholders regarding the model of interpersonal professionalism and its potential for further development as the basis for a psychometric measure of professionalism. It was anticipated that voting consensus would indicate the most important feedback themes to be addressed, but the data suggested very little consensus both within and across groups, limiting the utility of voting scores as a means of informing further development. In addition, despite specifically intending to target feedback regarding the model of interpersonal professionalism and its potential utility in an occupational context, this appeared in only two of the identified themes (3. *Evaluation of the model* and 5. *Measurement*), with participants instead focusing overwhelmingly on aspects that complemented/contravened their own views regarding professionalism.

The lack of consensus makes intuitive sense in the context of professionalism because as discussed in chapter 1, attempts to define professionalism garner similarly low levels of consensus (Anderson et al., 2014; Birden et al., 2014; Blake & Gutierrez, 2011; Buck et al., 2015; Evans, 2008; Finn et al., 2010; Freidson, 1994; Goldie, 2013; Marei et al., 2018; Mazor et al., 2007; Monrouxe et al., 2011; O’Flynn et al., 2014; Trathen & Gallagher, 2009; Wilkinson et al., 2009). The data of this study suggest that it may be difficult for participants to move away from their personal definitions of professionalism towards one provided for them. The tendency of participants to focus on their personal theories of professionalism supports the assertions made in chapters 6 and 7 regarding the use of explicit research methods in relation to professionalism. Evidence reported in these chapters suggests that asking stakeholders explicit questions about professionalism may result in the receipt of espoused theories in response (Argyris & Schön, 1974). Although the focus of the present study was consultation regarding a specific model of professionalism, it appears that the nominal group methodology may still be susceptible to this effect and so future similar research may benefit from considering alternative methods of consultation.

It was further observed that espoused theories tended to feature more heavily in the ideas generated by trainees than experienced occupational psychologists, whereby they provided statements of what they felt professionalism ought to look like that appeared unconnected to the content of the model (e.g. group on, idea two “*It is important to have boundaries*”). One reason for this may lie in the potential reported in chapter 6 for participants who have not given the subject matter of professionalism significant previous thought to fall back on providing ideas perceived as ‘correct’ or espoused theories. Although occupational psychology has largely neglected professionalism, experienced occupational psychologists are more likely to have encountered related issues during their career than trainees, such as those relating to performance enhancement. As such, although not under the guise of professionalism per se, they may have previously reflected on relevant issues in more detail than those early in their occupational psychology career. This would account for the apparent disparity between trainees and qualified occupational psychologists in terms of providing espoused theories of professionalism. This

finding adds further evidence that seeking alternative methods for stakeholder consultation regarding professionalism may be appropriate in future research.

8.4.1 Implications for the model of interpersonal professionalism.

Within theme 3. *Evaluation of the model*, the most prominent subtheme was 3b. *Oversimplifications and omissions*. A total of 31 related ideas were contributed across all four groups. These ideas were highly idiosyncratic, suggesting that the model required extending to include a wide range of additional factors. Resulting from the lack of consensus demonstrated amongst participants, no clear priorities emerged in this regard. Responding to such a wide range of perceived omissions may result in an overly complex theory that, if used as the basis of a future measure of professionalism, would require that measure to be similarly long and detailed. This is in line with other research findings that suggest that measures of professionalism attempting to respond to the full breadth of espoused theories are indeed prohibitively resource intensive to complete and impractical to use (Aguilar et al., 2011; Blake & Gutierrez, 2011; Kelley et al., 2011). This suggests once more that an alternative approach to consulting stakeholders regarding the model is needed in future in order to provide more targeted feedback fit for implementation.

The second most prominent subtheme encompassed aspects of the model that made intuitive sense to participants (3g. *Aspects of the model that make sense/I like/I think are important*). Ideas associated with this subtheme praised the multi-factorial nature of the model, agreeing that professionalism was likely to involve factors relating to mood, situational and interpersonal context, and emotional intelligence. However, this issue was also questioned as a potential limitation of the model within theme 5. *Measurement*. Specifically, participants suggested that the model represents a process rather than construct definition, and that any attempt to measure that process of professionalism using a standard psychometric tool would constitute an oversimplification of the issue.

It is worth noting that this feedback was received only from participants in the groups including qualified occupational psychologists, suggesting that their enhanced understanding of the mechanisms of psychometric measurement was

relevant. As discussed in chapters 1 and 5, the majority of published literature seeking to measure professionalism as a psychological attribute is generated by the field of medical education and undertaken by non-specialists in psychometric theory (Bertolami, 2004; Birden et al., 2014; Cruess, 2006; Evetts, 2006; Veloski et al., 2005). Such authors are unlikely to have extensive applied experience in the fields of occupational psychology and psychometry. As such, they may have been errant from the outset in their interpretation of professionalism as an attribute that may be psychometrically measured at all. Given the high numbers of tools published as measures of professionalism claiming psychometric validation and used for the purposes of assessment in education and training, this possibility is concerning. Further data is required to explore this issue in more detail, but future research should be undertaken without an *a priori* assumption that professionalism actually lends itself to psychometric measurement, in order to avoid biased conclusions.

8.4.2 Study evaluation.

The findings of the present study should be understood in the context of their limits. This study included small numbers of participants and groups to determine the level of consensus regarding feedback ideas. Although this enabled the collection of rich and detailed data, it also means that the conclusions drawn may not be generalised to other participants or groups. The present study provides a pilot-scale consultation with only one relevant stakeholder group and should therefore not be interpreted as constituting a full or extensive consultation.

In methodological terms, sampling was undertaken assuming that the unit of relevance was number of groups, not participants within those groups, as recommended by practice guidelines (Carlson & Glenton, 2011). Such guidelines recommend that three to six groups are sufficient to identify 90% of extant feedback themes and all those most prevalent (Cyr, 2019; Carlson & Glenton, 2011; Guest et al., 2016). The process of thematic analysis supported this as only one new theme was identified following the final group, suggesting that data saturation had been achieved. Data saturation refers to the point at which no new information is generated by additional groups (Guest et al., 2016) and has the somewhat controversial legacy of being commonly used to provide *post hoc* justification for

small sample sizes (Carlson & Glenton, 2011). Although the present study was small, the described use of thematic analysis and adherence to best practice guidelines regarding the nominal group technique constituted steps taken to ensure that the concept of data saturation was not used in this way.

A second methodological issue was that one group involved only two participants. The minimum advised participants per group within the literature is around four, although instances are available in the published literature where smaller groups have yielded useful data (McMillan et al., 2014, 2015; McMillan et al., 2016). The data from group four was retained within the present study for two reasons. Firstly, the ideas and votes provided useful insight relevant to the research question and so were deemed worthy of inclusion under the pragmatic research paradigm. Secondly, the interpersonal context within the nominal group technique is credited with increasing the amount of data elicited from a group via constructive tension and hitchhiking, as discussed in chapter 4. This was interpreted to mean that groups with fewer participants would fail to capitalise on these benefits to the same extent as other groups, rather than affect the integrity of the data elicited. As such, it was deemed appropriate to retain the group's data for further analysis.

A final methodological issue relates to the exclusion of group two's votes due to incorrect use of the voting procedure. This was identified as resulting from a training issue with that group's facilitator that was rectified before any further groups were carried out. However, it did mean that data from one group was inadmissible in terms of voting scores. This was deemed to have had limited impact on the results of this study, however, as the lack of consensus among and across the remaining groups was sufficiently low to render it of limited utility.

8.4.3 Future research directions.

The present study was a small-scale pilot stakeholder consultation to gather feedback regarding the potential utility of the model of interpersonal professionalism for further development in the context of psychometry. However, the findings provided greater insight into the process of stakeholder consultation than the utility of the model. This was unexpected but provides useful directions for future research.

Specifically, data suggested that undertaking a stakeholder consultation regarding a model of professionalism using the nominal group technique was susceptible to the effects of espoused theories of professionalism, particularly where stakeholders had limited expertise in or experience of the concept. In order to gain useable insight, alternative research methods should be explored.

In terms of the model itself, the most striking finding was that although the model was generally positively received by experienced occupational psychology professionals, concerns were raised as to whether it could be used as a basis for psychometric measurement without excessively oversimplifying the issue. Further empirical scrutiny of the model is required in order to determine whether a specific construct of professionalism is evident, but it should be acknowledged that the results of this scrutiny could have profound implications for current practice in the teaching and learning of professionalism and the ways that it is assessed and measured.

8.5 Chapter Summary and Conclusions

This chapter described a pilot stakeholder consultation study using the nominal group technique. The aim of this study was to gather initial feedback from key stakeholders regarding the acceptability and usability of the model of interpersonal professionalism, as a basis for deciding whether a full stakeholder consultation was warranted. The data suggests that both the model and method of consultation require reconsidering before such consultation takes place. The second aim was to inform the materials to be used to undertake full stakeholder consultation by indicating the highest priorities to be addressed via voting consensus. This aim was not achieved as the level of consensus across participants and groups was lower than expected.

The data did highlight unexpected insights into the nominal group technique method and the potential for the model as a basis of psychological measurement. Based on the data gathered, the former is not recommended for further stakeholder consultation regarding professionalism, as the explicit nature of the method tends to result in espoused theories rather than feedback focused on the model itself. The

latter was criticised by experienced occupational psychologists as not constituting a definition of the construct of professionalism sufficient to enable psychometric assessment. In light of this feedback, further research would be required to further evaluate the model of interpersonal professionalism.

Part IV: Overall Discussion, Conclusions, and Recommendations

The research reported in this thesis explored the concept of professionalism from a psychometric perspective and recommendations are made for the future of the field. Several new contributions to knowledge are made, both in terms of psychological theory and measurement, and methodological insights. This thesis is believed to be the first programme of research to approach professionalism specifically from a psychometric perspective. It constitutes the early groundwork required for a broader, long-term strategy to support further psychological research into professionalism and its development, and related improvements to teaching and learning practices.

9.1 Contributions Relevant to Psychological Theory and Measurement

Systematic and comprehensive evidence is provided as to the current state of the art regarding the measurement of professionalism as a psychological attribute. Based on this evidence, recommendations are made to improve related assessment and measurement. Findings suggest that current practice in measuring professionalism as a psychological attribute is poor, with a return to basic psychometric principles required to pave the way for improvements. To this end, creating a theoretical account of professionalism as the basis for a construct definition was identified as the critical first step.

A series of empirical studies were undertaken in response to this identified need. Data were gathered using the RGT to explore the ways that individuals construe professionalism. This data indicated that the theories-in-use used by individuals making real-world judgement decisions when perceiving professionalism, are different to the espoused theories explicated by previous research. Previous research has generally relied upon top-down approaches to defining the construct of professionalism, such as by asking experts to describe the way that they perceive it (Birden et al., 2014). However, the RGT data suggest that defining professionalism in this way may contradict the ways that it is actually perceived by individuals on a day-to-day basis. As such, findings indicate that efforts

to improve perceptions of professionalism based on top-down, espoused theories may be ineffective, as they are likely to target different characteristics to those actually being used to make such judgements.

The RGT data further suggest that perceived professionalism may actually be a negative characteristic that is judged primarily by the absence of unprofessionalism rather than the presence of its logical opposite. In addition, the roles of gender and personal likeability were found to be relevant to perceptions of professionalism. The data support the assertion that the gender of the appraiser may play a role in professionalism judgements sufficient to dictate categorisation into extreme poles (extremely professional or extremely unprofessional) and that personal likeability may play a role in appraisals only where the other party is perceived as unprofessional. These findings may have profound implications for the assessment of professionalism.

The Q data confirmed that different individuals view professionalism in different ways, a finding borne out within the related literature (Finn et al., 2010; Goldie, 2013; Monrouxe et al., 2011). However, it also provided the novel finding that different subjective viewpoints may actively contradict one another, suggesting that efforts to generate a normative definition of the construct of professionalism by aggregating the views of multiple individuals may be inappropriate. This may mean that the strategy of various regulatory bodies of basing their professionalism guidelines on consensus statements following public consultation may be equally inappropriate (e.g. General Dental Council, 2013; General Medical Council, 2013, 2016). The findings reported based on Q data provide evidence supporting a move away from normative definitions of professionalism and a shift of emphasis onto the social processes that facilitate the effective navigation of divergent viewpoints regarding professionalism during real-world interactions. These findings have implications for professionalism training, which currently focuses on understanding and implementing the 'rules' of a profession, rather than developing relevant requisite psychosocial skills.

Based on the combined Q and RGT data, and with contributions from existing psychological theory, a theoretical model of interpersonal professionalism was proposed and stakeholder consultation undertaken. The model explicated the

empirical findings of the RGT and Q studies within the context of the CAPS theory and theories of emotional intelligence, focussing on the interpersonal mechanisms that the RGT and Q data suggested may be relevant. The results of stakeholder consultation suggested that the model did not yet constitute a definition of the construct of professionalism sufficient to form the basis of a psychometric measurement tool, but highlighted areas where the model may be used to support improvements in perceived professionalism. Taken together, the findings of this thesis provide the foundations that were required to identify avenues for future research. The recommendations for further research made in this thesis are based on new, previously undiscovered insight into the processes underlying perceived professionalism.

9.2 Methodological Insights

This thesis reports a programme of research using methodological approaches previously untested in studying professionalism. It was the first to ground methods for studying professionalism in psychological theory such as the lexical hypothesis theory (Goldberg, 1992; John & Srivastava, 1999; Poropat & Corr, 2015), personal construct theory (Kelly, 1955), and Q methodology (Brown, 1993; Stephenson, 1935b; Watts & Stenner, 2012). This provided opportunities to use research methods never before applied to the problem of professionalism, resulting in new insights.

The use of the RGT and Q methodology both resulted in data that did not replicate previous findings or those of each other, suggesting that both may make distinct contributions in the study of perceived professionalism. The former indicates that the method of enquiry determines whether espoused versus theories-in-use are elicited, meaning that the method should be tailored to the intended use of the data. For example, where one seeks to improve perceptions of professionalism, there may be little value in using methods that elicit espoused theories regarding the construct, as these differ in content to those actually underlying its perception. The latter finding provides evidence that the conceptually distinct theoretical underpinnings of each methodology are correct in assuming that their related methods do indeed access different psychological phenomena, thus supporting the validity of each

theory. In terms of analysis, the RGT study also provided tentative proof of concept evidence for cluster analysis using numerical word vectors. Although beyond the remit of this thesis, the findings suggest key directions to develop this approach to exploring the semantic content of RGT data.

The use of the NGT confirmed the potential for explicit research methods to elicit espoused theories more readily than theories-in-use, when applied to the topic of professionalism, adding weight to the evidence already provided by the Q and RGT data. Findings indicated that this effect might be more pronounced when such methods are used with participants naïve to, or inexperienced in, the area of specialism under study. This means that that before further stakeholder consultation is attempted, alternative consultation methods should be pilot-tested for effectiveness in delivering the specific information required. Such research could explore non-explicit methods of stakeholder consultation, including the RGT and Q methodology.

9.3 Evaluation of the Approach of this Thesis

The mixed methods approach of this thesis was grounded in pragmatism, with methodological decisions made according to best practice and to maximise the integrity of the approach. The thesis was based on the assertion that professionalism is a complex, socially constructed phenomenon, the likes of which MMR is particularly well suited to exploring (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010b). Professionalism was conceived of as an issue involving both objective and subjective components, the former relating to observable behaviour and objective measurement, and the latter to the subjectively perceived and ideographically constructed nature of interpersonal interactions. The research described sought intersubjectivity by capturing objective accounts of subjectively experienced professionalism. In line with best practice recommendations, the selected research methods were deemed to offer the best options to achieving this aim and answering the specific research questions.

Employing mixed methods within this thesis capitalised on the advantage of enabling the research questions to dictate the methods chosen, rather than adapting

the questions specifically towards quantitative or qualitative research (Tashakkori & Teddlie, 2010b). This allowed the thesis to tell the story emerging from the data, rather than that required by paradigmatic allegiance. The field of professionalism is dominated by qualitative research, with explicit methods taking centre stage (e.g. Carter et al., 2015; Finn et al., 2010; Monrouxe et al., 2011). Taking pragmatism to be the guiding principle in this thesis enabled a more flexible methodological approach undominated by the historical prevalence of qualitative methods. This resulted in the application of methods previously untested in exploring professionalism, enabling a new and previously untold story of it as a subjectively perceived construct.

Another strength of this thesis is its firm grounding in psychological theory. Despite a significant shift towards measuring professionalism as a psychological attribute (see discussion in chapter 1), previous research in this field has not been grounded in full and detailed understanding of psychological theory relevant to both professionalism and psychological measurement (e.g. Jha et al., 2007; Li et al., 2017; Lynch et al., 2004; de Mendonça et al., 2016; Veloski et al., 2005). The methodological approach of this thesis is grounded in well-developed and established psychological theory such as the lexical hypothesis (Goldberg, 1992; John & Srivastava, 1999; Poropat & Corr, 2015) and personal construct theory (Kelly, 1955), and Q methodology (Brown, 1993; Stephenson, 1935b; Watts & Stenner, 2012), with interpretations of findings adhering closely to the theoretical underpinnings thereof. Interpretations were also grounded in psychological theories of psychometry, behaviour, and cognition (see chapter 4).

The most pertinent limitation of the approach of this thesis relates to the limits of the interpretations of data that could be made, rather than a weakness per se. The study of professionalism from the perspective of psychology is a relatively untouched subject area. This means that little relevant theory was available from which to derive testable hypotheses that would provide generalisable findings. Positivist research provides generalisable findings, but only where specific hypotheses may be derived from general theories of the issue under study (Burke Johnson & Gray, 2010; Creswell & Plano Clark, 2011; Morgan, 2007; Neuman, 2000; Teddlie & Tashakkori, 2003, 2010). The lack of psychological theory relating to professionalism meant that no positivist methods were recommended for use in

this thesis and, by extension, that none of the findings may be generalised to other people, times, or contexts. The findings of this thesis should be interpreted accordingly and therefore situate their value in prompting theoretical developments that will, in future research, enable specific hypotheses to be derived and tested quantitatively such that findings may be more broadly extrapolated from.

9.4 Future Research Directions

Psychological research into the topic of professionalism is in its infancy, so there are a wide range of directions in which it might progress. This thesis has provided evidence to narrow this range somewhat. The study reported in chapter 5 demonstrates the need for further research targeting the psychometric measurement of professionalism. Although a measurable construct is not provided within this thesis, the model described in chapter 8 has the potential to lead to such a definition following further empirical study. Further research continuing to explore the potential for professionalism to be measured as a psychological attribute is therefore required, but without *a priori* expectations that it can be. Further data may endeavour to reveal the psychometric measurability of professionalism before progressing onto the development of such a measure, should it be recommended.

The study described in chapter 6 suggests ways that professionalism may be construed by individuals and possible areas where such construing is shared across multiple individuals. Future research should seek to confirm whether the construing reported in this thesis is also shared within the wider population, to support generalisable conclusions. Each of the principal components described in this chapter also suggest further areas for research: PC1 suggests merit in exploring professionalism as a negative construct, as something judged by the absence of unprofessionalism rather than presence of professionalism; PC2 suggests that further research into the role of gender in appraisals of professionalism is warranted; and PC3 suggests that the role of personal likeability requires further investigation, in order to further explicate its part in judgements of professionalism. The cluster analysis approach of this study also requires further study to explore in detail the potential reasons for its failure to provide useful results in this instance.

The study reported in chapter 7 suggests factors accounting for shared subjectivity in perceptions of professionalism. Further research exploring whether these factors are shared among the general population will further explain their relevance and potential applications in supporting professional development. This might be achieved by employing the Q2S (Q methodology to survey) methodology, which is specifically designed to estimate the presence and extent of non-generalisable Q factors within a representative sample, thus enabling generalisable conclusions (Baker et al., 2010; Baker, Wildman, Mason & Donaldson, 2014; Mason et al., 2016; Mason et al., 2018). The findings of the study described in chapter 8 suggest that future research is required to explore alternative methods for undertaking stakeholder consultation, particularly where the subject matter is susceptible to the influence of espoused theories.

9.5 Recommendations for the Teaching and Learning of Professionalism

The findings of this thesis suggest that there is no gold-standard psychometric tool currently available that may be recommended for educational assessment and qualification purposes. It is recommended that no extant measure be used for formative or summative assessment and that where tools are used for purely developmental purposes, this be in full acknowledgement and transparent discussion of the implications of their weak validation arguments. In terms of progressing towards a resolution to this issue, the empirical data reported in this thesis do not yet enable a definition of the construct of professionalism sufficient to develop a new measure with a stronger validation argument. Indeed, the data remain inconclusive as to whether such a definition is possible. In order to progress towards a generalisable conclusion in this regard, it is recommended that educators with an interest in professionalism work collaboratively and proactively with the related research agenda. Specifically, collaborations are recommended between vocational education institutions and psychological research programmes. Findings suggest that research to date undertaken outside the field of psychology regarding the potential to assess professionalism psychometrically fails to meet the required standards of rigour, resulting in a gap within the literature. Although this thesis has taken steps to begin rectifying this, ongoing collaboration between educators, relevant regulatory bodies,

and researchers specialising in occupational psychology and applied psychometric theory is required to respond to this currently unmet need.

A resolution to the above issue ought to be recognised as likely to require the commitment of a number of years work and so in the meantime, recommendations are also made for more immediate teaching and learning concerns. It is recommended that a curricula shift is required away from indoctrinating students into espoused theories of professionalism and towards greater understanding of, and skills relevant to, the dynamic and interpersonal nature of perceived professionalism. This might serve to address claims that current educational practices encourage students to reproduce rote learning and develop their ability to accurately predict correct answers relating to professionalism, rather than enhancing their related real-world behaviour in any way.

The model of interpersonal professionalism provides a number of opportunities for development at which teaching interventions might be effectively targeted. For example, interventions may seek to explore the impact of enhancing emotional intelligence and psychological flexibility, and the psychological antecedents of the effective and reflective use of feedback (e.g. Jellicoe & Forsythe, 2019). Fundamentally, educational policy is recommended to acknowledge that attempts to measure professionalism as a psychological attribute are, at this point, premature, and instead focus on factors affecting the processes at play during professional interactions and the development opportunities they offer, while making an ongoing commitment to supporting long-term research in this area.

9.6 Overall Conclusions

The research reported in this thesis makes novel contributions to the understanding of professionalism and ways that it might be measured as a psychological attribute. Referring to the research aims and objectives stated in chapter 3, this thesis provides a comprehensive and systematic methodological review of the state of the art in measuring professionalism with reference to contemporary psychometric best practice (aim 1; objective I). No extant measures were recommended for use and the first priority action for addressing this identified

as creating a theoretically grounded definition of the construct of professionalism (aim 2).

Empirical research was reported using established psychological research methods never before applied to the study of professionalism (aim 3). Findings suggested that the phenomenology of professionalism is distinct from documented espoused theories and that its complexity may be best understood as an interpersonal process rather than a discrete psychological attribute (objective II). These findings were used to propose a model of interpersonal professionalism (aim 4; objective III) that was presented to key stakeholders for initial consultation (objective IV).

Overall, there remains further work to be done in exploring professionalism and its development. However, this thesis has laid foundations that identify potentially fruitful directions for such progress while also recommending more immediate avenues for change within educational policy that may enhance the perceived professionalism of trainees and graduates. By extension, this could also serve to improve the reputational standing of both educating institutions and the vocational professions they serve. It is acknowledged that the politically charged history of professionalism may result in such change receiving challenge or even resistance amongst advocates of the sociological approach to understanding professionalism and role modelling approaches to its development. However, in the face of the apparent failure to date of these approaches to provide a theoretically and empirically supported solution to the increasing professionalism crisis, this thesis recommends that identified needs should finally be permitted to outweigh historical allegiances and enable the field to modernise.

References

- Abadel, F. T. & Hattab, A. S. (2014). Patients' assessment of professionalism and communication skills of medical graduates. *BMC Medical Education*, *14*(28), 1-8. DOI: 10.1186/1472-6920-14-28
- Abdi, H., Williams, L. J. & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*, 149-179. DOI: 10.1002/wics.1246
- Adam, J., Bore, M., Childs, R., Dunn, J., Mckendree, J., Munro, D. & Powis, D. (2015). Predictors of professional behaviour and academic outcomes in a UK medical school: a longitudinal cohort study. *Medical Teacher*, *37*(9), 868-880. DOI: 10.3109/0142159X.2015.1009023
- Adams-Webber, J. R. (2001). Cognitive complexity and role relationships. *Journal of Constructivist Psychology*, *14*, 43-50. DOI: 10.1080/10720530125762
- Aguilar, A. E., Stupans, L. & Scutter, S. (2011). Assessing students' professionalism: considering professionalism's diverging definitions. *Education for Health*, *24*(3), 1-10
- Akhund, S., Shaikh, Z. A. & Ali, S. A. (2014). Attitudes of Pakistani and Pakistani heritage medical students regarding professionalism at a medical college in Karachi, Pakistan. *BMC Research Notes*, *7*(150), 1-6. DOI: 10.1186/1756-0500-7-150
- Al Ansari, A., Al Khalifa, K., Al Azzawi, M., Al Almer, R., al Sharqi, D., Al-Mansoor, A. & Munshi, F. M. (2015). Cross-cultural challenges for assessing medical professionalism among clerkship physicians in a Middle Eastern country (Bahrain): feasibility and psychometric properties of multisource feedback. *Advances in Medical Education and Practice*, *6*, 509-515. DOI: 10.2147/AMEP.S86068
- Alberti, G. (2014). Modeling group size and scalar stress by logistic regression from an archaeological perspective. *PLoS ONE*, *9*(3), e91510. DOI: 10.1371/journal.pone.0091510
- Alcolta, M., Ruiz de Gauna, P. & González, F. E. (2016). Professionalism in dental education. *Revista Facultad de Odontología Universidad de Antioquia*, *28*(1), 158-178. DOI: 10.17533/udea.rfo.v28n1a9

- Al-Eraky, M. & Marei, H. (2016). A fresh look at Miller's pyramid: assessment at the 'Is' and 'Do' levels. *Medical Education*, 50, 1253-1257. DOI: 10.1111/medu.13101
- Allport, G. W. & Odbert, H. S. (1936). Trait names: a psycho-lexical study. *Psychological Monographs*, 47(1), i-171. DOI: 10.1037/h0093360
- Altirkawi, K. (2014). Teaching professionalism in medicine: what, why and how? *Sudanese Journal of Paediatrics*, 14(1), 31-38
- American Board of Internal Medicine (n.d.). *Mini-CEX clinical evaluation for trainees: direct observation assessment tool*. Retrieved from <https://www.abim.org/~media/ABIM%20Public/Files/pdf/paper-tools/mini-cex.pdf>
- American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201-238. DOI: 10.1037/h0053479
- Anderson, J. R., Taylor, L. F. & Gahimer, J. E. (2014). Assessing the impact of a short-term service-learning clinical experience on the development of professional behaviors of student physical therapists: a pilot study. *Journal of the Scholarship of Teaching and Learning*, 14(4), 130-143. DOI: 10.14434/josotl.v14i4.12788
- Aproxima Gesellschaft für Markt- und Sozialforschung Weimar (2015). HTMLQ (version 1.0.4). Available from <https://github.com/aproxima/htmlq/releases>
- Aramesh, K., Mohebbi, M., Jessri, M. & Sanagou, M. (2009). Measuring professionalism in residency training programs in Iran. *Medical Teacher*, 31(8), e356-e361. DOI: 10.1080/01421590802638022
- Aranda, G. & Finch, E. (2003). Using repertory grids to measure changes in risk-taking behaviour. *Journal of Construction Research*, 4(1), 101-114. DOI: 10.1142/S1609945103000340
- Argyris, C. & Schön, D. A. (1974). *Theory in practice: increasing professional effectiveness*. San Francisco (CA), USA: Jossey-Bass
- Argyris, C. & Schön, D. A. (1978). *Organizational learning: a theory of action perspective*. Reading (MA), USA: Addison-Wesley
- Armstrong, N., Hearnshaw, H., Powell, J. & Dale, J. (2007). Stakeholder perspectives on the development of a virtual clinic for diabetes care: a

- qualitative study. *Journal of Medical Internet Research*, 9(3), e23. DOI: 10.2196/jmir.9.3.e23
- Arndt, A. D., Evans, K. R., Zahedi, Z. & Khan, E. (2019). Competent or threatening? When looking like a “salesperson” is disadvantageous. *Journal of Retailing and Consumer Services*, 47, 166-176. DOI: 10.1016/j.jretconser.2018.11.012
- Arnold, L. (2002). Assessing professional behaviour: yesterday, today, and tomorrow. *Academic Medicine*, 77(6), 502-515
- Arnold, L. & Stern, D. T. (2006). What is medical professionalism? In D. T. Stern (Ed.), *Measuring medical professionalism* (pp. 15-37). New York (NY), USA: Oxford University Press
- Aryadoust, V. (2017). Understanding the role of likeability in the peer assessments of university students’ oral presentation skills. A latent variable approach. *Language Assessment Quarterly*, 14(4), 398-419. DOI: 10.1080/15434303.2017.1393820
- Ashton, M. C. & Lee, K. (2005). A defence of the lexical approach to the study of personality structure. *European Journal of Personality*, 19, 5-24. DOI: 10.1002/per.541
- Auerswald, M. & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions. *Psychological Methods*, advance online publication. DOI: 10.1037/met0000200
- Aukes, L. C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. P. & Sleats, J. P. J. (2007). The development of a scale to measure personal reflection in medical practice and education. *Medical Teacher*, 29(2-3), 177-182. DOI: 10.1080/01421590701299272
- Baernstein, A., Oeschlager, A. E. A. & Chang, T. (2009). Learning professionalism: perspectives of preclinical medical students. *Academic Medicine*, 84, 574-581. DOI: 10.1097/ACM.0b013e31819f5f60
- Baker, R., van Exel, J., Mason, H. & Stricklin, M. (2010). Connecting Q & surveys: three methods to explore factor membership in large samples. *Operant Subjectivity*, 34(1), 38-58. DOI: dx.doi.org/10.15133/j.os.2010.003

- Baker, R., Wildman, J., Mason, H. & Donaldson, C. (2014). Q-ing for health – a new approach to eliciting the public’s views on health care resource allocation. *Health Economics*, 23, 283-297. DOI: 10.1002/hec.2914
- Baldwin, D. C., Daugherty, S. R. & Rowley, B. D. (1998). Unethical and unprofessional conduct observed by residents during their first year of training. *Academic Medicine*, 73(11), 1195-1200
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs (NJ), USA: Prentice-Hall
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50, 248-287
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York (NY), USA: W. H. Freeman
- Bandura, A. (1999a). Social cognitive theory: an agentic perspective. *Asian Journal of Social Psychology*, 2, 21-41
- Bandura, A. (1999b). Social cognitive theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 154-196). New York (NY), USA: Guilford
- Barchard, K. A. & Hakstian, A. R. (1997). The effects of sampling model on inference with coefficient alpha. *Educational and Psychological Measurement*, 57(6), 893-905
- Barker, C. & Pistrang, N. (2005). Quality criteria under methodological pluralism: implications for conducting and evaluating research. *American Journal of Community Psychology*, 35(3/4), 201-212. DOI: 10.1007/s10464-005-3398-y
- Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M. & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, 31(5), 625-640. DOI: 10.1080/07294360.2012.702735
- Ben-David, M. F., Snadden, D. & Hesketh, A. (2004). Linking appraisal of PRHO professional competence of junior doctors to their education. *Medical Teacher*, 26(1), 63-70. DOI: 10.1080/0142159032000156524
- Bergman, M. M. (2008). Introduction: whither mixed methods? In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 1-7). London, UK: SAGE

- Bertolami, C. N. (2004). Why our ethics curricula don't work. *Journal of Dental Education*, 68(4), 414-425
- Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioural research* (2nd ed., pp. 95-118). Thousand Oaks (CA), USA: SAGE
- Birden, H., Glass, N., Wilson, I., Harrison, M., Usherwood, T. & Nass, D. (2014). Defining professionalism in medical education: a systematic review. *Medical Teacher*, 36, 47-61. DOI: 10.3109/0142159X.2014.850154
- Birt, L., Scott, S., Cavers, D., Campbell, C. & Walter, F. (2016). Member checking: a tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research*, 26(13), 1802-1811. DOI: 10.1177/1049732316654870
- Black, S., Estrada, C., de la Fuente, M. C., Orozco, A., Trabazo, A., de la Vega, S. & Gutshe Jr, R. E. (2019). "Nobody really wants to be called bossy or domineering": feminist critique of media "industry speak". *Journalism Practice*, 13(1), 35-31. DOI: 10.1080/17512786.2017.1399812
- Blackall, G. F., Melnick, S. A., Shoop, G. H., George, J., Lerner, S. M. Wilson, P. K. ... Kreher, M. (2007). Professionalism in medical education: the development and validation of a survey instrument to assess attitudes toward professionalism. *Medical Teacher*, 29(2-3), e58-e62. DOI: 10.1080/01421590601044984
- Blake, R. & Gutierrez, O. (2011). A semantic analysis approach for assessing professionalism using free-form text entered online. *Computers in Human Behavior*, 27, 2249-2262. DOI: 10.1016/j.chb.2011.07.004
- Block, J. (1956). A comparison of the forces and unforced Q-sorting procedures. *Educational and Psychological Measurement*, 16(4), 481-493. DOI: 10.1177/001316445601600406
- Block, J. (2008). *The Q sort in character appraisal: encoding subjective impressions of persons quantitatively*. Washington DC, USA: American Psychological Association
- Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & Quantity*, 36, 391-409

- Bolland, J. M. (1985). The search for structure: an alternative to the forced Q-sort technique. *Political Methodology*, 11(1/2), 91-107
- Bonke, B. (2006). Unprofessional or problematic behaviour of medical students outside the learning environment. *Medical Teacher*, 28(5), 440-442. DOI: 10.1080/01421590600626955
- Borsboom, D. (2005). *Measuring the mind: conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press
- Borsboom, D., Mellenburgh, G. J. & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071. DOI: 10.1037/0033-295X.111.4.1061
- Boyle, G. J., Stakov, L., Martin, N. G., Petrides, K. V., Eysenck, M. W. & Ortet, G. (2016). Hans J. Eysenck and Raymond B. Cattell on intelligence and personality. *Personality and Individual Differences*, 103, 40-47. DOI: 10.1016/j.paid.2016.04.029
- Brackett, M. A., Rivers, S. E. & Salovey, P. (2011). Emotional intelligence: implications for personal, social, academic, and workplace success. *Social and Personality Psychology Compass*, 5(1), 88-103
- Brennan, R. L. (2011). Generalizability theory and classical test theory. *Applied Measurement in Education*, 24, 1-21. DOI: 10.1080/08957347.2011.532417
- Brereton, L., Ingleton, C., Gardiner, C., Goyder, E., Mozygemba, K., Lysdahl, K. B. ... Oortwijn, W. (2017). Lay and professional stakeholder involvement in scoping palliative care issues: methods used in seven European countries. *Palliative Medicine*, 31(2), 181-192. DOI: 10.1177/0269216316649154
- Brewerton, P. & Millward, L. (2001). *Organizational research methods*. London, UK: SAGE
- Brincat, C. A. (2006). How medical training mangles professionalism: the prolonged death of compassion. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 199-210). Boston (MA), USA: Springer
- Bringle, R. G. & Hatcher, J. A. (1996). Implementing service learning in higher education. *The Journal of Higher Education*, 67(2), 221-239. DOI: 10.2307/2943981

- Bringmann, L. F. & Eronen, M. I. (2016). Heating up the measurements debate: what psychologists can learn from the history of physics. *Theory & Psychology, 26*(1), 27-43. DOI: 10.1177/095935431561725
- Brinkman, W. B., Geraghty, S. R., Lanphear, B. P., Khoury, J. C., Gonzalez, J. A., DeWitt, T. G. & Britto, M. T. (2007). Effect of multisource feedback on resident communication skills and professionalism. *Archives of Pediatrics and Adolescent Medicine, 161*, 44-49
- Brophy, S., Fransella, F. & Reed, N. (2003). The power of a good theory. In F. Fransella (Ed.), *International handbook of personal construct theory* (pp. 329-338). West Sussex, UK: Wiley
- Brouwer, M. (1999). Q is accounting for tastes. *Journal of Advertising Research, 39*(2), 35-39
- Brown, S. R. (1970). On the use of variance designs in Q methodology. *The Psychological Record, 20*, 179-189. DOI: 10.1007/BF03393928
- Brown, S. R. (1971). Experimental design and the structuring of theory. *Experimental Study of Politics, 1*(1), 1-41
- Brown, S. R. (1972). A fundamental incommensurability between objectivity and subjectivity. In S. R. Brown & D. J. Brenner (Eds.), *Science, psychology, and communication: essays honouring William Stephenson* (pp. 57-94). New York (NY), USA: Teachers College Press
- Brown, S. R. (1980). *Political subjectivity*. New Haven (CT), USA: Yale University Press
- Brown, S. R. (1993). A primer on Q methodology. *Operant Subjectivity, 16*(3/4), 91-138
- Brown, S. R. (1996). Q methodology and qualitative research. *Qualitative Health Research, 6*(4), 561-567
- Brown, N. & Bhugra, D. (2007). 'New' professionalism or professionalism derailed? *Psychiatric Bulletin, 31*, 281-283. DOI: 10.1192/pb.bp.107.016543
- Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/merge/fuse quantitative and qualitative research? In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 87-100). London, UK: SAGE

- Buck, E., Holden, M. & Szauter, K. (2015). A methodological review of the assessment of humanism in medical students. *Academic Medicine*, 90(11), S14-S23. DOI: 10.1097/ACM.0000000000000910
- Bumgarner, G. W., Spies, A. R., Asbill, S. & Prince, V. T. (2007). Using the humanities to strengthen the concept of professionalism among first-professional year pharmacy students. *American Journal of Pharmaceutical Education*, 7(2), 1-6
- Burke Johnson, R. & Gray, R. (2010). A history of philosophical and theoretical issues for mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (2nd ed., pp. 69-94). Thousand Oaks (CA), USA: SAGE
- Burke Johnson, R. & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. DOI: 10.3102/0013189X033007014
- Burke Johnson, R., Onwuegbuzie, A. J. & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133. DOI: 10.1177/1558689806298224
- Burns, A. & Danyluk, P. (2017). Applying Kolb's model to a nontraditional preservice teaching practicum. *Journal of Experiential Education*, 40(3), 249-263. DOI: 10.1177/1053825917696832
- Bursch, B., Fried, J. M., Wimmers, P. F., Cook, I. A., Baillie, S., Zackson, H. & Stuber, M. L. (2013). Relationships between medical student perceptions of mistreatment and mistreatment sensitivity. *Medical Teacher*, 35, e998-e1002. DOI: 10.3109/0142159X.2012.733455
- Burt, C. (1940). *The factors of the mind*. London, UK: University of London Press
- Burt, C. (1972). The reciprocity principle. In S. R. Brown & D. J. Brenner (Eds.), *Science, psychology, and communication: essays honouring William Stephenson* (pp. 39-56). New York (NY), USA: Teachers College Press
- Butler, R. (2009a). Coming to terms with personal construct theory. In R. Butler (Ed.), *Reflections in personal construct theory* (pp. 3-20). West Sussex, UK: Wiley
- Butler, R. (2009b). Preface. In R. Butler (Ed.), *Reflections in personal construct theory* (pp. XV-XXIV). West Sussex, UK: Wiley

- Cameron, R. (2011). Mixed methods research: the five Ps framework. *Electronic Journal of Business Research Methods*, 9(2), 96-108
- Campbell, E. G., Regan, S., Gruen, R. L., Ferris, T. G., Rao, S. R., Cleary, P. D. & Blumenthal, D. (2007). Professionalism in medicine: results of a national survey of physicians. *Annals of Internal medicine*, 147, 795-802
- Campbell, J. L., Richards, S. H., Dickens, A., Greco, M., Narayanan, A. & Brearley, S. (2008). Assessing the professional performance of UK doctors: an evaluation of the utility of the General medical Council patient and colleague questionnaires. *Quality & Safety in Health Care*, 17, 187-193. DOI: 10.1136/qshc.2007.024679
- Campbell, J. L., Roberts, M., Wright, C., Hill, J., Greco, M., Taylor, M. & Richards, S. (2011). Factors associated with variability in the assessment of UK doctors' professionalism: analysis of survey results. *BMJ*, 343, 1-11. DOI: 10.1136/bmj.d6212
- Candela, A. G. (2019). Exploring the function of member checking. *The Qualitative Report*, 24(3), 619-628
- Carey, J. R. & Ness, K. K. (2001). Erosion of professional behaviors in physical therapist students. *Journal of Physical Therapy Education*, 15(3), 20-22
- Carlson, B. & Glenton, C. (2011). What about N? A methodological study of sample-size reporting in focus group studies. *BMC Medical Research Methodology*, 11(26). DOI: 10.1186/1471-2288-11-26
- Carmona, S., Iyer, G. & Reckers, P. M. J. (2014). Performance evaluation bias: a comparative study on the role of financial fixation, similarity-to-self and likeability. *Advances in Accounting, incorporating Advances in International Accounting*, 30, 9-17. DOI: 10.1016/j.adiac.2014.04.001
- Carney, O., McIntosh, J. & Worth, A. (1996). The use of the nominal group technique in research with community nurses. *Journal of Advanced Nursing*, 23(5), 1024-1029. DOI: 10.1046/j.1365-2648.1996.09623.x
- Carter, M., Hesselgreaves, H., Rothwell, C., Crampton, P., Burford, B., McLachlan, J. & Illing, J. (2015). *Measuring professionalism as a multi-dimensional construct: professionalism and conscientiousness in healthcare professionals – study 2*. Retrieved from <https://www.hcpc-uk.org/resources/reports/2016/measuring-professionalism-as-a-multi->

[dimensional-construct---professionalism-and-conscientiousness-in-healthcare-professionals-study-2/](#)

- Castallani, B. & Hafferty, F. W. (2006). The complexities of medical professionalism. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 3-24). Boston (MA), USA: Springer
- Catania, G. & Randall, R. (2015). The repertory grid technique as a method for eliciting tacit knowledge. In F. Bezzina & V. Cassar (Eds.), *Proceedings of the 14th annual European conference on research methodology for business and management studies, University of Malta, Valletta* (pp. 104-110). Reading, UK: Academic Conferences and Publishing International
- Cattell, R. B. (1943). The description of personality: basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 38(4), 476-506. DOI: 10.1037/h0054116
- Cattell, R. B. (1945a). The description of personality: principles and findings in a factor analysis. *The American Journal of Psychology*, 58(1), 69-90. DOI: 10.2307/1417576
- Cattell, R. B. (1945b). The principal trait clusters for describing personality. *Psychological Bulletin*, 42(3), 129-161
- Cattell, R. B. (1965). *The scientific analysis of personality*. Chicago (IL), USA: Aldine Press
- Centre for Reviews and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in healthcare*. York, UK: University of York
- Chan, A. T., S., Chan, E. H. W. & Scott, D. (2007). Evaluation of Hall's professionalism scale for professionals in the construction industry. *Psychological Reports*, 100, 1201-1217
- Chang, S., van Witteloostuijn, A. & Eden, L. (2010). From the editors: common method variance in international business research. *Journal of International Business Studies*, 41, 178-184. DOI: 10.1057/jibs.2009.88
- Chin Fatt, C. R., Beaton, D. & Abdi, H. (2013). MExPosition: multi-table ExPosition [Computer software]. Retrieved from <https://CRAN.R-project.org/package=MExPosition>
- Chisholm, M. A., Cobb, H. Duke, L., McDuffie, C. & Kennedy, W. K. (2006). Instructional design and assessment: development of an instrument to

- measure professionalism. *American Journal of Pharmaceutical Education*, 70(4), 1-6
- Chisholm-Burns, M. A., Spivey, C. A., Jaeger, M. C., Williams, J. & George, C. (2017). Development of an instrument to measure pharmacy student attitudes toward social media professionalism. *American journal of Pharmaceutical Education*, 81(4), 1-10
- Cizek, G. J., Rosenberg, S. L. & Noons, H. H. (2008). Sources of validity evidence for educational tests. *Educational and Psychological Measurement*, 68(3), 397-412. DOI: 10.1177/0013164407310130
- Committee on Standards in Public Life (1995). *The 7 principles of public life*. Retrieved from <https://www.gov.uk/government/publications/the-7-principles-of-public-life>
- Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., ... Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the recovering quality of life measure (ReQoL). *Quality of Life Research*, 27, 1893-1902. DOI: 10.1007/s11136-018-1847-y
- Cook, D. A. & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine*, 119, 116.e7-116.e16. DOI: 10.1016/j.amjmed.2005.10.036
- Cook, L. L., Eignor, D. R. & Taft, H. L. (1988). A comparative study of the effects of recency of instruction on the stability of IRT and conventional item parameter statistics. *Journal of Educational Measurement*, 25(1), 31-45
- Cooper, C. (2010). *Individual differences and personality* (3rd ed.). London, UK: Routledge
- Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practice Assessment, Research & Evaluation*, 10(7), 1-9
- Coulehan, J. (2006). You say self-interest, I say altruism. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 103-128). Boston (MA), USA: Springer

- Crandell, C. E., Wiegand, M. R. & Broksy, J. A. (2013). Examining the role of service-learning on development of professionalism in doctor of physical therapy students: a case report. *Journal of Allied Health*, 42(1), e25-e32
- Creswell, J. W. (2010). Mapping the developing landscape of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioural research* (2nd ed., pp. 45-68). Thousand Oaks (CA), USA: SAGE
- Creswell, J. W. & Creswell, J. D. (2018). *Research design: qualitative, quantitative and mixed methods approaches* (5th ed.). Thousand Oaks (CA), USA: SAGE
- Creswell, J. W. & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks (CA), USA: SAGE
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York (NY), USA: Harper Collins
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioural measurements: theory of generalizability for scores and profiles*. New York (NY), USA: Wiley
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302
- Crossley, J. & Vivekananda-Schmidt, P. (2009). The development and evaluation of a professional self identity questionnaire to measure evolving professional self-identity in health and social care students. *Medical Teacher*, 31(12), e603-e607. DOI: 10.3109/01421590903193547
- Cruess, S. R. (2006). Professionalism and medicine's social contract with society. *Clinical Orthopaedics and Related Research*, 449, 170-176. DOI: 10.1097/01.blo.0000229275.66570.97
- Cruess, R. L. & Cruess, S. R. (1997). Teaching medicine as a profession in the service of healing. *Academic Medicine*, 72, 941-952
- Cruess, R. L., Cruess, S. R. & Steinhert, Y. (2016). Amending Miller's pyramid to include professional identity formation. *Academic Medicine*, 91(2), 180-185. DOI: 10.1097/ACM.0000000000000913
- Cruess, R., McIlroy, J. H., Cruess, S., Ginsburg, S. & Steinert, Y. (2006). The professionalism mini-evaluation exercise: a preliminary investigation.

- Academic Medicine*, 81(10), S74-S78. DOI: 10.1097/00001888-200610001-00019
- Cyr, J. (2019). *Focus groups for the social science researcher*. Cambridge, UK: Cambridge University Press
- de Haas, J. C. J. M., Oort, F., Oosterveld, P. & ten Cate, O. (2001). Assessment of medical students' communicative behaviour and attitudes: estimating the reliability of the use of the Amsterdam attitudes and communication scale through generalisability coefficients. *Patient Education and Counseling*, 45, 35-42
- de Mendonça, E. T., Cotta, R. M. M., Lelis, V. de P. & Carvalho Junior, P. M. (2016). Assessment of professionalism in student of health-related courses: a systematic review. *Interface - Comunicação, Saúde, Educação*, 20(58), 679-690. DOI: 10.1590/1807-57622015.0274
- de Ruyter, K. (1996). Focus versus nominal group interviews: a comparative analysis. *Marketing Intelligence & Planning*, 14(6), 44-50. DOI: 10.1108/02634509610131153
- Delbecq, A. L., Van de Ven, A. H. & Gustafson, D. H. (1975). *Group techniques for program planning: a guide to nominal group and delphi processes*. Glenview (IL), USA: Scott, Foresman and Company
- Denzin, N. K. & Lincoln, Y. S. (2008). Introduction: the discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research* (3rd ed., pp. 1-44). Los Angeles (CA), USA: SAGE
- Dickson, R., Cherry, M. G. & Boland, A. (2017). Carrying out a systematic review as a Master's thesis. In A. Boland, M. G. Cherry & R. Dickson (Eds.), *Doing a systematic review: a student's guide* (pp. 1-20). London, UK: SAGE
- Doreian, P. (2004). Cluster analysis. In M. S. Lewis-Beck, A. Bryman & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods* (pp. 129-130). Thousand Oaks (CA), USA: SAGE
- Dougherty, M. R., Slevc, L. R. & Gran, J. A. (2019). Making research evaluation more transparent: aligning research philosophy, institutional values, and reporting. *Perspectives on Psychological Science*, 14(3), 361-375. DOI: 10.1177/1745691618810693

- Doukas, D. J. (2006). Professionalism: curriculum goals and meeting their challenges. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 43-59). New York (NY), USA: Springer
- du Toit, D. (1995). A sociological analysis of the extent and influence of professional socialization on the development of nursing identity among nursing students at two universities in Brisbane, Australia. *Journal of Advanced Nursing*, 21, 134-171
- Duckor, B. (2017). Beyond immanent and transcendent critique: exploring Maul's argument within the NRC (2001). *Measurement: Interdisciplinary Research and Perspectives*, 15(2), 98-102. DOI: 10.1080/15366367.2017.1369812
- Dwyer, T., Takahashi, S. G., Hynes, M., Herold, J., Wasserstein, D., Nousiainen, M. ... Ogilvie-Harris, D. (2014). How to assess communication, professionalism, collaboration and the other intrinsic CanMEDS roles in orthopedic residents: use of an objective structured clinical examination (OCSE). *The Journal of the Canadian Chiropractic Association*, 57(4), 230-236. DOI: 10.1503/cjs.018813
- Dyrbye, L. N., Harper, W., Moutier, C., Durning, S. J., Power, D. V., Massie, S. ... Shanafelt, T. D. (2012) A multi-institutional study exploring the impact of positive mental health on medical students' professionalism in an era of high burnout. *Academic Medicine*, 87, 1024-1031. DOI: 10.1097/ACM.0b013e31825cfa35
- Dyrbye, L. N., Massie, F. S., Eacker, A., Harper, W., Power, D., Durning, S. J. ... Shanafelt, T. D. (2010). Relationship between burnout and professional conduct and attitudes among US medical students. *JAMA*, 304(11), 1173-1180
- Easterby-Smith, M. (1980). How to use repertory grids in HRD. *Journal of European Industrial Training*, 4(2), 2-32. DOI: 10.1108/eb002339
- Easterby-Smith, M., Thorpe, R. & Holman, D. (1996). Using repertory grids in management. *Journal of European Industrial Training*, 20(3), 3-30. DOI: 10.1108/03090599610114138
- Eccleston, C., De, A. C., Williams, C. & Stainton Rogers, W. (1997). Patients' and professionals' understandings of the causes of chronic pain: blame,

- responsibility and identity protection. *Social Science & Medicine*, 45(5), 699-709. DOI: 10.1016/S0277-9536(96)00404-2
- Embretson, S. E. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin*, 93(1), 179-197. DOI: 10.1037/0033-2909.93.1.179
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: application to abstract reasoning. *Psychological Methods*, 3(3), 380-396. DOI: 10.1037/1082-989X.3.3.380
- Embretson, S. E. (2004). The second century of ability testing: some predictions and speculations. *Measurement*, 2(1), 1-32
- Embretson, S. & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38(4), 343-368
- Eva, K. W., Rosenfeld, J., Reitter, H. I. & Norman, G. R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education*, 38, 314-326. DOI: 10.1046/j.1365-2923.2004.01776.x
- Evans, L. (2008). Professionalism, professionalism and the development of education professionals. *British journal of Educational Studies*, 56(1), 20-38. DOI: 10.1111/j.1467-8527.2007.00392.x
- Everitt, B. S., Landau, S. & Leese, M. (2001). *Cluster analysis* (4th ed.). London, UK: Arnold
- Evers, A., Hagemester, C., Høstmælingen, A., Lindley, P., Muñiz, J. & Sjöberg, A. (2013). *EFPA review model for the description and evaluation of psychological and educational tests*. Brussels, Belgium: European Federation of Psychologists' Associations
- Evers, A., Sijtsma, K., Luassen, W. & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: history, procedure, and results. *International Journal of Testing*, 10(4), 295-317. DOI: 10.1080/15305058.2010.518325
- Evetts, J. (2003). The sociological analysis of professionalism: occupational change in the modern world. *International sociology*, 18(2), 395-415
- Evetts, J. (2006). Trust and professionalism: challenges and occupational changes. *Current Sociology*, 54(4), 515-531. DOI: 10.1177/0011392106065083

- Evetts, J. (2014). The concept of professionalism: professional work, professional practice and learning. In S. Billet, C. Harteis & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 29-56). Dordrecht, The Netherlands: Springer
- Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381
- Fassin, Y. & Van Rossem, A. (2009). Corporate governance in the debate on CSR and ethics: sensemaking of social issues in management by authorities and CEOs. *Corporate Governance: An International Review*, 17(5), 573-593. DOI: 10.1111/j.1467-8683.2009.00738.x
- Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), 6-16. DOI: 10.1177/1558689809349691
- Financial Conduct Authority (2019). *Statements of principle and code of practice for approved persons*. Retrieved from <https://www.handbook.fca.org.uk/handbook/APER.pdf>
- Finn, G., Garner, J. & Sawdon, M. (2010). 'You're judged all the time!' Students' views on professionalism: a multicentre study. *Medical Education*, 44, 814-825. DOI: 10.1111/j.1365-2923.2010.03743.x
- Fischer, R. & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research*, 3(1), 88-96. DOI: 10.21500/20112084.852
- Fisher Jr., W. P. (2017). Suggestions for rethinking validation. *Measurement: Interdisciplinary Research and Perspectives*, 15(2), 86-90. DOI: 10.1080/15366367.2017.1369778
- Fochtmann, L. J. (2006). Professionalism and the Heisenberg Uncertainty Principle. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 233-254). New York (NY), USA: Springer
- Forster, J. R. (1992). Eliciting personal constructs and articulating goals. *Journal of Career Development*, 18(3), 175-185
- Fransella, F., Bell, R. & Bannister, D. (2004). *A manual for repertory grid technique* (2nd ed.). West Sussex, UK: Wiley

- Freidson, E. (1994). *Professionalism reborn: theory, prophecy, and policy*.
Cambridge, UK: Polity Press
- Freidson, E. (2001). *Professionalism: the third logic*. Cambridge, UK: Polity Press
- Friedman, V. J. & Rogers, T. (2009). There is nothing so theoretical as good action research. *Action Research*, 7(1), 31-47. DOI: 10.1177/1476750308099596
- Frohna, A. (2006). Medical students' professionalism. *Medical Teacher*, 28(1), 1-2.
DOI: 10.1080/01421590600568280
- Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. Los Angeles (CA), USA: SAGE
- Furr, R. M. & Bacharach, U. R. (2014). *Psychometrics: an introduction* (2nd ed.).
Thousand Oaks (CA), USA: SAGE
- Gallagher, K. & Porock, D. (2010). The use of interviews in Q methodology.
Nursing Research, 59(4), 295-300
- Galton, F. (1884). Measurement of character. *Fortnightly Review*, 42, 179-185
- Garin O. (2014) Ceiling Effect. In A. C. Michalos (Ed.) *Encyclopedia of Quality of Life and Well-Being Research* (pp. 631-633). Dordrecht, The Netherlands: Springer
- Gauttier, S. E. J. (2017). *Making (inter)subjectivity transparent in phenomenological research, an illustration using Q-method*. Paper presented at the 33rd Annual Conference of the International Society for the Scientific Study of Subjectivity, Glasgow: UK
- General Dental Council (2013). *Standards for the dental team*. Available from <https://standards.gdc-uk.org/>
- General Medical Council (2013). *Good medical practice*. Available from <https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice>
- General Medical Council (2016). *Achieving good medical practice: guidance for medical students*. Available from <https://www.gmc-uk.org/education/standards-guidance-and-curricula/guidance/achieving-good-medical-practice>
- Gill, A. C., Nelson, E. A., Mian, A. I., Raphael, J. L., Rowley, D. R. & McGuire, A. L. (2015). Responding to moderate breaches in professionalism: an

- intervention for medical students. *Medical Teacher*, 37(2), 136-139. DOI: 10.3109/0142159X.2014.911270
- Gillespie, C., Paik, S., Ark, T., Zabar, S. & Kalet, A. (2009). Residents' perceptions of their own professionalism and the professionalism of their learning environment. *Journal of Graduate Medical Education*, 1(2), 208-215. DOI: 10.4300/JGME-D-09-00018.1
- Ginsberg, A. (1989). Construing the business portfolio: a cognitive model of diversification. *Journal of Management Studies*, 26(4), 417-438. DOI: 10.1111/j.1467-6486.1989.tb00737.x
- Girden, E. R. & Kabacoff, R. I. (2011). *Evaluating research articles: from start to finish*. Thousand Oaks (CA), USA: SAGE
- Giroux, H. A. & Penna, A. N. (1979). Social education in the classroom: the dynamics of the hidden curriculum. *Theory and Research in Social Education*, VII(1), 21-42
- Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. Hawthorne (NY), USA: Aldine Press
- Gleeson, P. B. (2007). Understanding generational competence related to professionalism: misunderstandings that lead to a perception of unprofessional behaviour. *Journal of Physical Therapy Education*, 21(3), 23-28
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26-42
- Goldie, J. (2013). Assessment of professionalism: a consolidation of current thinking. *Medical Teacher*, 35, e952-e956. DOI: 10.3109/0142159X.2012.714888
- Goleman, D. (1995). *Emotional intelligence: why it can matter more than IQ*. London, UK: Bloomsbury
- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213-228. DOI: 10.1080/02671520701296189
- Gough, D., Oliver, S. & Thomas, J. (2012). *Introduction to systematic reviews*. London, UK: SAGE

- Grant, K. A. (2007). Tacit knowledge revisited – we can still learn from Polanyi. *The Electronic Journal of Knowledge Management*, 5(2), 173-180
- Green, C. D., (1992). Of immortal mythological beasts: operationism in psychology. *Theory & Psychology*, 2(3), 291-320
- Greene, J. C. & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In S. Tashakkori & C. Teddlie (Eds.), *Handbook for mixed methods in social and behavioural research* (pp. 91-110). Thousand Oaks (CA), USA: SAGE
- Grewal, D. & Salovey, P. (2005). Feeling smart: the science of emotional intelligence. *American Scientist*, 93, 330-339
- Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 17-28). Newbury Park (CA), USA: SAGE
- Guba, E. G. & Lincoln, Y. S. (2008). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research* (3rd ed., pp. 255-286). Los Angeles (CA), USA: SAGE
- Guest, G., Namey, E. & McKenna, K. (2016). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, 29(1), 3-22. DOI: 10.1177/1525822X16639015
- Guion, R. M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11(3), 385-398
- Gullisken, H. (1950). Intrinsic validity. *The American Psychologist*, 5(10), 511-517. DOI: 10.1037/h0054604
- Hafferty, F. W. (1998). Beyond curriculum reform: confronting medicine's hidden curriculum. *Academic Medicine*, 73, 403-407
- Hafferty, F. W. (2006). Definitions of professionalism: a search for meaning and identity. *Clinical Orthopaedics and Related Research*, 449, 193-204. DOI: 10.1097/01.blo.0000229273.20829.d0
- Hafferty, F. W. & Franks, R. (1994). The hidden curriculum, ethics teaching, and the structure of medical education. *Academic Medicine*, 69, 861-871
- Hall, J. & Ashcroft, D. (2011). What characterises professionalism in pharmacy students? A nominal group study. *Pharmacy Education*, 11(1), 65-70

- Halpin, P. F. (2017). How can we improve the measurement of “non-academic” competencies? *Measurement: Interdisciplinary Research and Perspectives*, 15(2), 70-72. DOI: 10.1080/15366367.2017.1367245
- Hambleton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 38-47
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park (CA), USA: SAGE
- Hammer, D. (2006). Improving student professionalism during experiential learning. *American Journal of Pharmaceutical Education*, 70(3), 1-6
- Han, J., Kamber, M. & Pei, J. (2011). *Data mining: concepts and techniques*. Waltham (MA), USA: Elsevier
- Hare, A. P. (1952). A study of interaction and consensus in different sized groups. *American Sociological Review*, 17(3), 261-267. DOI: 10.2307/2088071
- Harnett, J. (2012). Reducing discrepancies between teachers’ espoused theories and theories-in-use: an action research model of reflective professional development. *Educational Action Research*, 20(3), 367-384. DOI: 10.1080/09650792.2012.697397
- Harvey, L. (2015). Beyond member-checking: a dialogic approach to the research interview. *International Journal of Research & Method in Education*, 38(1), 23-38. DOI: 10.1080/1743727X.2014.914487
- Harvey, N. & Holmes, C. A. (2012). Nominal group technique: an effective method for obtaining group consensus. *International Journal of Nursing Practice*, 18(2), 188-194. DOI: 10.1111/j.1440-172X.2012.02017.x
- Hassall Thomsen, L. (2015). Silvio Waisboard: reinventing professionalism: journalism and news in a global perspective. *Journal of Media and Communication Research*, 31(58), 150-153. DOI: 10.7146/mediekultur.v31i58.18079
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164. DOI: 10.1177/014662168500900204
- Hendelman, W. & Byszewski, A. (2014). Formation of medical student professional identity: categorizing lapses of professionalism, and the learning

- environment. *BMC Medical Education*, 1(139). DOI: 10.1186/1472-6920-14-139
- Hennink, M. M. (2014). *Understanding focus group discussions*. New York (NY), USA: Oxford University Press
- Heru, A., Gagne, G. & Strong, D. (2009). Medical student mistreatment results in symptoms of posttraumatic stress. *Academic Psychiatry*, 33(4), 302-306
- Hicks, L. K., Lin, Y., Robertson, D. W., Robinson, D. L. & Woodrow, S. I. (2001). Understanding the clinical dilemmas that shape medical students' ethical development: questionnaire survey and focus group study. *BMJ*, 322, 709-710. DOI: 10.1136/bmj.322.7288.709
- Higgins J. P. T. & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Available from www.handbook.cochrane.org
- Hill, K., Wittkowski, A., Hodgkinson, E., Bell, R. & Hare, D. J. (2016). Using the repertory grid technique to examine trainee clinical psychologists' construal of their personal and professional development. *Clinical Psychology and Psychotherapy*, 23, 425-437. DOI: 10.1002/cpp.1961
- Hinojosa, A. S., Gardner, W. L., Walker, H. J., Coglisier, C. & Gullifor, D. (2017). A review of cognitive dissonance theory in management research: opportunities for further development. *Journal of Management*, 43(1), 170-199. DOI: 10.1177/0149206316668236
- Ho, M. J. (2011). Does one size fit all? Building a framework for medical professionalism. *Academic Medicine*, 86(11), 1407-1414. DOI: 10.1097/ACM.0b013e31823059d1
- Hodges, D., McLachlan, J. C. & Finn, G. M. (2009). Exploring reflective 'critical incident' documentation of professionalism lapses in a medical undergraduate setting. *BMC Medical Education*, 9, 44. DOI: 10.1186/1472-6920-9-44
- Hogan, T. P. & Agnello, J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational and Psychological Measurement*, 64(4), 802-812. DOI: 10.1177/0013164404264120
- Honicke, T. & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: a systematic review. *Educational Research Review*, 17, 63-84. DOI: 10.1016/j.edurev.2015.11.002

- Houchens, G., Hurt, J., Stobaugh, R. & Keedy, J. L. (2012). Double-loop learning: a coaching protocol for enhancing principal instructional leadership. *Qualitative Research in Education, 1*(2), 135-178. DOI: 10.4471/qre.2012.08
- Hoyle, E. (2001). Teaching: prestige, status and esteem. *Educational Management & Administration, 29*(2), 139-152. DOI: 10.1177/0263211X010292001
- Humphrey-Murto, S., Varpio, L., Gonsalves, C. & Wood, T. J. (2017). Using consensus group methods such as delphi and nominal group in medical education research. *Medical Teacher, 39*(1), 14-19. DOI: 10.1080/0142159X.2017.1245856
- Humphris, G. M. & Kaney, S. (2000). The objective structured video exam for assessment of communication skills. *Medical Education, 34*, 939-945. DOI: 10.1046/j.1365-2923.2000.00792.x
- Humphry, S. M. (2011). The role of the unit in physics and psychometrics. *Measurement, 9*, 1-24. DOI: 10.1080/15366367.2011.558442
- Humphry, S. M. (2017). Psychological measurement: theory, paradoxes, and prototypes. *Theory & Psychology, 27*(3), 407-418. DOI: 10.1177/0959354317699099
- Hurst, Y. K., Precott-Clements, L. E. & Rennie, J. S. (2004). The patient assessment questionnaire: a new instrument for evaluating the interpersonal skills of vocational dental practitioners. *British Dental Journal, 197*(8), 497-500. DOI: 10.1038/sj.bdj.4811750
- Hutchings, H. A., Rapport, F. L., Wright, S. & Doel, M. A. (2013). Obtaining consensus from mixed groups: an adapted nominal group technique. *British Journal of Medicine & Medical Research, 3*(3), 491-502. DOI: 10.9734/BJMMR/2013/2625
- Igundunasse, A. (2016). The impact of small samples sizes in factor analytical studies. *IFE Psychologia: An International Journal, 24*(1), 54-65
- Iivari, N. (2018). Using member checking in interpretive research practice: a hermeneutic analysis of informants' interpretation of their organizational realities. *Information Technology & People, 31*(1), 111-133. DOI: 10.1108/ITP-07-2016-0168

- Information Commissioner's Office (n.d.). *Information commissioner's office code of conduct*. Retrieved from https://ico.org.uk/media/about-the-ico/policies-and-procedures/1845/code_of_conduct.pdf
- Inui, T. S., Cottingham, R. M., Frankel, D. K., Litzelman, D. L., Mossbarger, A. L., Suchman, T. ... Williamson, P. R. (2006). Educating for professionalism at Indiana University School of Medicine: feet on the ground and fresh eyes. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 165-184). New York (NY), USA: Springer
- Jankowicz, D. (2004). *The easy guide to repertory grids*. West Sussex, UK: Wiley
- Jellicoe, M. & Forsythe, A. (2019). The development and validation of the feedback learning scale (FLS). *Frontiers in Education*, 4, 84. DOI: 10.3389/feduc.2019.00084
- Jha, V., Bekker, H. L., Duffy, S. R. G. & Roberts, T. E. (2007). A systematic review of studies assessing and facilitating attitudes towards professionalism in medicine. *Medical Education*, 41, 822-829. DOI: 10.1111/j.1365-2923.2007.02804.x
- John, O. P., Angleitner, A. & Ostendorf, F. (1988). The lexical approach to personality: a historical review of trait taxonomical research. *European Journal of Personality*, 2, 171-203. DOI: 10.1002/per.2410020302
- John, O. P. & Srivastava, S. (1999). The big five trait taxonomy; history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102-138). New York (NY), USA: Guildford
- Jones, L. V. & Thissen, D. (2007). A history of psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26: psychometrics* (pp. 1-28). Amsterdam, The Netherlands: Elsevier Science B. V
- Jonson, J. L. & Plake, B. S. (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*, 58(5), 736-753
- Jordan, K., Capdevila, R. & Johnson, S. (2005). Baby or beauty: a Q study into post pregnancy body image. *Journal of Reproductive and Infant Psychology*, 23(1), 19-31. DOI: 10.1080/02646830512331330965

- Kačániová, M. & Szabová, E. (2014). Perception of international web pages: a repertory grid approach. *Studia Ekonomiczne*, 205, 59-71
- Kane, M. (2004). Certification testing as an illustration of arguments-based validation. *Measurement*, 2(3), 135-170. DOI: 10.1207/s15366359mea0203_1
- Kane, M. T. (1990). *An argument-based approach to validation*. Retrieved from <https://eric.ed.gov/?id=ED336428>
- Kane, M. T. (2001). *Current concerns in validity theory*. Retrieved from <https://eric.ed.gov/?id=ED446094>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73
- Karnieli-Miller, O., Vu, R, Holtman, M., Clyman, S. G. & Inui, T. S. (2010). Medical students' professionalism narratives: a window on the informal and hidden curriculum. *Academic Medicine*, 85, 124-133
- Karukivi, M., Korteganas-Savolainen, O., Saxén, U. & Haapasalo-Pesu, K. (2015). Professionalism mini-evaluation exercise in Finland: a preliminary investigation introducing the Finnish version of the P-MEX instrument. *Journal of Advances in Medical Education & Professionalism*, 3(4), 154-158
- Kaufman, T. J. & Ricci, P. (2014). Creation of a professionalism scale for hospitality students: an exploratory study. *SAGE Open*, 1-6. DOI: 10.1177/2158244014559016
- Kelle, U. & Erzberger, C. (2004). Qualitative and quantitative methods: not in opposition. In U. Flick, E. von Kardoff & I. Steinke (Eds.), *A companion to qualitative research* (pp. 172-177). London, UK: SAGE
- Kelley, K. A., Stanke, L. D., Rabi, S. M., Kuba, S. E. & Janke, K. K. (2011). Cross-validation of an instrument for measuring professionalism behaviors. *American Journal of Pharmaceutical Education*, 75(9), 1-10
- Kelly, G. A. (1955). *The psychology of personal constructs, volume 1: A theory of personality*. New York (NY), USA: Norton
- Kenny, N. P., Mann, K. V. & MacLeod, H. (2003). Role modelling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Academic Medicine*, 78, 1203-1210. DOI: 10.1097/00001888-200312000-00002

- Khan, Y., Fazli, G., Henry, B., de Villa, E., Tasamis, C., Grant, M. & Schwartz, B. (2015). The evidence base of primary research in public health emergency preparedness: a scoping review and stakeholder consultation. *BMC Public Health*, 15, 432. DOI: 10.1186/s12889-015-1750-1
- Kim, S. S., Kaplowitz, S. & Johnston, M. V. (2004). The effects of physician empathy on patient satisfaction and compliance. *Evaluation & the Health Professions*, 27(3), 237-251. DOI: 10.1177/0163278704267037
- Klemenc-Ketis, Z. & Vrecko, H. (2014). Development and validation of a professionalism assessment scale for medical students. *International Journal of Medical Education*, 5, 205-211. DOI: 10.5116/ijme.544b.7972
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London, UK: Routledge
- Kline, P. (1994). *An easy guide to factor analysis*. London, UK: Routledge
- Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*. Upper Saddle River (NJ), USA: Prentice-Hall
- Kolb, D. A., Boyatzis, R. E. & Mainemelis, C. (2001). Experiential learning theory: previous research and new directions. In R. J. Sternberger & L. Zhang (Eds.), *Perspectives on thinking, learning, and cognitive styles* (pp. 227-247). Mahwah (NJ), USA: Lawrence Erlbaum Associates
- Kolb, A. Y. & Kolb, D. A. (2005). Learning styles and learning spaces: enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4(2), 193-212. DOI: 10.5465/AMLE.2005.17268566
- Krauss, S. R., Idris, K., Tamam, E., Suandi, T., Ismail, I. A., Bandar, N. F. A. & Dahalan, D. (2012). Exploring professionalism among youth work practitioners in Malaysia: a measurement development study. *Young* 20(3), 297-322. DOI: 10.1177/110330881202000305
- Krosnick, J. A., Judd, C. M. & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change*. Mahwah (NJ), USA: Erlbaum
- Krueger, R. A. & Casey, M. A. (2015). *Focus groups: a practical guide for applied research* (5th ed.). Thousand Oaks (CA), USA: SAGE
- Kuczewski, M. (2006). The problem with evaluating professionalism: the case against the current dogma. In D. Wear & J. M. Aultman (Eds.),

- Professionalism in medicine: critical perspectives* (pp. 185-198). New York (NY), USA: Springer
- Kulac, E., Sezik, M., Asci, H. & Doguc, D. K. (2013). Medical students' participation in and perception of unprofessional behaviors: comparison of preclinical and clinical phases. *Advances in Physiology Education*, 37(4), 298-302. DOI: 10.1152/advan.00076.2013
- Langford, B. E., Schoenfeld, G. & George, I. (2002). Nominal grouping sessions vs focus groups. *Qualitative Market Research: An International Journal*, 5(1), 58-70. DOI: 10.1108/13522750210414517
- Laughlin, P. R., Bonner, B. L. & Miner, A. G. (2006). Groups perform better than the best individuals on letter-to-numbers problems: effects of group size. *Organizational Behavior and Human Decision Processes*, 88, 605-620
- Lees-Haley, P. R. (1996). Alice in validityland, or the dangerous consequences of consequential validity. *American Psychologist*, 51(9), 981-983. DOI: 10.1037/0003-066X.51.9.981
- Lennon, R. T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement*, 16(3), 294-304. DOI: 10.1037/0003-066X.51.9.981
- Lewis, B. (2006). Medical professionals and the discourse of professionalism: teaching implications. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 149-161). New York (NY), USA: Springer
- Li, H., Ding, N., Zhang, Y., Liu, Y. & Wen, D. (2017). Assessing medical professionalism: a systematic review of instruments and their measurement properties. *PLoS ONE*, 12(5), e0177321. DOI: 10.1371/journal.pone.0177321
- Lie, D., Trial, J., Schaff, P., Wallace, R. & Elliott, D. (2013). "Being the best we can be": medical students' reflections on physician responsibility in the social media era. *Academic Medicine*, 88(2), 240-245. DOI: 10.1097/ACM.0b013e31827bf5fc
- Lievens, F. & Coetsier, P. (2002). Situational tests in student selection: an examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, 10(4), 245-257. DOI: 10.1111/1468-2389.00215

- Livianiere, V. & De Raad, B. (2016). The factor structure of Lithuanian personality-descriptive adjectives of the highest frequency of use. *International Journal of Psychology*. DOI: 10.1002/ijop.12247
- Lloyd, S. (2011). Applying the nominal group technique to specify the domain of a construct. *Qualitative Market Research*, 14(1), 105-121. DOI: 10.1108/13522751111099346
- Lombarts, K. M. J. M. H., Plochg, T., Thompson, C. A. & Arah, O. A. on behalf of the DUQuE Project Consortium (2014). Measuring professionalism in medicine and nursing: results of a European survey. *PLoS ONE*, 9(5), e97069. DOI: 10.1371/journal.pone.0097069
- Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I. & Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and Social Psychology Bulletin*, 30(8), 1018-1034. DOI: 10.1177/0146167204264762
- Lopes, P. N., Salovey, P., Côté, S. & Beers. M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion*, 5(1), 113-118. DOI: 10.1037/1528-3542.5.1.113
- Lopes, P. N., Salovey, P. & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, 35(3), 641-658. DOI: 10.1016/S0191-8869(02)00242-8
- Lord, F. M. (1952). *A theory of test scores (psychometric monograph no. 7)*. Richmond (VA), USA: Psychometric Corporation. Retrieved from <http://www.psychometrika.org/journal/online/MN07.pdf>
- Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13(4), 517-549. DOI: 10.1177/001316445301300401
- Lumsden, J. (1976). Test Theory. *Annual Review of Psychology*, 27(1), 251-280
- Lynch, D. C., Surdyk, P. M. & Eiser, A. R. (2004). Assessing professionalism: a review of the literature. *Medical Teacher*, 26(4), 366-373. DOI: 10.1080/01421590410001696434
- MacDonald, P. & Paunonan, S. V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62(6), 921-943. DOI: 10.1177/0013164402238082

- Madill, A. & Sullivan, P. (2018). Mirrors, portraits, and member checking: managing difficult moments of knowledge exchange in the social sciences. *Qualitative Psychology*, 5(3), 321-339. DOI: 10.1037/qup0000089
- Manners Jr., G. E. (1975). Another look at group size, group problem solving, and member consensus. *Academy of Management Journal*, 18(4), 715-724. DOI: 10.2307/255374
- Marchalik, D. (2015). The art of medicine: saving the professionalism course. *The Lancet*, 385, 2346-2347
- Marei, H. F., Al-Eraky, M. M., Almasoud, N. N., Donkers, J. & Van Merriënboer, J. J. G. (2018). The use of virtual patient scenarios as a vehicle for teaching professionalism. *European Journal of Dental Education*, 22, e253-e260. DOI: 10.1111/eje.12283
- Markus, K. A. & Borsboom, D. (2013). *Frontiers of test validity theory: measurement, causation, and meaning*. New York (NY), USA: Routledge
- Mason, H., Collins, M., McHugh, N., Godwin, J., van Exel, J., Donaldson, C. & Baker, R. (2018). Is “the end of life” a special case? Connecting Q with survey methods to measure societal support for views on the value of life-extending treatments. *Health Economics*, 1-13. DOI: 10.1002/hec.3640
- Mason, H., van Exel, J., Baker, R., Brouwer, W., Donaldson, C. & EuroVaQ team (2016). From representing views to representativeness of views: illustrating a new (Q2S) approach in the context of health care priority setting in nine European countries. *Social Science & Medicine*, 166, 205-213. DOI: 10.1016/j.socscimed.2016.08.036
- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*, 4, 1-13. DOI: 10.3389/fpsyg.2013.00169
- Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, 15(2), 51-69. DOI: 10.1080/15366367.2017.1348108
- Maul, A., Irribarra, D. T. & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, 79, 311-320. DOI: 10.1016/j.measurement.2015.11.001
- May, T. & Perry, B. (2017). *Reflexivity: the essential guide*. London, UK: SAGE

- Mayer, J. D., Caruso, D. R. & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27(4), 267-298. DOI: 10.1016/S0160-2896(99)00016-1
- Mayer, J. D., Caruso, D. R. & Salovey, P. (2016). The ability model of emotional intelligence: principles and updates. *Emotion Review*, 8(4), 290-300. DOI: 10.1177/1754073916639667
- Mayer, J. D., DiPaolo, M. & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: a component of emotional intelligence. *Journal of Personality Assessment*, 54(3-4), 772-781. DOI: 10.1080/00223891.1990.9674037
- Mayer, J. D., Perkins, D. M., Caruso, D. R. & Salovey, P. (2001). Emotional intelligence and giftedness. *Roeper Review*, 23(3), 131-137
- Mayer, J. D. & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, 17(4), 433-442. DOI: 10.1016/0160-2896(93)90010-3
- Mayer, J. D. & Salovey, P. (1995). Emotional intelligence and the construction and regulation of feelings. *Applied and Preventive Psychology*, 4(3), 197-208. DOI: 10.1016/S0962-1849(05)80058-7
- Mayer, J. D., Salovey, P. & Caruso, D. R. (2004a). A further consideration of the issues of emotional intelligence. *Psychological Inquiry*, 15(3), 249-255
- Mayer, J. D., Salovey, P. & Caruso, D. R. (2004b). Emotional intelligence: theory, findings, and implications. *Psychological Inquiry*, 15(3), 197-215
- Mayer, J. D., Salovey, P. & Caruso, D. R. (2008). Emotional intelligence: new ability or eclectic traits? *American Psychologist*, 63(6), 503-517. DOI: 10.1037/0003-066X.63.6.503
- Mayer, J. D., Salovey, P., Caruso, D. R. & Cherkasskiy, L. (2011). Emotional intelligence. In R. J. Sternberg & S. D Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 528-549). New York (NY), USA: Cambridge University Press
- Mayer, J. D., Salovey, P., Caruso, D. R. & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, 3(1), 97-105. DOI: 10.1037/1528-3542.3.1.97
- Mazor, K. M., Zanetti, M. L., Alper, E. J., Hatem, D., Barrett, S. V., Meterko, V. ... Pugnaire, M. P. (2007). Assessing professionalism in the context of an

- objective structured clinical examination: an in-depth study of the rating process. *Medical Teacher*, 41, 331-340. DOI: 10.1111/j.1365-2929.2006.02692.x
- McGrane, J. A. (2015). Stevens' forgotten crossroads: the divergent measurement traditions in the physical sciences from the mid-twentieth century. *Frontiers in Psychology*, 6, 431. DOI: 10.3389/fpsyg.2015.00431
- McKeown, B. & Thomas, D. B. (2013). *Q methodology* (2nd ed.). California (CA), USA: SAGE
- McLachlan, J. C., Finn, G. & Macnaughton, J. (2009). The conscientiousness index: a novel tool to explore students' professionalism. *Academic Medicine*, 84, 559-565
- McMillan, S. S., Kelly, F., Sav, A., Kendall, E., King, M. A., Whitty, J. A. & Wheeler, A. J. (2014). Using the nominal group technique: how to analyse across multiple groups. *Health Services and Outcomes Research Methodology*, 14, 92-108. DOI: 10.1007/s10742-014-0121-1
- McMillan, S. S., Kelly, F., Sav, A., Kendall, E., King, M. A., Whitty, J. A. & Wheeler, A. J. (2015). Consumers and carers versus pharmacy staff: do their priorities for Australian pharmacy services align? *Patient*, 8, 411-422. DOI: 10.1007/s40271-014-0105-9
- McMillan, S. S., King, M. & Tully, M. P. (2016). How to use the nominal group and delphi techniques. *International Journal of Clinical Pharmacy*, 38, 655-662. DOI: 10.1007/s11096-016-0257-x
- McNeil, D. E., Sandberg, D. A. & Binder, R. L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence. *Law and Human Behavior*, 22(6), 655-669. DOI: 10.1023/A:1025754706716
- McNeill, F. (2001). Developing effectiveness: frontline perspectives. *Social Work Education*, 20(6), 671-687. DOI: 10.1080/02615470120089843
- McWilliams, S. A. (2009). Williams James' pragmatism and PCP. *Personal Construct Theory & Practice*, 6, 109-119
- Mendenhall, A. N., Iachini, A. & Anderson-Butcher, D. (2013). Exploring stakeholder perceptions of facilitators and barriers to implementation of an

- expanded school improvement model. *Children & Schools*, 35(4), 225-234.
DOI: 10.1093/cs/cdt011
- Meng, L., Metro, D. G. & Patel, R. M. (2009). Evaluating professionalism and interpersonal and communication skills: implementing a 360-degree evaluation instrument in an anesthesiology residency program. *Journal of Graduate Medical Education*, 216-220. DOI: 10.4300/JGME-D-09-00014.1
- Mertens, D. M. (2007). Transformative paradigm: mixed methods and social justice. *Journal of Mixed Methods Research*, 1(3), 212-225. DOI: 10.1177/1558689807302811
- Messick, S. (1979). *Test validity and the ethics of assessment*. Princeton (NJ), USA: Educational Testing Service
- Messick, S. (1990). *Validity of test interpretation and use*. Princeton (NJ), USA: Educational Testing Service
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. DOI: 10.1037/0003-066X.50.9.741
- Michell, J. (1997). Quantitative science and the definition of *measurement* in psychology. *British Journal of Psychology*, 88, 355-383
- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology*, 10(5), 369-667
- Mid Staffordshire NHS Foundation Trust Public Inquiry (2013). *Report of the Mid Staffordshire NHS Foundation Trust public inquiry: executive summary*. Retrieved from <https://webarchive.nationalarchives.gov.uk/20150407084231/http://www.midstaffpublicinquiry.com/report>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dea, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 3111-3119
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), S63-S67
- Miller, L. A. & Lovler, R. L. (2016). *Foundations of psychological testing: a practical approach* (5th ed.). Thousand Oaks (CA), USA: SAGE

- Miller, S. (2006). Mixed methods as methodological innovations: problems and prospects. *Methodological Innovations Online*, 1(1), 29-33. DOI: 10.4256/mio.2006.0005
- Mischel, W. (1969). Continuity and change in personality. *American Psychologist*, 24(11), 1012-1018. DOI: 10.1037/h0028886
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80(4), 252-283
- Mischel, W. (1977). On the future of personality measurement. *American Psychologist*, 32(4), 246-254. DOI: 10.1037/0003-066X.32.4.246
- Mischel, W. (1979). On the interface of cognition and personality: Beyond the person-situation debate. *American Psychologist*, 34(9), 740-754. DOI: 10.1037/0003-066X.34.9.740
- Mischel, W. & Ayduk, O. (2002). Self-regulation in a cognitive-affective personality system: attentional control in the service of the self. *Self and Identity*, 1(2), 113-120. DOI: 10.1080/152988602317319285
- Mischel, W. & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246-268
- Mischel, W. & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, 49, 229-258. DOI: 10.1146/annurev.psych.49.1.229
- Mischel, W. & Shoda, Y. (1999). Integrating dispositions and processing dynamics within a unified theory of personality: the cognitive-affective personality system. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 197-218). New York (NY), USA: Guildford
- Mollaret, P. (2009). Using common psychological terms to describe other people: from lexical hypothesis to polysemous conception. *Theory & Psychology*, 19(3), 315-334. DOI: 10.1177/0959354309104157
- Monrouxe, L. V. (2010). Identity, identification and medical education: why should we care? *Medical Education*, 44, 40-49. DOI: 10.1111/j.1365-2923.2009.03440.x
- Monrouxe, L. V. & Rees, C. R. (2012). "It's just a clash of cultures": emotional talk within medical students' narratives of professionalism dilemmas. *Advances*

in Health Sciences Education, 17, 671-701. DOI: 10.1007/s10459-011-9342-z

- Monrouxe, L. V., Rees, C. & Hu, W. (2011). Differences in medical students' explicit discourses of professionalism: acting, representing, becoming. *Medical Education, 45*, 585-602. DOI: 10.1111/j.1365-2923.2010.03878.x
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks (CA), USA: SAGE
- Morgan, D. L. (2007). Paradigms lost and paradigms regained: methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*(1), 48-76. DOI: 10.1177/2345678906292462
- Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement, 7*(2), 191-205. DOI: 10.1177/001316444700700201
- Naidu, T. & Prose, N. (2018). Re-envisioning member checking and communicating results as accountability practice in qualitative research: a South African community-based organization example. *Forum: Qualitative Social Research, 19*(3), art. 26
- Neimeyer, G. J. & Neimeyer, R. A. (1981). Personal construct perspectives on cognitive assessment. In T. V. Merluzzi, C. R. Glass & M. Genest (Eds.), *Cognitive assessment* (pp. 18-232). New York (NY), USA: The Guildford Press
- Neimeyer, G. J. & Neimeyer, R. A. (1985). Relational trajectories: a personal construct contribution. *Journal of Social and Personal Relationships, 2*, 325-349
- Neuman, W. L. (2000). *Social research methods: qualitative and quantitative approaches* (4th ed.). Needham Heights (MA), USA: Allyn & Bacon
- Newman, I. & Ramlo, S. (2010). Using Q methodology and Q factor analysis in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioural research* (pp. 505-530). Thousand Oaks (CA), USA: SAGE
- Newton, P. E. & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London, UK: SAGE

- Niglas, K. (2010). The multidimensional model of research methodology: an integrated set of continue. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (2nd ed., pp. 215-236). Thousand Oaks (CA), USA: SAGE
- Norcini, J. (2006). Faculty observations of student professional behavior. In D. T. Stern (Ed.), *Measuring medical professionalism* (147-157). New York (NY), USA: Oxford University Press
- O'Flynn, S., Power, S., Horgan, M. & O'Tuathaigh, C. M. P. (2014). Attitudes towards professionalism in graduate and non-graduate entrants to medical school. *Education for Health*, 27(2), 200-204. DOI: 10.4103/1357-6283.143770
- O'Hare, M. M. (1987). Career decision-making models: espoused theory versus theory-in-use. *Journal of Counseling & Development*, 65(6), 301-303. DOI: 10.1002/j.1556-6676.1987.tb01290.x
- O'Sullivan, A. J. & Toohey, S. M. (2008). Assessment of professionalism in undergraduate medical students. *Medical Teacher*, 30(3), 280-286. DOI: 10.1080/01421590701758640
- O'Sullivan, H., van Mook, W., Fewtrell, R. & Wass. V. (2012). Integrating professionalism into the curriculum: AMEE guide no. 61. *Medical Teacher*, 34(2), e64-e77. DOI: 10.3109/0142159X.2012.655610
- Office for National Statistics (2011). *2011 Census: key statistics for England and Wales, March 2011*. London: Office for National Statistics
- Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776-783. DOI: 10.1037/h0043424
- Papadakis, M. A., Hodgson, C. S., Teherani, A. & Kohatsu, N. D. (2004). Unprofessional behaviour in medical school is associated with subsequent disciplinary action by a State Medical Board. *Academic Medicine*, 79, 244-249. DOI: 10.1097/00001888-200403000-0001
- Papadakis, M. & Loeser, H. (2006). Using critical incident reports and longitudinal observations to assess professionalism. In D. T. Stern (Ed.), *Measuring medical professionalism* (159-173). New York (NY), USA: Oxford University Press

- Papadakis, M. A., Teherani, A., Banach, M. A., Knettler, T. R., Rattner, S. L., Stern, D. T., ... Hodgson, C. S. (2005). Disciplinary action by medical boards and prior behaviour in medical school. *The New England Journal of Medicine*, 353(25), 2673-2682. DOI: 10.1056/NEJMsa052596
- Pareek, B., Batra, K. & Kalia, R. (2016). Development of professionalism assessment scale for nurse educator. *Asian Journal of Nursing Education & Research*, 6(4), 506-510. DOI: 10.5958/2349-2996.2016.00095.1
- Park, J., Woodrow, S. I., Reznick, R. K., Beales, J. & MacRae, H. M. (2010). Observation, reflection, and reinforcement: surgery faculty members' and residents' perceptions of how they learned professionalism. *Academic Medicine*, 85, 134-139
- Paulhus, D. L., Mysy, D. C. & Yik, M. S. M. (1998). Self-report measures of intelligence: are they useful as proxy IQ tests? *Journal of Personality*, 66(4), 525-554. DOI: 10.1111/1467-6494.00023
- Peeters, M. J. & Stone, G. E. (2009). An instrument to objectively measure pharmacist professionalism as an outcome: a pilot study. *The Canadian Journal of Hospital Pharmacy*, 62(3), 209-216
- Petticrew, M. & Roberts, H. (2006). *Systematic reviews in the social sciences: a practical guide*. Malden (MA), USA: Blackwell
- Pfadenhauer, M. (2006). Crisis or decline? Problems of legitimation and loss of trust in modern professionalism. *Current Sociology*, 54(4), 565-578. DOI: 10.1177/0011392106065088
- Pipere, A. (2007). Mapping the researcher's identity in the university context: dimensions of personal constructs. *Baltic Journal of Psychology*, 8(1/2), 21-36
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. DOI: 10.1037/0021-9010.88.5.879
- Poirier, T. I. & Gupchup, G. V. (2010). Assessment of pharmacy student professionalism across a curriculum. *American Journal of Pharmaceutical Education*, 74(4), 1-5
- Popper, K. R. (1972). *The logic of scientific discovery*. London, UK: Hutchinson

- Poropat, A. E. & Corr, P. J. (2015). Thinking bigger: the Cronbachian paradigm & personality theory integration. *Journal of Research in Personality, 56*, 59-69. DOI: 10.1016/j.jrp.2014.10.006
- Procter, H. (2009). The construct. In R. Butler (Ed.), *Reflections in personal construct theory* (pp. 21-40). West Sussex, UK: Wiley
- Qu, B., Zhao, Y. & Sun, B. (2010). Evaluation of residents in professionalism and communication skills in south China. *Saudi Medical Journal, 31*(11), 1260-1265
- Qu, B., Zhao, Y. & Sun, B. (2012). Assessment of resident physicians in professionalism, interpersonal and communication skills: a multisource feedback. *International journal of Medical Sciences, 9*(3), 228-236. DOI: 10.7150/ijms.3353
- Rankin, N. M., McGregor, D., Butow, P., White, K., Phillips, J. L., Young, J. M. ... Shaw, T. (2016). Adapting the nominal group technique for priority setting of evidence-practice gaps in implementation science. *BMC Medical Research Methodology, 16*, 110. DOI: 10.1186/s12874-016-0210-7
- Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329-353. DOI: 10.1207/s15327906mbr3204_2
- Reber, B. H., Kaufman, S. E. & Cropp, F. (2000). Assessing Q-assessor: a validation study of computer-based Q sorts versus paper sorts. *Operant Subjectivity, 23*(4), 192-209. DOI: dx.doi.org/10.15133/j.os.2000.009
- Reed, D. A., West, C. P., Mueller, P. S., Ficalora, R. D., Engstler, G. J. & Beckman, T. J. (2008). Behaviors of highly professional resident physicians. *Journal of the American Medical Association, 300*(11), 1326-1333
- Rees, C. R. & Knight, L. V. (2007). The trouble with assessing students' professionalism: theoretical insights from sociocognitive psychology. *Academic Medicine, 82*, 46-50. DOI: 10.1097/01.ACM.0000249931.85609.05
- Reichertz, J. (2004). Abduction, deduction and induction in qualitative research. In U. Flick, E. von Kardoff & I. Steinke (Eds.), *A companion to qualitative research* (pp. 159-164). London, UK: SAGE

- Reimer, D., Russell, R., Khallouq, B. B., Kauffman, C., Hernandez, C., Cendán, J. & Castiglioni, A. (2019). Pre-clerkship medical students' perceptions of medical professionalism. *BMC Medical Education*, 19, 239. DOI: 10.1186/s12909-019-1629-4
- Reitzschel, E. F., Nijstad, B. A. & Stroebe, W. (2006). Productivity is not enough: a comparison of interactive and nominal brainstorming groups on idea generation and selection. *Journal of Experimental Social Psychology*, 42(2), 244-251. DOI: 10.1016/j.jesp.2005.04.005
- Research Excellence Framework (2011). *Assessment framework and guidance on submissions (updated to include addendum published in January 2012)*. Retrieved from <file:///C:/Users/user/Zotero/storage/4ETE8U85/GOS%20including%20addendum.pdf>
- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis (MN), USA: University of Minnesota Press
- Rocco, T. S., Bliss, L. A., Gallagher, S., Perez-Prado, A., Alacaci, C., Dwyer, S. ... Pappamihiel, N. E. (2003). The pragmatic and dialectical lenses: two views of mixed methods use in education. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 3-50). Thousand Oaks (CA), USA: SAGE
- Rogers, B. & Ryals, L. (2007). Using repertory grid to access the underlying realities in key account relationships. *International Journal of Market Research*, 49(5), 595-612. DOI: 10.1177/147078530704900506
- Roos, M., Pfisterer, D., Krug, D., Ledig, T., Steinhauser, J., Sxecsényi, J. & Goetz, K. (2016). Adaptation, psychometric properties and feasibility of the professionalism scale Germany. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 113, 66-75. DOI: 10.1016/j.zefq.2016.04.002
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 28-49). Hillsdale (NJ), USA: Erlbaum
- Rosenthal, S., Howard, B., Schlüssel, Y. R., Herrigel, D., Smolarz, G., Gable, B. ... Kaufman, M. (2011). Humanism at heart: preserving empathy in third-year medical students. *Academic Medicine*, 86(3), 350-358. DOI: 10.1097/ACM.0b013e318209897f

- Rosete, D. & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership & Organization Development Journal*, 26(5), 388-399. DOI: 10.1108/01437730510607871
- Rosnow, R. L. (2002). The nature and role of demand characteristics in scientific inquiry. *Prevention & Treatment*, 5, 1-7. DOI: 10.1037/1522-3736.5.1.537c
- Roux, M. (2018). A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, 35, 345-366. DOI: 10.1007/s00357-018-9259-9
- RStudio Team (2016). RStudio: Integrated Development for R [Computer software]. Boston (MA), USA: RStudio, Inc
- Rubin, M. B. (2018). Professionalism and social media usage. *Annual Institute on Mineral Law*, 61, 63-116
- Rudner, L. M. (1983). A closer look at latent trait parameter invariance. *Educational and Psychological Measurement*, 43, 951-955
- Salovey, P. & Grewal, D. (2005). The science of emotional intelligence. *Current Directions in Psychological Science*, 14(6), 281-285
- Saltmarsh, J. (2005). The civic promise of service learning. *Liberal Education*, 50-55
- Sandberg, M. T. (2018). Nursing faculty perceptions of service-learning: an integrative review. *Journal of Nursing Education*, 57(10), 584-589. DOI: 10.3928/01484834-20180921-03
- Saville, P. & MacIver, R (2017). It is the way your write them. In B. Cripps (Ed.), *Psychometric testing: critical perspectives* (pp. 29-42). West Sussex, UK: Wiley
- Sawdon, M. A., Whitehouse, K., Finn, G. M., McLachlan, J. C. & Murray, D. (2017). Relating professionalism and conscientiousness to develop an objective, scalar, proxy measure of professionalism in anaesthetic trainees. *BMC Medical Education*, 17(49), 1-8. DOI: 10.1186/s12909-017-0891-6
- Scheer, J. (2013). Finished an unfinished business: the legacy of Fay Fransella. *Journal of Constructivist Psychology*, 26(1), 30-36. DOI: 10.1080/10720537.2013.732532
- Schmolck, P. (2014). PQMethod (version 2.35). Available at <http://schmolck.org/qmethod/downpqwin.htm>

- Scholte, R. H. J., van Lieshout, C. F. M. & de Wit, C. A. M. (2005). Adolescent personality types and subtypes and their psychosocial adjustment. *Merrill-Palmer Quarterly*, 51(3), 258-286. DOI: 10.1353/mpq.2005.0019
- Secolsky, C. (1987). On the direct measurement of face validity: a comment on Nevo. *Journal of Educational Measurement*, 24(1), 82-83
- Shapiro, J., Rucker, L. & Robitshek, D. (2006). Teaching the art of doctoring: an innovative medical student elective. *Medical Teacher*, 28(1), 30-35. DOI: 10.1080/01421590600568348
- Sharpe, D. & Whelton, W. (2016). Frightened by an old scarecrow: the remarkable resilience of demand characteristics. *Review of General Psychology*, 20(4), 349-368. DOI: 10.1037/gpr0000087
- Shaughnessy, J. J. & Zechmeister, E. B. (2015). *Research methods in psychology* (10th ed.). New York (NY), USA: McGraw-Hill
- Shirley, J. L. & Padgett, S. M. (2006). An analysis of the discourse of professionalism. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 25-41). New York (NY), USA: Springer
- Shoda, Y., Mischel, W. & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behaviour: incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, 67(4), 674-687. DOI: 10.1037/0022-3514.67.4.674
- Sijtsma, K. (2012a). Future of psychometrics: ask what psychometrics can do for psychology. *Psychometrika*, 77(1), 4-20. DOI: 10.1007/S11336-011-9242-4
- Sijtsma, K. (2012b). Psychological measurement between physics and statistics. *Theory & Psychology*, 22(6), 786-809. DOI: 10.1177/0959354312454353
- Sijtsma, K. & Emons, W. H. M. (2013). Separating models, ideas, and data to avoid a paradox: rejoinder to Humphry. *Theory & Psychology*, 23(6), 786-796. DOI: 10.1177/0959354313503724
- Simms, L. J. & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 240-258). New York (NY), USA: Guildford
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117

- Smith, J. A. (1995). Repertory grids: An interactive case-study perspective. In J. A. Smith, R. Harré & L. van Langenhove (Eds.), *Rethinking methods in psychology* (pp. 162-177). London, UK: SAGE
- Snider, K. T. & Johnson, J. C. (2014). Professionalism score and academic performance in osteopathic medical students. *Journal of the American Osteopathic Association*, 114(11), 850-589. DOI: 10.7556/jaoa.2014.171
- Sockett, H. T. (1996). Teachers for the 21st century: redefining professionalism. *NASSP Bulletin*, 80(580), 22-29. DOI: 10.1177/019263659608058004
- Solicitor's Regulation Authority (2019). *The SRA Handbook, version 21*. Retrieved from <https://www.sra.org.uk/solicitors/handbook/welcome/>
- Søndergaard, E., Ertmann, R. K., Reventlow, S. & Lykke, S. (2018). Using a modified nominal group technique to develop general practice. *BMC Family Practice*, 19, 117. DOI: 10.1186/s12875-018-0811-9
- Sonnentag, S. (1998). Identifying high performers: Do peer nominations suffer from a likeability bias? *European Journal of Work and Organizational Psychology*, 7(4), 501-515. DOI: 10.1080/135943298398547
- Souza, E. S., de Lorena, S. B., Ferreira, C. C. G., Amorim, A. F. C. & Peter, J. V. S. (2017). Ethics and professionalism in social media: the online behaviour of medical students. *Revista Brasileira de Educação Médica*, 41(4), 564-575. DOI: 10.1590/1981-52712015v41n3RB20160096
- Squier, R. W. (1990). A model of empathic understanding and adherence to treatment regimens in practitioner-patient relationships. *Social Science & Medicine*, 30(3), 325-339
- Stainton Rogers, R. (1995). Q methodology. In J. A. Smith, R. Harré & L. Langenhove (Eds.), *Rethinking methods in psychology* (pp. 178-192). London, UK: SAGE
- Stainton Rogers, W. (1991). *Explaining health and illness: an exploration of diversity*. Hemel Hempstead, UK: Harvester Wheatsheaf
- Starkweather, J. (2012). Step out of the past: stop using coefficient alpha; there are better ways to calculate reliability. *Research and Statistical Support*, 1-6
- Stenner, P. (2009). Q methodology as a constructivist method. *Operant Subjectivity*, 32(1-3), 46-69

- Stenner, P. H. D., Bianchi, G., Popper, M., Supeková, M., Lukšík, I. & Pujol, J. (2006). Constructions of sexual relationships. *Journal of Health Psychology, 11*(5), 669-684. DOI: 10.1177/1359105306066617
- Stenner, P. H. D., Cooper, D. & Skevington, S. M. (2003). Putting the Q into quality of life; the identification of subjective constructions of health-related quality of life using Q methodology. *Social Science & Medicine, 57*, 2161-2172. DOI: 10.1016/S0277-9536(03)00070-4
- Stenner, P. H. D., Dancey, C. P. & Watts, S. (2000). The understanding of their illness amongst people with irritable bowel syndrome: a Q methodological study. *Social Science & Medicine, 51*, 439-452. DOI: 10.1016/S0277-9536(99)00475-X
- Stenner, P. & Marshall, H. (1999). On developmentality: researching the varied meanings of 'independence' and 'maturity' extant amongst a sample of young people in East London. *Journal of Youth Studies, 2*(3), 297-316
- Stenner, P. & Stainton Rogers, R. (1998). Jealousy as a manifold of divergent understandings: a Q methodological investigation. *European Journal of Social Psychology, 28*, 71-94
- Stenner, P. & Stainton Rogers, R. (2004). Q methodology and qualiquantology: the example of discriminating between emotions. In Z. Todd, B. Nerlich, S. McKeown & D. D. Clarke (Eds.), *Mixing methods in psychology: the integration of qualitative and quantitative methods in theory and practice* (pp. 166-201). New York (NY), USA: Psychology Press
- Stephens, R. A. & Gammack, J. G. (1994). Knowledge elicitation for systems practitioners: a constructivist application of the repertory grid technique. *Systems Practice, 7*(2), 161-182. DOI: 10.1007/BF02169662
- Stephenson, W. (1935a). Correlating persons instead of tests. *Journal of Personality, 4*(1), 17-24. DOI: 10.1111/j.1467-6494.1935.tb02022.x
- Stephenson, W. (1935b). Technique of factor analysis. *Nature, 136*, 297
- Stephenson, W. (1952). Q-methodology and the projective techniques. *Journal of Clinical Psychology, 8*(3), 219-229. DOI: 10.1002/1097-4679(195207)8:3<219::aid-jclp2270080302>3.0.co;2-j
- Stephenson, W. (1953a). Postulates of behaviorism. *Philosophy of Science, 20*(2), 110-120

- Stephenson, W. (1953b). *The study of behaviour: Q-technique and its methodology*. Chicago (IL), USA: The University of Chicago Press
- Stephenson, W. (1977). Factors as operant subjectivity. *Operant Subjectivity*, 1(1), 3-16. DOI: 10.15133/j.os.1977.001
- Stephenson, W. (1986a). Protoconcurus: the concourse theory of communication, part 1. *Operant Subjectivity*, 9(2), 37-58
- Stephenson, W. (1986b). Protoconcurus: The concourse theory of communication, conclusion. *Operant Subjectivity*, 9(3), 73-96
- Stern, D. T. (2000). The development of professional character in medical students. *Hastings Center Report*, 30(4), S26-S29. DOI: 10.2307/3527661
- Stern, D. T. & Papadakis, M. (2006). The developing physician – becoming a professional. *The New England Journal of Medicine*, 355, 1794-1799
- Stewart, D. W., Shamdasani, P. N. & Rook, D. W. (2007). *Focus groups theory and practice* (2nd ed.). Thousand Oaks (CA), USA: SAGE
- Stewart, K., Wyatt, R. & Conway, J. (2011). Unprofessional behaviour and patient safety. *The International Journal of Clinical Leadership*, 17, 93-101
- Sudrajat (2017). Teachers professionalism and the challenge of education in a global era. *Al Murabbi*, 3(2), 211-227
- Suen, H. K. & Lei, P. (2007). Classical versus generalizability theory of measurement. *Educational Measurement*, 4
- Svensson, L. G. (2006). New professionalism, trust and competence. *Current Sociology*, 54(4), 579-593. DOI: 10.1177/0011392106065089
- Swaminathan, R. (2007). Educating for the “real world”: the hidden curriculum of community service-learning. *Equity & Excellence in Education*, 40, 134-143. DOI: 10.1080/10665680701246450
- Swick, H. M. (2000). Toward a normative definition of medical professionalism. *Academic Medicine*, 75, 612-616
- Symons, A. B., Swanson, A., McGuigan, D., Orrange, S. & Akl, E. A. (2009). A tool for self-assessment of communication skills and professionalism in residents. *BMC Medical Education*, 9(1). DOI: 10.1186/1472-6920-9-1
- Tan, F. B. & Hunter, M. G. (2002). The repertory grid technique: a method for the study of cognition in information systems. *MIS Quarterly*, 26(1), 39-57. DOI: 10.2307/4132340

- Tashakkori, A. & Teddlie, C. (2003). Issues and dilemmas in teaching and research methods courses in social and behavioural sciences: US perspective. *International Journal of Social Research Methodology*, 6(1), 61-77. DOI: 10.1080/13645570305055
- Tashakkori, A. & Teddlie, C. (2010a). Epilogue: current developments and emerging trends in integrated research methodology. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (2nd ed., pp. 803-826). Thousand Oaks (CA), USA: SAGE
- Tashakkori, A. & Teddlie, C. (2010b). Putting the human back in “human research methodology”: the researcher in mixed methods research. *Journal of Mixed Methods Research*, 4(4), 271-277. DOI: 10.1177/1558689810382532
- Taylor, C. L. & Grey, N. J. A. (2015). Professional behaviours demonstrated by undergraduate dental students using an incident reporting system. *British Dental Journal*, 10, 591-596. DOI: 10.1038/sj.bdj.2015.386
- Teddlie, C. & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 3-50). Thousand Oaks (CA), USA: SAGE
- Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oaks (CA), USA: SAGE
- Teddlie, C. & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioural research* (2nd ed., pp. 1-42). Thousand Oaks (CA), USA: SAGE
- Teddlie, C. & Tashakkori, A. (2012). Common “core” characteristics of mixed methods research: a review of critical issues and call for greater convergence. *American Behavioral Scientist*, 56(6), 774-788. DOI: 10.1177/0002764211433795
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J. ... de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42. DOI: 10.1016/j.jclinepi.2006.03.012

- The Institute of Engineering and Technology (2017). *Skills & demand in industry: 2017 survey*. Retrieved from <http://www.sciencecampaign.org.uk/resource/iet-skills-and-demands-in-industry-2017.html>
- Thomas, D. B. & Baas, L. R. (1992). The issue of generalization in Q methodology: “reliable schematics” revisited. *Operant Subjectivity*, 16(1/2), 18-36
- Thomas, M. R., Dyrbye, L. N., Huntington, M. P. H., Lawson, K. L., Novotny, M. S., Sloan, J. A. & Shanafelt, T. D. (2007). How do distress and well-being relate to medical student empathy? A multicentre study. *Journal of General Internal Medicine*, 22(2), 177-183. DOI: 10.1007/s11606-006-0039-6
- Tinsley, H. E. A. & Dawis, R. V. (1977). Test-free person measurement with the rasch simple logistic model. *Applied Psychological Measurement*, 1(4), 483-487. DOI: 10.1177/014662167700100404
- Tokuda, Y., Barnett, P. B., Norisue, Y., Konishi, R., Kudo, H. & Miragi, S. (2009). Questionnaire survey for challenging cases of medical professionalism in Japan. *Medical Teacher*, 31, 502-507. DOI: 10.1080/01421590802203488
- Torgerson, C. (2003). *Systematic reviews*. London, UK: Continuum International
- Trathen, A. & Gallagher, J. E. (2009). Dental professionalism: definitions and debate. *British Dental Journal*, 206(5), 249-253. DOI: 10.1038/sj.bdj.2009.164
- Travers, R. M. W. (1951). Rational hypotheses in the construction of tests. *Educational and Psychological Measurement*, 11(1), 128-137. DOI: 10.1177/001316445101100113
- Treviño, L. K., Weaver, G. R. & Reynolds, S. J. (2006). Behavioral ethics in organizations: a review. *Journal of Management*, 32(6), 951-990. DOI: 10.1177/0149206306294258
- Troman, G. (1996). The rise of the new professionals? The restricting of primary teachers’ work and professionalism. *British Journal of Sociology of Education*, 17(4), 473-487
- Tromp, F., Rademakers, J. J. D. J. M. & Ten Cate, T. J. (2007). Development of an instrument to assess professional behaviour of foreign medical graduates. *Medical teacher*, 29(2-3), 150-155. DOI: 10.1080/01421590601178014

- Tromp, F., Vernooij-Dassen, M., Kramer, A., Grol, R. & Bottema, B. (2010). Behavioural elements of professionalism: assessment of a fundamental concept in medical care. *Medical Teacher*, 32(4), e161-e169. DOI: 10.3109/01421590903544728
- Tsugawa, Y., Ohbu, S., Cruess, R., Cruess, S., Okubo, T., Takahashi, O. ... Fukui, T. (2011). Introducing the professionalism mini-evaluation exercise (P-MEX) in Japan: results from a multicentre, cross-sectional study. *Academic Medicine*, 86, 1026-1031. DOI: 10.1097/ACM.0b013e3182222ba0
- Tsugawa, Y., Tokuda, Y., Ohbu, S., Okubo, T., Cruess, T., Cruess, S. ... Fukui, T. (2009). Professionalism mini-evaluation exercise for medical residents in Japan: a pilot study. *Medical Education*, 43, 968-978. DOI: 10.1111/j.1365-2923.2009.03437.x
- Usman, E. A. (2016). Nurturing the law student's soul: why law schools are still struggling to teach professionalism and how to do better in an age of consumerism. *Marquette Law Review*, 99(4), 1021-1072
- Van de Ven, A. & Delbecq, A. L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. *Academy of Management Journal*, 14(2), 203-212. DOI: 10.2307/255307
- van Exel, J. & de Graaf, G. (2005). *Q methodology: a sneak preview*. Available from www.jobvanexel.nl
- van Mook, W. N. K. A., van Luijk, S. J., O'Sullivan, H., Wass, V., Schuwirth, L. W. & van der Vleuten, C. P. M. (2009). General considerations regarding assessment of professional behaviour. *European Journal of Internal Medicine*, 20(4), e90-e95. DOI: 10.1016/j.ejim.2008.11.011
- Veloski, J., Fields, S. K., Boes, J. R. & Blank, L. L. (2005). Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Academic Medicine*, 80(4), 366-370
- Verkerk, M. A., de Bree, M. J. & Mourits, M. J. E. (2007). Reflective professionalism: interpreting CanMEDS' "professionalism". *Journal of Medical Ethics*, 33, 663-666. DOI: 10.1136/jme.2006.017954
- Violato, C., Lockyer, J. M. & Fidler, H. (2008). Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme.

Medical Education, 42, 1007-1013. DOI: doi:10.1111/j.1365-2923.2008.03127.x

- Virtue, S. M., Pendergast, L., Tellez, M., Waldron, E. & Ismail, A. (2017). Identifying noncognitive skills that contribute to dental students' success: dental faculty perspectives. *Journal of Dental Education*, 81(3), 300-309
- Walker, B. M. & Winter, D. A. (2007). The elaboration of personal construct psychology. *Annual Review of Psychology*, 58, 453-477. DOI: 10.1146/annurev.psych.58.110405.085535
- Wang, J., He, B., Miao, X., Huang, X., Lu, Y. & Chen, J. (2017). The reliability and validity of a new professionalism assessment scale for young health care workers. *Medicine*, 96(25), e7058. DOI: 10.1097/MD.00000000000007058
- Wang, Z., Wang, S., Zhang, Y. & Jiang, X. (2019). Social media usage and online professionalism among registered nurses: a cross-sectional survey. *International Journal of Nursing Studies*, 98, 19-26. DOI: 10.1016/j.ijnurstu.2019.06.001
- Warren, B (2010). Kelly's personal construct psychology and Dewey's pragmatism: some direct and some 'intellectual context' aspects. *Personal Construct Theory & Practice*, 7, 32-40
- Watson, S. (2013). Understanding mathematics teachers' professional development from the perspective of social learning theory. In B. Ubuz, C. Haser & M. A. Mariotti (Eds.), *Proceedings of the 8th Congress of European Research in Mathematics Education (CERME-8)*, pp. 3287-3295). Ankara, Turkey: Middle East Technical University and ERME
- Watts, S. & Stenner, P. (2005). The subjective experience of partnership love: a Q methodological study. *British Journal of Social Psychology*, 44(1), 85-107. DOI: 10.1348/014466604X23473
- Watts, S. & Stenner, P. (2007). Q methodology: the inverted factor technique. *The Irish Journal of Psychology*, 28(1-2), 63-76
- Watts, S. & Stenner, P. (2012). *Doing Q methodological research: theory, method and interpretation*. London, UK: SAGE
- Wear, D. (2006). Respect for patients: a case study of the formal and hidden curriculum. In D. Wear & J. M. Aultman (Eds.), *Professionalism in medicine: critical perspectives* (pp. 87-102). Boston (MA), USA: Springer

- Weaver, M. J., Ow, C. L., Walker, D. J. & Degenhardt, E. F. (1993). A questionnaire for patients' evaluations of their physicians' humanistic behaviors. *Journal of General Internal Medicine*, 8(3), 135-139. DOI: 10.1007/bf02599758
- Webb, N. M., Shavelson, R. J. & Haertel, E. H. (2007). Reliability coefficients and generalizability theory. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26: psychometrics* (pp. 81-124). Amsterdam, The Netherlands: Elsevier Science B. V.
- Wilkinson, T. J., Wade, W. B. & Knock, L. D. (2009). A blueprint to assess professionalism: results of a systematic review. *Academic Medicine*, 84, 551-558
- Williams, J., Field, C. & James, K. (2011). The effects of a social media policy on pharmacy students' Facebook security settings. *American Journal of Pharmaceutical Education*, 75(9), 177. DOI: 10.5688/ajpe759177
- Windle, G., Bennett, K. M. & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes*, 9(8). DOI: 10.1186/1477-7525-9-8
- Wise, H. H. & Yuen, H. Y. (2013). Effect of community-based service learning on professionalism in student physical therapists. *Journal of Physical Therapy Education*, 27(2), 58-64. DOI: 10.1097/00001416-201301000-00013
- Wolming, S. & Wikström, C. (2010). The concept of validity in theory and practice. *Assessment in Education: Principles, Policy & Practice*, 17(2), 117-132. DOI: 10.1080/09695941003693856
- Yanai, H. & Ichikawa, M. (2007). Factor analysis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26: psychometrics* (pp. 257-296). Amsterdam, The Netherlands: Elsevier Science B. V.
- Yang, Y. & Green, S. B. (2011). Coefficient alpha: a reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4), 377-397. DOI: 10.1177/0734282911406668
- Yang, A. & Taylor, M. (2014). A global perspective on public relations professionalism: mapping the structure of public relations associations' international networks. *Journalism & Mass Communication Quarterly*, 91(3), 508-529. DOI: 10.1177/1077699014538831

- Yim, O. & Ramdeen, K. T. (2015). Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *The Quantitative Methods for Psychology, 11*(1), 8-21. DOI: 10.20982/tqmp.11.1.p008
- Young, M. E., Cruess, S. R., Cruess, R. L. & Steinert, Y. (2014). The professionalism assessment of clinical teachers (PACT): the reliability and validity of a novel tool to evaluate professional and clinical teaching behaviors. *Advances in Health Sciences Education, 19*, 99-113. DOI: 10.1007/s10459-013-9466-4
- Zanetti, M., Keller, L., Mazor, K., Carlin, M., Alper, E., Hatem, D. ... Pugnaire, M. (2010). Using standardized patients to assess professionalism: a generalizability study. *Teaching and Learning in Medicine, 22*(4), 274-279. DOI: 10.1080/10401334.2010.512542
- Zhao, Y., Zhang, X., Chang, Q. & Sun, B. (2013). Psychometric characteristics of the 360° feedback scales in professionalism and interpersonal and communication skills assessment of surgery residents in China. *Journal of Surgical Education, 70*(5), 628-635. DOI: 10.1016/j.jsurg.2013.04.004
- Zijlstra-Shaw, S., Robinson, P. G. & Roberts, T. (2012). Assessing professionalism within dental education; the need for a definition. *European Journal of Dental Education, 16*, e128-e136. DOI: 10.1111/j.1600-0579.2011.00687.x
- Zumbo, B. D. (2007). Validity: foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26: psychometrics* (pp. 45-80). Amsterdam, The Netherlands: Elsevier Science B. V.

Appendices

Appendix A: Ethical Approvals



Health and Life Sciences Committee on Research Ethics (Psychology, Health and Society)

26 May 2017

Dear Dr Forsythe,

I am pleased to inform you that your application for research ethics approval has been approved. Details and conditions of the approval can be found below:

Reference:	1922
Project Title:	Defining professionalism: An exploration of subjectivity
Principal Investigator/Supervisor:	Dr Alex Forsythe
Co-Investigator(s):	Ms Katie Cunliffe, Prof Luke Dawson
Lead Student Investigator:	-
Department:	School of Psychology (including DClinPOsych)
Reviewers:	Prof Ben Ambridge, Dr Georg Meyer
Approval Date:	26/05/2017
Approval Expiry Date:	Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions

- All serious adverse events must be reported via the Research Integrity and Ethics Team (ethics@liverpool.ac.uk) within 24 hours of their occurrence.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the research, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form using the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Health and Life Sciences Committee on Research Ethics (Psychology, Health and Society)

iphsec@liverpool.ac.uk

0151 795 5420



Health and Life Sciences Committee on Research Ethics (Psychology, Health and Society)

24 August 2017

Dear Dr Forsythe,

I am pleased to inform you that the amendment to your study has been approved. Details and conditions of the approval can be found below:

Reference: 1922 (amendment)
Project Title: Defining professionalism: An exploration of subjectivity
Principal Investigator: Dr Alex Forsythe
Co-Investigator(s): Ms Katie Cunliffe, Prof Luke Dawson
Student Investigator(s): -
Department: School of Psychology (including DClinPOsych)
Approval Date: 24/08/2017

The amendment was **APPROVED** subject to the following conditions:

Conditions

- All serious adverse events must be reported to the Committee within 24 hours of their occurrence, via the Research Integrity and Ethics Officer (ethics@liv.ac.uk).
- If it is proposed to extend the duration of the study beyond the expiry date listed above, the Committee should be notified.
- If it is proposed to make an amendment to the research, you should notify the Committee by following the Notice of Amendment procedure.
- If the named Principal Investigator or Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore please contact the Committee (details below) in order to notify them of a change in Principal Investigator or Supervisor.

Kind regards,

Health and Life Sciences Committee on Research Ethics (Psychology, Health and Society)

iphsrec@liverpool.ac.uk

0151 795 5420



Health and Life Sciences Research Ethics Committee (Psychology, Health and Society)

2 August 2019

Dear Dr Forsythe

I am pleased to inform you that your application for research ethics approval has been approved. Application details and conditions of approval can be found below. Appendix A contains a list of documents approved by the Committee.

Application Details

Reference: 4987
Project Title: The model of interpersonal professionalism
Principal Investigator/Supervisor: Dr Alex Forsythe
Co-Investigator(s): Ms Katie Cunliffe, Prof Luke Dawson
Lead Student Investigator: -
Department: School of Dentistry
Approval Date: 02/08/2019
Approval Expiry Date: Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions of approval

- All serious adverse events must be reported to the Committee (ethics@liverpool.ac.uk) in accordance with the procedure for reporting adverse events.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the study, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor changes, or leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Health and Life Sciences Research Ethics Committee (Psychology, Health and Society)

iphsrec@liverpool.ac.uk

0151 795 5420

Appendix - Approved Documents

(Relevant only to amendments involving changes to the study documentation)

The final document set reviewed and approved by the committee is listed below:

Document Type	File Name	Date	Version
Interview Schedule	Interview Schedule 26.03.19	26/03/2019	1
Advertisement	Advert 26.03.19	26/03/2019	1
Participant Information Sheet	Participant Information Sheet 22.07.19	22/07/2019	2
Participant Consent Form	Consent Form 22.07.19	22/07/2019	2

Appendix B: Repertory Grid Technique Defining Elements

Participant	element	poleA	xprof	someprof	neutral	someunprof	xunprof	ideal	Self	poleB
1 a	Conscientious						3	5		Driven
1 b	Reliable						4	5		Inconsistent
1 c	Discourteous						1	1		Helpful
1 d	Inconsistent						4	1		Consistent
1 e	Guarded						1	5		More relaxed
1 f	Proactive						5	5		Unhelpful
1 g	Engaged						5	5		Disengaged
1 h	Upbeat						5	5		Dour
1 i	Apathetic						1	1		Eager
1 j	Reactive						4	3		Measured
1 k	Relatable						5	5		Uninspiring
1 l	Inquisitive						5	5		Disinterested
1 m	Thorough						5	5		Idle
1 n	Relaxed						3	2		Uptight
2 a	Smartly dressed			1	2	2		5	1	Untidy appearance
2 b	Boiterous			5				1		Quiet
2 c	Assertive			3				1		Persuadable
2 d	Efficient			1				3		Busy fool
2 e	Supportive			3				3		Unavailable
2 f	Organised			1				5		Late
3 a	Casual			4	2	3		5	4	Strict
3 b	Understanding							5		Hard
3 c	Decent							5		Aggressive
3 d	Friendly							5		Nasty
3 e	Tolerant							5		Impatient
3 f	Social							5		Independent
3 g	Jolly							4		Miserable
3 h	Accepting							5		Finds enemies
3 i	Humour							5		Sour
3 j	Bully							1		Friend
3 k	Authority							1		Easy going
3 l	Unforgiving							1		Chilled
3 m	Reliable							5		Untrustworthy
3 n	Stubborn							1		Tolerant
3 o	Scruffy							3		Presentable
4 a	Professional			1	3	3	4	4	1	Unprofessional
4 b	Objective			5				5		People pleaser
4 c	Sly						1	1		Honest
4 d	Friendly						5	4		False
4 e	Self preserving						1	1		Selfless
4 f	Optimistic						4	5		Pessimistic
5 a	Laid back			2	4	3		3	3	Tightly wound
5 b	Focused							4	1	Distracted
5 c	Confidential							5	2	Non-confidential
5 d	Driven							4	1	Drifting
5 e	Mature							5	1	Immature
5 f	Gossipy							1	4	Appropriate
7 a	More concerned with self			5	3	1	1	1	5	Concerned with others
7 b	Rule application			4				5		Spirit rather than letter
7 c	Listens			1				5		Talks
7 d	Women			1				5		Male
7 e	Family			1				5		Not related
7 f	Doing PhD's			1				5		Not studying
7 g	Understand mental health			1				5		No understanding
7 h	Good looking			1				5		Average and overweight
7 i	Tenacious in the face of difficulty			1				2		Don't know well enough t
7 j	Medicine/biology			1				3		Accounting
7 k	Confidential			1				5		Anxious and insecure in a
7 l	Ignore things not directly interested in			5				2		Wide range of interests
7 m	Second jobs			1				5		One job
7 n	People seek their opinion			1				4		Feel my opinion not alwa
7 o	Only know in a single capacity (e.g. work)			4				1		This is me, so in all spher

8 a	Dry humour					5			Mean humour
8 b	Smartly dressed					1			Casual
8 c	Forward					1			Emotional
8 d	Lazy					1			Efficient
8 e	Fair					5			Ruthless
8 f	Helpful					5			Unhelpful
8 g	Open					3			Confidential
9 a	Wisdom					5			Decadence
9 b	Sense of humour					5			Hypocrisy
9 c	Hard-working					5			Ambivalent
9 d	Stability					5			Moody
9 e	Punctuality					5			Randomness
9 f	to the point					5			Distracted
9 g	German discipline					5			Black sense of humour ov
9 h	curious					5			Plan
9 i	Sense of humour/some clumsiness in every life settings					5			Too coordinated
9 j	Thinking critically					5			Going with the flow
9 k	Self-discipline					5			No self-discipline/stealing
9 l	Some level of laziness					5			Cleverness to the point w
9 m	Vulnerability					5			Indestructability
9 n	Organization of a lot of people					5			Using other people's labo
9 o	Kind					5			Rude
10 a	Kind					5			Rude
10 b	Flexible					4			Resistant
10 c	Helpful					5			Selfish
10 d	Enthusiastic					5			Bored
10 e	Hardworking					5			Lazy
10 f	Intelligent					3			
10 g	Thoughtful					5			
11 a	Dedication					5			Ignorant
11 b	Humour					5		1	Selfish
11 c	Focused					5		1	Bully
11 d	Detail driven					5		5	Creative
11 e	control					5		5	Abstract
11 f	Support and guidance					1		5	Development
11 g	Creative					1		1	Personal gain
11 h	Global thinker					1		1	Detail driven
11 i	Honest					1		1	Narrow thinker
11 j	Transparent					1		1	Dishonest
11 k	Myopic to detail					5		1	Manipulative
11 l	Driven					5		5	Flexible
11 m	Appropriate networks to suppor others					5		5	Relaxed
11 n	Rigid					1		1	Network to support self
11 o	Wide knowledge					5		5	Flexible
12 a	Focused					5		1	Narrow focus
12 b	Competent					5			Liar
12 c	Liars					5			Lack of knowledge
12 d	Focused					1			Serious
12 e	Focused					3			Knowledge
14 a	Caring					5			Liar
14 b	Fair					5			Self-absorbed
14 c	Detail-oriented					4			Lets feelings get in the w
14 d	Hard-working					3			Rushed
14 e	Focussed					3			Does the minimum
14 f	Interested					3			Distractable
14 g	Competent					4			Disinterested
14 h	Intelligent					5			Incompetent
14 i	Competent					4			Unfocused
14 j	Willing to learn					5			Incompetent
14 k	Polite					5			Arrogant
14 l	Loyal					5			Rude
14 m	Gossipy					3			Takes advantage
14 n	Knowledgeable					3			Discrete
14 o	Literate					5			Ignorant
15 a	Friendly					3			Bad at grammar
15 b	Academic					3			Very extrovert
15 c	Warm					5			Relaxd
15 d	Impulsive					5			Too chilled
15 e	Relaxed					5			Rational
15 f	Understanding					5			Irritating
15 g	Lack of problem solving					5			Judgemental
15 h	Friendly					5			Problem solver
15 i	Intellectually dynamic					5			Unfriendly
15 j	Overly friendly					5			Lack of lateral thinking
15 k	Highly intelligent					3			Appropriately reserved
15 l	Young					3			Moderately intelligent
						3			Mature

17 a	Organised	1	2	5	5	1	Lazy
17 b	Gain knowledge			5			Uninterested
17 c	Researcher			5	3		Workman
17 d	Quick, diligent			5	4		Undiligent
17 e	Motivated			5	5		Unmotivated
17 f	Ambitious			5	5		Bored
17 g	Full of ideas			3	4		Without concept
17 h	Intelligent			5	3		Stupid
17 i	Dainty			3	5		Imprecise
17 j	Do their job			4	5		Cheating
17 k	Get things done faster			4	4		Little bit slow
17 l	Think about it			5	4		Just does his job
17 m	Passionate			5	4		Just does the job
17 n	Motivated			5	5		Unmotivated
18 a	Ethical sense	1	3		5		Harassing attitudes
18 b	Team spirit						Selfishness
18 c	Fairness						Double-faced
18 d	Willing to help and easily approachable						Difficult to approach and
18 e	Egocentrism and badly authoritative						Easier to talk to
18 f	Gread leaders						Not a strong leadership p
18 g	High experience						Less experienced
18 h	Easier to talk to, more flexible						Very bossy
18 i	Fairness towards who is "below them", kind and helpful						Demanding, denigrating
18 j	Not highly professional but still nice to relate to						Unpleasant to relate to w
18 k	Understanding towards who is less experienced, willing to be a guide						Irresponsible and not und
18 l	Not totally fair or precise in their work					1	Incredibly fair, experience
18 m	Not completely candid in their work					2	Fair and crystalline with t
18 n	Highly influential					5	Not so influential
18 o	Bossy					1	More on the same level v
19 a	Educated	1	1	3	5	1	Shallow
19 b	Eloquence	4			3		Values
19 c	Extravert	5			1		Quiet
19 d	Serene	1			5		Party lover
19 e	Strong	1			4		Weak
19 f	Ambitious	1			5		Flaky
19 g	Outward oriented	5			1		Deep
21 a	Knowledgeable		2	4		1	Unconfident
21 b	Reserved				2	4	Talkative
21 c	Observant				2	1	Unaware
21 d	Male				3	3	Female
21 e	Warm and friendly				2	5	Professional but not wann
21 f	Careful				4	1	Impulsive
21 g	Judicious with words				4	1	Says what they think
21 h	Goes out of their way to help				5	1	Less eager to jump in
21 i	Confident in their job				4	1	Unsure about some thing
21 j	Work appropriate conversation				5	1	Speaks like a teenager
21 k	Friendly				4	1	Quieter
21 l	Good sense of humour				2	5	Warm but not necessarily
21 m	Cares about what they do				5	1	Bored and dismissive
21 n	Self motivated				4	1	Needs help and support
21 o	Makes people feel welcome				5	1	Clique-ish
22 a	Focused	1	1	2	5		Tangential
22 b	Academic				2		Intuitive
22 c	Unemotional				1		Unpredictable
22 d	Clarity of purpose				5		Unclear on goal
22 e	Angry				5		Easy going
22 f	Makes things happen				5		Tries too hard
22 g	Unassertive				1		Assertive
22 h	Looks for direction				1		Sets directions
22 i	Manipulative				3		Direct
22 j	Can be informal				3		Rigidly formal
22 k	Develops others				3		Controls others
22 l	Lacks self awareness				1		Very self aware
22 m	Impatient				4		Takes it in her stride
22 n	Content				4		Dissatisfied
22 o	Desires approval				1		Desires money

23 a	Short tempered	5	3	3		1		Thoughtful
23 b	focused					1		Organised
23 c	Leader					4		focused
23 d	Organised					5		Narrow minded
23 e	Hard working					5		Ignorant
23 f	Target orientated					3		Lazy
23 g	Democratic					5		Arrogant
23 h	Friendly					5		Loathed
23 i	Helpful					5		Closed
23 j	Listening					5		Ignorant
23 k	Threatening					1		Caring
23 l	Valued					5		Tolerated
23 m	Liked					5		Detested
23 n	Leader					1		Equal
23 o	Role model					5		Tolerated
26 a	Organised	1	2	5	5	5		Disloyal
26 b	Selfish					1		Team player
26 c	Dishonest, selfish					1		Honest, team player
26 d	Respectful					5		Disrespectful
26 e	Hardworking					5		Gossiper
26 f	Hardworking					4		Lazy
26 g	Hardworking					4		Lazy
26 h	Tidy					4		Messy
26 i	Hardworking					4		Disorganised
27 a	Work ethic	1	1	4		5		Selfish
27 b	Sensitive					5		Hard headed
27 c	Caring					5		Uncaring
27 d	Lazy					3		Stressed
27 e	Organised					4		Childish
27 f	Hot headed					1		Approachable
27 g	Work at home					5		Sarcastic
27 h	Good with discussions					5		Two faced
27 i	Work ethic					3		Passes tasks onto others
27 j	Sticks to one system					4		Gives up easily
27 k	Good with paperwork					5		Sly
27 l	Can twist people words					3		Honest
27 m	Gets too involved with work					3		Complains about work
28 a	Highly communicative	1	4	3		5		Guarded/almost secretive
28 b	Full of business ideas					5		Embarassing conduct
28 c	Consistent in approach					5		Random, erratic
28 d	Charismatic					5		Inconsistent
28 e	People centred					5		Selfish-needy
28 f	Well presented					5		Slovenly
28 g	Decisions makers					5		Dithering
28 h	Energetic					5		Lazy
28 i	Intellectually dynamic					5		Unaware
28 j	Serious					5		Joker
28 k	Shoddy					1		Attentio to every detail
28 l	Custoemr focused					3		Technology focused
28 m	Punctuality					5		Random
28 n	Knowledgeable					5		Disinterested
28 o	Distrusted					1		Honourable
29 a	Not proactive	5	5	3		1		Engaged
29 b	Confidence					5		Insecurity
29 c	Focused					4		Unclear
29 d	Experienced in their field					4		Poor leadership skills
29 e	Enthusiasm					3		Formality
29 f	Organised					5		Mood Hoover
29 g	Expresses unhappiness about work					1		Enthusiastic
29 h	Accountable					5		Avoids work
29 i	Respected within organisation					4		Unapproachable

33 a	Skilled personal networker				3	3		
33 b	Trustworthy				3	4		Untrustworthy
33 c	Credible				4	4		Lacks integrity
33 d	Extremely experienced				2	4		An up and coming novice
33 e	Believable				4	4		Untrustworthy
33 f	Gossipy				1	2		Trustworthy
33 g	Dependable				4	4		Scatty
33 h	Trustworthy				4	4		Not completely rustwort
33 i	Passionate				2	3		
33 j	Excellent listener				4	4		Sometmes difficult to talk
33 k	Cautious				4	1		Maverick
33 l	Visionary				3	5		Un-realistic
33 m	Excellent sense of humour				2	5		Quite serious
33 n	Focused and finishers				4	4		Unfocused, scatty, dream
33 o	Good mentor				5	5		Terrible mentor
34 a	Results				2	2		Knowledge transfer
34 b	Presentation skills				2	2		difficulty delegating
34 c	Skip proces steps				1	1		Conscientiously
34 d	Managing expectations				4	4		Uncommunicative
34 e	Difficulty receiving criticism				2	2		Welcome criticism
34 f	Responsible				5	5		Does not see the necessit
36 a	Warm				3	3		Too ambitious
36 b	Not too emotional				4	4		No boundaries
36 c	Friendly but slightly guarded				3	3		Heart on sleeve
36 d	Down to earth				5	5		Volatile
36 e	Defensive				1	1		Confidential
36 f	Secure				5	5		Insecure
36 g	Confident				5	5		Insecure
36 h	Pragmatic				5	5		Chaotic
36 i	Confident				5	5		Insecure
36 j	Reliable				5	5		Unreliable
36 k	Approachable				5	5		Too sensitive
36 l	Sensitive				1	1		Less sensitive
36 m	Reliable				3	3		Too chatty
37 a	Dutiful				3	3		Giving appropriate feedbi
37 b	Responsible				3	3		Expected student behavi
37 c	Accurate feedback to change behaviours				3	3		Avoids conflict and fails to
37 d	Hold values recognised by the medical profession				3	3		Holds values recognised t
37 e	Avoid conflict with students				3	3		Embraces conflict (when
37 f	Try to develop future medical professionals				3	3		Try to develop future scie
37 g	Acts as a good role model				5	5		Acts as a poor role model
37 h	Puts the University first				5	5		Puts the NHS first, thoug
37 i	Driven by ethics appropriate for patient care				5	5		Drive by self interest
37 j	Puts others first				5	5		Puts self first
37 k	Share a love of learning and instill this in students				3	3		Teaches by tradition and
37 l	Uses an evidence base to inform decisions				5	5		Uses populare culture to
37 m	Behaviour always appropraite to the situation				5	5		Behaviour not often appr
37 n	Approachable				5	5		Aloof
39 a	Honest				5	5		Untrustworthy
39 b	Direct				4	4		Obtuse
39 c	Act in others best interests				5	5		Self obsessed
39 d	Reliable				5	5		Unpredictable
39 e	Comply with regulation				5	5		Will choose what to com
39 f	Altruistic				5	5		Motivated by personal ga
39 g	Respected by all peers				5	5		Only respected by some p
39 h	Mentor others selflessly				5	5		Not interested in develop
39 i	Emotionall stable in crises				5	5		Can overreact
39 j	Work towards most benefit for most people				5	5		Works to most benefit fo
39 k	Manage most people well				5	5		Polarises opinion
39 l	Effective leaders				5	5		Ineffective leader
39 m	Receptive to alternate views				5	5		Fixed views
39 n	Will compromise standards under duress				1	1		Maintains standards at al
39 o	Not always emotionally intelligent				1	1		Aware of others' feelings

Appendix C: Repertory Grid Technique Cluster Analysis Results

Cluster 1

Work approach			Relates to others			Personal		
capable	desires approval (unclear)	ambivalent	confidential (6)	boisterous (unclear)	abrupt	capable of letting go	dainty (unclear)	decadent
cautious	intuitive (unclear)	apathetic	dependable		controls others	deep	dissatisfied (unclear)	dithering
committed (2)	known attitude (unclear)	careless	develops others (2)		critical	discrete	extravert (unclear)	dour
conscientious	sticks to one system (unclear)	chaotic	forward		discourteous	down to earth	tightly wound (unclear)	haughty
content		detail-oriented (3)	relateable		disloyal	dutiful	too coordinated (unclear)	hot headed
customer focused		disengaged			distrusted	eloquent		obtuse
dedicated		disorganised (3)			false	genuine		reactive (unclear)
detail-oriented		distracted (2)			finds enemies	has a conscience		scattered
driven (3)		flippant			gossipy (4)	honourable		shallow
exhibits self-discipline		imprecise			gruff	ordinary		slovenly
extremely experienced (2)		impulsive (2; mixed)			guarded (2)	secure		unaware (2)
flexible (3)		inexperienced (2)			hard	serene		unpredictable (2)
focused (10)		lapses in consistency			loathed	steadfast		volatile (2)
inquisitive		late			self-absorbed	strong		vulnerable
literate		limited			sly (2)	thrill seeking (unclear)		workman
measured		maverick			superior			young
novel		messy			tolerated (2)			
obedient		opportunist			unforgiving			
organised (9)		rebellious			uninspiring			
pragmatic		resistant						
successful		rigid (2)						
thorough		rushed						
tidy		shoddy						
transparent		sloppy						
visionary		struggles						
wisdom		tangential						
		traditional						
		unreliable						

Cluster 2

Work approach			Relates to others			Personal		
academic	academic (mixed)	an up and coming novice	accurate feedback to change behaviours	independent	authoritative	altruistic (2)	human (unclear)	arrogant (3)
accountable	planner (unclear)	bored (2)	aware of the feelings of others	social	bored and dismissive	clear sense of right and wrong	takes it in their stride	family
assertive (2)		casual	concerned with others		bossy (2)	ethical sense		ignorant (4)
behaviour always appropriate to the situation		cheats	considers of new suggestions		can twist the words of others	stubborn (3)		joker
can be informal (unclear)		desires money	democratic		controls others	morals		lacks self awareness
cares about what they do		does not consider alternative ways if it creates more work	easy to talk to (2)		difficulty receiving criticism	only know in a single capacity (e.g. work)		party lover
casual		does not see the necessity (of responsibility)	effective leader		dishonest (4)	selfless		third parties
competent (3)		doesn't take things seriously	effective leaders		dismissive of new suggestions	tenacious in the face of difficulty		women
compliant		drifting	equal to others (2)		doesn't value opinions of others	upbeat		
consistent (4)		fixed views	fair		egoism and authoritarian	self aware		
educated (2)		formal	fair and direct		emotionally unintelligent			
focused and finishers		ignore things not directly interested in	fair and transparent		ineffective leader (2)			
gets things done faster		incompetent (2)	friend		liar			
global thinker (unclear)		inconsistent (4)	good mentor		manipulative (3)			
good customer service		lacks integrity	good role model		mood hoover			
good with paperwork		lazy (3)	good with discussions		not interested in developing others			
highly influential		motivated by personal gain	honest		only respected by some peers			
intellectually dynamic (2)		narrow focus/minded/thinker (3, mixed)	honest, team player		passes tasks onto others			
intelligent (6)		needs help and support	incredibly fair, experienced and precise		people pleaser			
interested in gaining knowledge		not always rule following (unclear)	leader (2)		poor mentor			
invested in their work		not completely candid in their work	loyal		poor role model (2)			
knowledgeable (6)		not influential	people centred		says what they think			
maintains standards at all times		not really involved in their work	people seek their opinion		selfish (10)			
makes decisions		outward oriented	positive role model		steals other people's work and takes credit			
never makes mistakes		personal gain	prioritises knowledge transfer i.e. development of others		two-faced			
objective		professional but not warm	provides appropriate feedback		unapproachable			
planner		random, erratic (3)	provides equal treatment to everyone		unapproachable			
presentable (2)		rigidly formal (unclear)	provides support and guidance		unfair and imprecise			
problem solver		rule follower	respected by all		unpleasant to relate to			
professional (3)		scruffy	strict		untrustworthy (3)			
puts the university first		set in their ways (unclear)	team player (2)		uses others to cover their own laziness			
rational (2)		skip proces steps	tolerant (2)		works to benefit self			
reliable (6)		teaches by tradition and does not engage	trustworthy (3)					
responsible (2)		tries too hard	understanding (4)					
rule follower		trouble maker	unprofessional but nice to relate to					
set in their ways (unclear)		unclear on goal	values opinions of others					
sets direction		uneducated	warm					
smartly dressed (2)		unintelligent (2)	welcomes criticism					
spirit rather than letter		untidy appearance (2)	willing to help and approachable					
thinks critically		uses popular culture to inform decisions	works to benefit most others					
uses an evidence base to inform decisions								
willing to learn								

Cluster 3

Work approach			Relates to others			Personal		
charismatic		anxious and insecure in ability	acts in the best interests of others		avoids conflict to detriment of development (2)	abstract		
clarity of purpose		complains about work (2)	embraces conflict (when needed) to support development		behavior towards colleagues varies based on sympathy	black sense of humour over self-damage due to too much discipline	accounting	
confident in their job		difficult delegating	engages in work appropriate conversation		not a strong leadership personality	emotionally stable in crises	sense of humour/some clumsiness in every life settings	average and overweight
does their job (cheats)		does the minimum (2, unclear)	fairness towards who is "below them", kind and helpful		ruke in the workplace	hold values recognised by the medical profession		doing a PhD
driven by ethics appropriate for patient care		emotional in work (2, unclear)	manages people well		uses the labour of others excessively	not studying		wears their heart on their sleeve
experienced		engaged student behaviour standards (unclear)	managing expectations			personality		negative personality
full of business ideas		exploits position (unclear)	mentors others selflessly			researcher		second jobs
german discipline		holds values recognised by the scientific profession (unclear)	networks to support others			understand mental health		self obsessed
mind wanders (unclear)		lack of lateral thinking (unclear)	respected within organization			work at home		this is me, so in all spheres of life
passionate about their work		lack of problem solving skills (unclear)	respectful of colleagues					
presentation skills (unclear)		lacks knowledge	share a love of learning and install this in students					
proactive (2)		let's emotions guide professional decisions (unclear)	short tempered					
problem solver		network to support self (unclear)	try to develop future medical professionals					
punctual (2)		not proactive						
realistic in work related topics		just the the first, though paid by the university						
skilled networker		self preserving						
stable		technology focused						
target orientated		try to develop future scientists in a medical programme						
unemotional at work		unfocused, scatty, dreamer						
willing to work extra for the best possible result		will choose what to comply with						
work ethic (2)		will compromise standards under duress						

Cluster 4

	Work approach			Relates to others			Personal	
ambitious (3)	laid back	too ambitious		accepting	abusive (3)	calm (3)	female	angry
appropriately reserved		too chilled		approachable (6)	aggressive (3)	good sense of humour (5)	less sensitive	overreacts
chilled/relaxed (2)		defensive		caring (6)	aloof	good looking	quiet (2)	childish
confident (2)		embarrassing conduct		communicative	bully (2)	jolly	sensitive	immature
direct		emotional		considerate	closed	male (2)	shy	insecure (4)
eager		flaky		direct	demanding, denigrating	mature (2)	values	irritable
energetic (2)		flustered		empathetic	disrespectful	optimistic (2)	extravert	miserable
engaged (2)		nervous		excellent listener	harassing attitudes	quiet (2)		moody
enthusiastic (3)		no boundaries		friendly (7)	hypocrisy	uptight		quiet
hardworking (6)		pessimistic (2)		friendly but slightly guarded	impatience (3)			sarcastic
interested (2)		poor grammar		good listener (5)	irresponsible and not understanding at all			sensitive
judicious with words		pushy		humour	irritable (2)			sour
motivated (2)		speaks like a teenager		kind (3)	irritating			too sensitive
not too emotional		unconfident (2)		liked	judgemental			
passionate (2)		uninterested (5)		open (2)	mean humour			
positive (2)		unmotivated (4)		patient (2)	overly friendly			
prioritises results over knowledge transfer/development		unprofessional		polite (2)	passive aggressive			
self motivated				polite and honest	poor communicator			
stressed				respectful (2)	poor listener			
unemotional				supportive (3)	reserved			
unrealistic				talkative	rude (5)			
				valued	shy			
				warm (2)	talkative			
				warm and friendly	threatening			
					too chatty			
					unapproachable (2)			
					unfriendly			

Cluster 5

	Work approach			Relates to others			Personal	
always on time	ideal	avoids work		believable	know all	decent		don't know well enough to comment
appropriate		busy fool		credible		not related		little bit slow
attention to detail		doesn't think flexibly		fair (2)		one job		no understanding
careful (2)		easy going		generous				presumptuous
clever/smart (2)		gets too involved with work		helpful (7)				stupid
creative (2)		gives up easily		honest (5)				weak
curious		goes with the flow		makes people feel welcome				
dedicated (6)		impractical		puts others first				
easy going (2)		just does the job (2)						
efficient (2)		just wants to get through the day						
full of ideas		lazy (8)						
hard working (2)		lets feelings get in the way						
makes things happen		liar (2)						
observant		looks for direction						
open to new ideas (2)		negative						
organisation of a lot of people		late						
practical		serious						
quick, diligent		unclear						
serious (2)		unsure about some things						
thoughtful (3)		weak (2)						
to the point		without concept						
well presented								
wide range of interests								

Cluster 6

Work approach		Relates to others				Personal	
adaptable		distractable	trustworthy		clique-ish		medicine/biology
conscientious		persuadable			detested		spontaneous
hardworking		relaxed			double-faced		
indestructable		scatty			non-confidential		
rapacious		unassertive			selfish-needy		
		undiligent			unavailable		
		unfocused			uncaring		
					uncommunicative		
					untrustworthy (2)		

Interpretation Notes

Alphabetical order													
Removed duplicates (total in brackets)													
Corrected spelling													
Re-worded to describe an individual e.g. dedication to dedicated													
Meaning unknown, as related to professionalism													
Separated into positive/negative/unclear columns, based on ratings i.e. if Xprof & Ideal rated on something, considered positive by participant													
	unclear - single user, ratings inconsistent, but 'ideal' rating is the trump column is decided by ideal, but unclear noted if appears inconsistent in reality												
	mixed - multiple users, apparently disagreeing over whether it is a positive or negative attribute												

Appendix D: Q Set Development

Statements

Themes: 19

Recommended number: 30-50 statements (currently 42)

Condition of instruction: Importance (debate regarding whether agreement indicates ranking, and whether it has a true mid-point); should be derived from research question:

1. For a person to be professional, they must...
2. When I meet a professional, they should...
3. In my opinion, a professional is...
4. In my opinion, professionalism is...

Where 'customer', read any individual requiring a service from the professional (customer/client/patient/service user).

Statements: Can be a mixture of positively and negatively loaded statements, but need to make sense in the context of the condition of instruction

Theme	Professional	Neutral/both	Unprofessional
Interpersonal skills	Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful		
Communication	Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang		

	Be able to express themselves clearly, and in a focused and articulate way		
Image management	Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others		Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)
Relationships with colleagues	Recognise that having positive relationships with colleagues is part of being good at your job	Not be dishonest, but it is OK if there are times when they do not tell the whole story	
	Work collaboratively with colleagues, rather than competitively		
	Have a keen awareness of where humour can become inappropriate		
	Be someone that demonstrates loyalty, someone that a colleague can rely on		
Leadership/being led	Be appropriately assertive and able to influence decisions		Not be resentful of work-related scrutiny and challenge
Humanistic	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone		Not come across as uncaring or callous
Social orientation	View themselves and other professionals as appointed leaders in society		
	Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs		

Sociological view (professionals are different) & how professionalism has changed	Take pride in their profession, and defend it to others	Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	
		Be defined by their quality or character, rather than their occupation or job title; professionalism less about what you do, and more about how you do it	
Reflection and self-awareness	Acknowledge, accept, and learn from their own mistakes		Not be defensive or easily upset in response to workplace events or change, or when being challenged
	Be perceived as demonstrating the required professional standards by those around them		
Independence and initiative	Be able to work independently and confidently according to their own initiative		
Relationships with customers	Build rapport with customers, in a way that encourages mutual respect and partnership		
	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice		
Boundaries	Keep their personal lives separate from their professional - personal issues or experiences should not influence actions or attitudes in the workplace		

Congruence	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only		Behave responsibly outside of the workplace, including being a positive role model for society
Competency/effectiveness and professional development/ education	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training		
	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done		
	Have a well-rounded set of skills – they must demonstrate both technical expertise and interpersonal skills		
	Be aware of the limits of their own competence, inform others about them, and remain within them		
Ethics and morality	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved		
	Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly		
Good/bad worker		Demonstrate a strict adherence to policy and procedural guidance at all times	

Wellbeing	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work		
Personal qualities and characteristics	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity		Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance
	Be able to manage stressful situations with objectivity rather than emotion		
	Embody characteristics such as honour, duty, justice, and courage		
Accountability	Welcome a sense that they are accountable to others		
	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers		

Appendix E: HTMLQ Coding

Configuration Settings

```
1  <?xml version="1.0" encoding="UTF-8" ?>
2
3  <!--
4  //
5  // =====
6  // FlashQ by Christian Hackert and Gernot Braehler
7  // =====
8  // Configuration file.
9  // #author      Christian Hackert <www.hackert.biz/flashq>
10 // #version     FlashQ version 1.0
11 // =====
12 //
13 -->
14
15 <config version="1.0" htmlParse="false">
16
17   <!-- title of the study -->
18   <item id="studyTitle">Defining professionalism: A comparison of subjectivity using the Repertory Grid Technique (RGT) and Q-methodology</item>
19
20   <!-- text-align property (left|right) -->
21   <item id="textAlign">left</item>
22
23   <!-- shuffle cards (true|false) -->
24   <item id="shuffleCards">true</item>
25
26   <!-- login required (true|false) -->
27   <item id="loginrequired">false</item>
28   <!-- login with common password (leave blank if not required) -->
29   <item id="loginPassword"></item>
30   <!-- URL to individual login script (leave blank if not required) -->
31   <item id="loginUrl"></item>
32   <!-- request mode (post|get) -->
33   <item id="loginUrlMethod"></item>
34
35   <!-- activate/deactivate optional steps (true|false) -->
36   <item id="showStep3">true</item>
37   <item id="showStep4">true</item>
38   <item id="showStep5">true</item>
39
40   <!-- HtmlQ only: disable the in-window back button -->
41   <item id="disableBackButton">true</item>
42
43   <!-- define form elements for step5; only displayed if showStep5 is true (delete tags if not required) -->
44   <!-- add label: <label>[STRING]</label> -->
45   <!-- add note: <note>[STRING]</note> -->
46   <!-- add input field: <input type="[text|textarea|radio|select|checkbox|rating2|rating5|rating10]" [VALUE]</input> -->
47   <!-- optional attributes: bg, id, maxlength, restricted, required, scale -->
48   <item id="form">
49     <label></label>
50     <label></label>
51     <label></label>
52     <label></label>
53     <label></label>
54
55     <label>Overall, what are some of the most common factors you thought about when completing the task? Did you find there were any themes that consistently came up in your reason</label>
56     <input type="textarea" required="false"></input>
57
58     <label>Please enter your age.</label>
59     <input type="text" required="false" maxlength="3" restricted="0-9"></input>
60
61     <label>What gender do you identify with?</label>
62     <input type="radio" required="false">Female;Male;Other</input>
63
64     <label>What is your nationality?</label>
65     <input type="textarea" required="false"></input>
66
67     <label>Is the majority of your profession or occupation conducted in your native language?</label>
68     <input type="radio" required="false">Yes;No</input>
69
70     <label>What is your current or most recent occupation/profession?</label>
71     <input type="textarea" required="false"></input>
72
73     <label>For how many years have you been in your current or most recent occupation/profession?</label>
74     <input type="text" required="false" maxlength="2" restricted="0-9"></input>
75
76     <label>For how many years have you been in the workplace in total, within your current or most recent, and any other, occupations/professions?</label>
77     <input type="text" required="false" maxlength="2" restricted="0-9"></input>
78
79     <label>Which of the following categories does your current occupation/profession most closely resemble?</label>
80     <input type="radio" required="false">Administration e.g. administrator, office manager;Dental profession e.g. dentist;Education profession e.g. teacher, lecturer;Health care pro:
81
82     <label>If you answered "Other" to the question above, please type your occupation/profession below.</label>
83     <input type="textarea" required="false"></input>
84
85     <label>What is the highest level of education you have completed ?</label>
86     <input type="select" required="false">Level 1 e.g. GCSEs with grades D-G, level 1 award/certificate/diploma/ESOL/essential skills/functional skills, NVQ level 1, or equivalent;L
87
88   </item>
89
90   <!-- URL for data transmission via POST/GET (leave blank if not required) -->
91   <item id="submitUrl">exe.php?do=save</item>
92
93   <!-- request mode (post|get|firebase) -->
94   <item id="submitUrlMethod">firebase</item>
95
96   <!-- data transmission via e-mail, user must have an e-mail program like Outlook (leave blank if not required) -->
97   <item id="submitMail">yourdomain.com</item>
98 </config>
```

Language Settings

```
1 <?xml version="1.0" encoding="UTF-8"?>
2
3 <!--
4 //
5 // In order to translate HtmlQ or to customize the texts,
6 // just change the text between the opening and the closing
7 // tag. The attribute 'id' is the name of the variable
8 // and must remain unchanged.
9 //
10 // You can use HTML-tags, but you have to braces instead
11 // of the normal angle brackets, otherwise the import will fail.
12 // Example: {b}bold{/b}
13 //
14 // If your are using German Umlauts and other special
15 // characters, make sure, that you save this XML-file
16 // as an UTF-8-document. White-space inbetween the
17 // nodes are ignored. Any white-space within the text-
18 // nodes, including leading or trailing whitespace and
19 // line breaks, is preserved.
20 //
21 //
22 // =====
23 -->
24
25 <language version="1.0" htmlParse="true">
26
27 <!-- misc -->
28 <item id="btnContinue">Continue...</item>
29 <item id="btnClose">Close</item>
30 <item id="btnHelp">Help</item>
31 <item id="btnAgreement">{br}{br}{br}Agree</item>
32 <item id="btnNeutral">{br}{br}{br}In my opinion, for a person to be professional they must...{br}{br}Neutral</item>
33 <item id="btnDisagreement">{br}{br}{br}Disagree</item>
34 <item id="btnTransfer">Submit data</item>
35 <item id="btnMail">Send via email</item>
36 <item id="btnPrint">Print data</item>
37 <item id="btnExit">Exit</item>
38 <item id="selectItem">Please select...</item>
39
40 <!-- Warning when user is trying to leave the page (this item was added for HtmlQ) -->
41 <item id="leaveSiteWarning">If you leave the site, your answers will be lost.{br}{br}However, you can re-start the survey and complete a new one at any time using the same link.</item>
42
43 <!-- HtmlQ only: In-App Back Button -->
44 <item id="backButton">Back</item>
45 <item id="fillInRequiredFields">Please complete all required fields.</item>
46
47 <!-- errors -->
48 <item id="errorHead">Error!</item>
49 <item id="errorWindowTooSmall">Please maximize your browser for using this application.</item>
50
51 <!-- welcome screen (leave blank to skip screen) -->
52 <item id="welcomeHead">Welcome!</item>
53 <item id="welcomeText">Participant information sheet 25/04/17.{br}{br}Title of study: Defining professionalism: A comparison of subjectivity using the Repertory Grid Technique
54 <!-- user login, only displayed if access is restricted (see config-file) -->
55 <item id="loginHead">User code</item>
56 <item id="loginText">Please enter your user code. Please note, that in this demo any user code will be accepted.{br}{br}{i}Tip: This is an optional step and you can deactivate it i
57 <item id="loginFormHead">User code</item>
58 <item id="loginInput">Please insert your user code.</item>
59 <item id="loginInvalidInput">User code invalid</item>
60 <item id="loginNoConnection">Connection to server failed. Please try again.</item>
61
62 <!-- introduction (leave blank to skip popup) -->
63 <item id="introHead">Participant consent form</item>
64 <item id="introText">Please read the following information carefully. By clicking the 'continue' button you are confirming your agreement to all of the following statements.{br}{br}
65
66 <!-- step1: rough sorting into three piles (leave blank to skip popup)-->
67 <item id="step1Head">Step 1 of 5</item>
68 <item id="step1Text">You will be presented with a series of statements which people might make about professionalism. You might agree with some, disagree with others, and find some
69
70 <!-- step2: sorting (leave blank to skip popup) -->
71 <item id="step2Head">Step 2 of 5</item>
72 <item id="step2Text">You will now be presented with the statements from your three boxes again. Please look at the statements in your "AGREE" box first and read them again. You can
73
74 <!-- step3: check sorting, only displayed if showStep3= true (see config-file; leave blank to skip popup)-->
75 <item id="step3Head">Step 3 of 5</item>
76 <item id="step3Text">Now you have placed all statements on the sorting sheet, please go over your responses once more to check that you are happy with them. You can move statements
77
78 <!-- step4: comments on best/worst rated statements, only displayed if showStep4= true (see config-file; leave blank to skip popup) -->
79 <item id="step4Head">Step 4 of 5</item>
80 <item id="step4Text">Below you will find the responses that you most strongly agreed and disagreed with. Looking at your responses, please explain why you felt so strongly about th
81
82 <!-- step5: dditional questions, only displayed if showStep5= true (see config-file; leave blank to skip popup)-->
83 <item id="step5Head">Step 5 of 5</item>
84 <item id="step5Text">Finally, please tell me a little about yourself by answering the following questions. If you would prefer not to answer any or all of the questions, please lei
85
86 <!-- data transfer, only displayed if submitUrl is not blank (see config-file) -->
87 <item id="transferHead">Submit Data</item>
88 <item id="transferText">Thank you for taking part in this study. You have now completed the survey so please remember to submit your data by clicking the submit button.</item>
89 <item id="transferFailed">Data submission failed. Please try again or mail your results via email/post.</item>
90 <item id="transferOk">Thank you - your responses have been recorded.{br}{br}You can now close your browser window.</item>
91
92 <!-- send data via mail, only displayed if submitUrl is blank (see config-file) -->
93 <item id="mailHead">Submit Data</item>
94 <item id="mailText">You can either submit your data either via email or post. In the former case you'll need an email client like Outlook, in the latter a printer.</item>
95
96 <!-- e-mail-body -->
97 <item id="mailBody">Thank you for participating in our survey. Please do not modify the following text:</item>
98
99 <!-- print out -->
100 <item id="printoutText">Somebody{br}Some street 100{br}Some city{br}{br}Please send this printout to address mentioned above. Thanks for you help.</item>
101
102 </language>
```

Map Settings

```
1 <?xml version="1.0" encoding="UTF-8"?>
2
3 <!--
4 //
5 // In this file you can set up the score sheet. FlashQ can
6 // handle all types of distribution (including non-forced).
7 //
8 // EXAMPLE
9 // Let's have a look at a (very) simple example with
10 // three columns and a bell-shaped distribution:
11 //
12 // |=====|
13 // | -1 | 0 | +1 |
14 // |=====|
15 // | | | |
16 // |=====|
17 // |
18 // |=====|
19 //
20 //
21 // We have three columns, so we need three column-
22 // nodes in our XML-file. The column heads (-1, 0,
23 // +1) are saved in an attribute named 'id'. The
24 // number of rows per column (1, 2, 1) goes between
25 // the opening and the closing column-tag. If you
26 // want to have coloured column heads, use the 'colour'-
27 // attribute (hex code). All column-nodes are surrounded
28 // by a parent-node called 'map'. The finished file
29 // looks like the following:
30 //
31 // <map version="1.0" htmlParse="false">
32 //   <column id="-1" colour="FFD5D5">1</column>
33 //   <column id="0" colour="E9E9E9">2</column>
34 //   <column id="+1" colour="9FDFBF">1</column>
35 // </map>
36 //
37 // =====
38 //
39 --->
40
41 <map version="1.0" htmlParse="false">
42   <column id="+6" colour="8E9CFC">2</column>
43   <column id="+5" colour="8E9CFC">3</column>
44   <column id="+4" colour="8E9CFC">3</column>
45   <column id="+3" colour="8E9CFC">4</column>
46   <column id="+2" colour="8E9CFC">4</column>
47   <column id="+1" colour="8E9CFC">5</column>
48   <column id="+1" colour="8E9CFC">5</column>
49   <column id="+2" colour="8E9CFC">4</column>
50   <column id="+3" colour="8E9CFC">4</column>
51   <column id="+4" colour="8E9CFC">3</column>
52   <column id="+5" colour="8E9CFC">3</column>
53   <column id="+6" colour="8E9CFC">2</column>
54
55 </map>
```

Statements Settings

```
1 <?xml version="1.0" encoding="UTF-8"?>
2
3 <!--
4 //
5 // This is the place for your statements. The maximum
6 // number of statements is theoretically unlimited, but
7 // less than 50 statements are advisable.
8 // If you are using German Umlauts and other special
9 // characters, make sure, that you save this XML-file
10 // as an UTF-8-document. White-space inbetween the
11 // nodes are ignored. Any white-space within the text-
12 // nodes, including leading or trailing whitespace and
13 // line breaks, is preserved.
14 //
15 // EXAMPLE
16 // Let's define some statements concerning FlashQ:
17 // Statement 1: FlashQ is very usefull.
18 // Statement 2: I prefer offline surveys.
19 // Statement 3: FlashQ is too complicated to configure.
20 //
21 // Each statement requires a separate statement-node
22 // consisting out of an opening and a closing tag. The
23 // statement itself goes between this two tags.
24 // The unique statement id is saved as an
25 // attribute called 'id'. We have three example
26 // statements and we need therefore three statement-
27 // nodes and one parent-node.
28 //
29 // <statements version="1.0" htmlparse="false">
30 // <statement id="1">FlashQ is very usefull.</statement>
31 // <statement id="2">I prefer offline surveys.</statement>
32 // <statement id="3">FlashQ is too complicated to configure.</statement>
33 // </statements>
34 //
35 // =====
36 //
37 // -->
38
39 <statements version="1.0" htmlParse="false">
40 <statement id="1">Take the perspectives of others, regardless of status or discipline, to ensure that their own conduct is appropriate and respectful.</statement>
41 <statement id="2">Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang.</statement>
42 <statement id="3">Be dressed in a way appropriate to the task they are performing - casual clothing is OK for some tasks, but not others.</statement>
43 <statement id="4">Be able to express themselves clearly, and in a focused and articulate way.</statement>
44 <statement id="5">Work collaboratively with colleagues, rather than competitively.</statement>
45 <statement id="6">Not come across as uncaring or callous.</statement>
46 <statement id="7">Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.).</statement>
47 <statement id="8">Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone.</statement>
48 <statement id="9">Have a sense of authenticity and credibility - professionalism reflects the whole person, and is not a role to be played in certain situations only.</statement>
49 <statement id="10">Recognise that having positive relationships with colleagues is part of being good at your job.</statement>
50 <statement id="11">Not be dishonest, but it is OK if there are times when they do not tell the whole story.</statement>
51 <statement id="12">Be someone that demonstrates loyalty, someone that a colleague can rely on.</statement>
52 <statement id="13">Have a keen awareness of where humour can become inappropriate.</statement>
53 <statement id="14">Be appropriately assertive and able to influence decisions.</statement>
54 <statement id="15">Not be resentful of work-related scrutiny and challenge.</statement>
55 <statement id="16">View themselves and other professionals as appointed leaders in society.</statement>
56 <statement id="17">Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs.</statement>
57 <statement id="18">Be free to exercise their expertise and judgement, without answering to a hierarchy of managers.</statement>
58 <statement id="19">Take pride in their profession, and defend it to others.</statement>
59 <statement id="20">Acknowledge, accept, and learn from their own mistakes.</statement>
60 <statement id="21">Embody characteristics such as honour, duty, justice, and courage.</statement>
61 <statement id="22">Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it.</statement>
62 <statement id="23">Keep their personal and professional lives separate - personal issues or experiences should not influence actions or attitudes in the workplace.</statement>
63 <statement id="24">Be perceived as demonstrating the required professional standards by those around them.</statement>
64 <statement id="25">Not be defensive or easily upset in response to workplace events or change, or when being challenged.</statement>
65 <statement id="26">Be able to work independently and confidently according to their own initiative.</statement>
66 <statement id="27">Build rapport with customers, in a way that encourages mutual respect and partnership.</statement>
67 <statement id="28">Centre all of their activities on the needs of the customer, even if this requires self-sacrifice.</statement>
68 <statement id="29">Behave responsibly outside of the workplace, including being a positive role model for society.</statement>
69 <statement id="30">Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training.</statement>
70 <statement id="31">Have a well-rounded set of skills - they must demonstrate both technical expertise and inter-personal skills.</statement>
71 <statement id="32">Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved.</statement>
72 <statement id="33">Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done.</statement>
73 <statement id="34">Be aware of the limits of their own competence, inform others about them, and remain within them.</statement>
74 <statement id="35">Demonstrate a strict adherence to policy and procedural guidance at all times.</statement>
75 <statement id="36">Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly.</statement>
76 <statement id="37">Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work.</statement>
77 <statement id="38">Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers.</statement>
78 <statement id="39">Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity.</statement>
79 <statement id="40">Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance.</statement>
80 <statement id="41">Be able to manage stressful situations with objectivity rather than emotion.</statement>
81 <statement id="42">Welcome a sense that they are accountable to others.</statement>
82
83 </statements>
```

Appendix F: Q Factor Interpretation Crib Sheets

Factor Interpretation Crib Sheet

Summary		
<p>Factor 1 has an eigenvalue of 8.40 and explains 12% of the study variance. 6 participants were significantly associated with this factor. They were 5 males, with 1 participant declining to provide their gender. They had an average age of 39.17 years. Participants were drawn from a range of occupational sectors with 2 participants categorising themselves as working in education, 1 in academic research, 1 in leadership/management, 1 in sales, and 1 in dentistry. No participants shared the same job title, which were Regional Manager, PhD student, College Tutor, Master of Science in Education, External Sales, and Dentist. The mean number of years participants had been within their current role was 9.17 years (range 2-30 years), and in the workplace in total was 16.33 years (range 5-35 years). In terms of education levels completed, 1 had completed level 8 education (doctorate or equivalent), 3 had completed level 7 education (Masters degree or equivalent), and 1 had completed level 2 education (GCSEs or equivalent), with a further 1 participant declining to answer.</p>		
Items ranked at 12		
26	Be able to work independently and confidently according to their own initiative	
9	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	
<u>Abductive significance</u>		
Items ranked higher by F1 than by any other factor (excludes all ties)		
14	Be appropriately assertive and able to influence decisions	7
15	Not be resentful of work-related scrutiny and challenge	9
26	Be able to work independently and confidently according to their own initiative	12
39	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	9
41	Be able to manage stressful situations with objectivity rather than emotion	11
Items ranked lower by F1 than by any other factor (excludes all ties)		
3	Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	2
11	Not be dishonest, but it is OK if there are times when they do not tell the whole story	1
20	Acknowledge, accept, and learn from their own mistakes	8
22	Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	4

24	Be perceived as demonstrating the required professional standards by those around them	4
27	Build rapport with customers, in a way that encourages mutual respect and partnership	6
Items ranked at 1		
11	Not be dishonest, but it is OK if there are times when they do not tell the whole story	
7	Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	
<u>Statistical significance</u>		
Significant distinguishing statements not captured above		
4	Be able to express themselves clearly, and in a focused and articulate way	7
8	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	6
23	Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	4
29	Behave responsibly outside of the workplace, including being a positive role model for society	2
Significant distinguishers: P<0.05 P<0.01		
<u>Additional statements of interest (abductive)</u>		
5	Work collaboratively with colleagues, rather than competitively	10
10	Recognise that having positive relationships with colleagues is part of being good at your job	8
12	Be someone that demonstrates loyalty, someone that a colleague can rely on	6
13	Have a keen awareness of where humour can become inappropriate	5
17	Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs	5
21	Embody characteristics such as honour, duty, justice, and courage	7
25	Not be defensive or easily upset in response to workplace events or change, or when being challenged	11
30	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	8
31	Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills	10
32	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	11

33	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	10
37	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	9
38	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	7
<u>Comments</u>		
Comments of significant loaders by statement		
1	Wording of 'their' causes issue. Agree if taking all viewpoints is the intention.	
3	<p>Professionalism is expressed by behaviours and results, not by how people look. Some effort can be made in certain occasions, but it should not be a strict rule.</p> <p>Professionalism is not about personal preferences in my view, and the way one dress does not say anything about the way one think and do according to evidence-based knowledge.</p> <p>Ambiguous – what is casual?</p>	
5	Competitiveness between colleagues can hinder relationships and lead to stress, which can impact the results at work.	
6	Unsure as to how important this is. Coming across is different, but it does put barriers up so would agree	
7	<p>Professionality is expressed by behaviours and results, not by how people look.</p> <p>This is very superficial. People should learn to pay more attention to the way others act, not what they look like. The public can be so gullible and will respect anyone who wears a suit and company badge. They should overcome these stereotypes and use their critical thinking skills.</p> <p>Right to expression and protest, but cultural norms mean that this is too culturally defined. Varies by culture</p>	
8	'Unconditional' is too strong a word	
9	<p>A job is part of a person's life and it should come across at all times. A person should not live up to professionalism, rather professionalism should come natural as a result from a long process of personal and professional development.</p> <p>People can spot phoniness easily and it's very obvious when someone acts a certain way to ingratiate themselves with other or to show off etc.</p> <p>Can fake both</p>	
10	Where a breakdown occurs, there are exceptions, but agree generally	
11	<p>It depends... For example a manager can't always tell their staff what decisions are being made at a higher level before the decision is finalized. But you shouldn't lie to your customers just to sell a product. The purpose of the company is to provide the best service, not just to make profits. Eventually, lying or not telling the whole story will also have negative implications for the profits of the company.</p> <p>Context required; need to tell the whole story, but make it understandable to the patient.</p>	

12	Agree in principle, but would disagree with blind loyalty without context
14	Some firmness is needed when making evidence-based decisions. It should not be a discussion where you are led by the patient or you are not using your expert training
16	A professional could be a window cleaner, and not required to be any leader. 'Self-appointed' is problematic Says who? There to provide a service, not lead – leading is a different issue to professionalism.
18	I believe in order, and people who normally does this, does not follow rules, or agreements. Public purse and the big picture need to be taken into account
19	Disagree as this precludes challenging the profession
21	I think I was kind of thinking about leadership rather than just professionalism. But there are a lot of situations at work where staff and/ or superiors bully each other. A lot of the time, a majority of the staff are also entitled and lethargic and only care about how much money they earn etc, as opposed to the actual purpose of their work. In these situations it's important to have these characteristics. Very difficult not to agree! Keeps drawing the eye. Can all be misplaced e.g. bravery of a martyr. Happy with placement generally though
22	Disagree, as a combination of expertise and character are needed; must be good at the job and get outcomes as well as acting professionally
23	I find some people will share their personal lives, specially dramas with whoever, when you are at work, you are at it, not watching Jeremy Kyle.
25	when dealing with customers, they can make comments which can make you think they don't appreciate all the hard work you have done. being professional about it, means that you take it as way of improving yourself, and not let it get you.
26	In addition to statement 33, the ability to apply knowledge and skills should be done independently to reflect the professionalism of a person.
28	Professionalism is not about people pleasing in my view, but about making rational choices and actions. Agree, but within limits
29	Generally agree, but have had colleagues in the past who are professional in work but not outside. The GDC would disagree that they are professional. The GDC prize this, and to knock them is to knock us all so must agree
32	Must be an appropriate compass. Non-discriminatory, inclusive, and liberal values. Do the best you can in a humanist sense. Same for everyone, regardless of personal feelings.
33	To be professional is to me to have higher education and to be able to use the education to make evidence-based choices, therefore this statement is important for me in regard to professionalism.
34	One of the hardest things to do regarding professionalism. Need to listen to everyone. Overconfidence ignores criticism and becomes dangerous.
35	Policy & procedure are valuable but need to be questioned in order to improve and avoid 'fossilisation'. Nothing improves otherwise.

Overall comments of significant loaders

- I thought of people I consider as high professionals.
- I noticed themes such as communication with the colleagues, communication with the costumers, supporting other on the workplace, sense of integrity, appearance, and skills.
- I think I thought about leadership rather than just professionalism.
- I mainly thought about interactions with colleagues as they tend to be more demanding in my job than those with 'customers' (students).
- The association of professionalism with higher education, objectivity and evidence-based practice.
- Overall, [agree end] both share 'unselfishness'.
- [Agree end relates to] Professionalism but not in the old sense of a conspiracy to exploit people based on better expertise or knowledge.
- Agree with most instantaneously; others require a second look
- Quite a lot are mixed; there are some positive aspects to most
- Disagree ones are quite personal i.e. slightly different wording would change my view on them
- No issues with wording per se, but context – different internationally or in a primary vs secondary care setting; culturally defined
- Overall, "doing the right thing without regard to personal benefit"

Factor Interpretation Crib Sheet

Summary		
<p>Factor 2 has an eigenvalue of 2.35 and explains 13% of the study variance. 11 participants were significantly associated with this factor, although the following data are based on only 10 as 1 participant declined to complete the demographic questionnaire. They were 7 females and 3 males. They had an average age of 37.40 years. Participants were drawn from a range of occupational sectors with 3 participants categorising themselves as working in administration, 2 in leadership/management, 3 in sales, 1 in education, and 1 in project management. 2 participants shared the same job title as Project Managers, and the remainder were a Field Sales Engineer, Library and E-learning Assistant, Tours Coordinator, Financial Director, PhD Student, Waitress, and Leader, with a further 1 participant declining to provide their job title. The mean number of years participants had been within their current role was 6.20 years (range 1-30 years), and in the workplace in total was 14.80 years (range 1-40 years). In terms of education levels completed, 3 had completed level 7 education (Masters degree or equivalent), 4 had completed level 6 education (first degree or equivalent), and 3 had completed level 3 education (A levels or equivalent).</p>		
Items ranked at 12		
22	Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	
20	Acknowledge, accept, and learn from their own mistakes	
<u>Abductive significance</u>		
Items ranked higher by F1 than by any other factor (excludes all ties)		
3	Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	7
4	Be able to express themselves clearly, and in a focused and articulate way	11
6	Not come across as uncaring or callous	7
11	Not be dishonest, but it is OK if there are times when they do not tell the whole story	5
12	Be someone that demonstrates loyalty, someone that a colleague can rely on	10
13	Have a keen awareness of where humour can become inappropriate	9
20	Acknowledge, accept, and learn from their own mistakes	12
22	Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	12
38	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	8
Items ranked lower by F1 than by any other factor (excludes all ties)		
1	Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	7

8	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	3
9	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	2
21	Embody characteristics such as honour, duty, justice, and courage	4
29	Behave responsibly outside of the workplace, including being a positive role model for society	1
30	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	6
32	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	5
34	Be aware of the limits of their own competence, inform others about them, and remain within them	5
36	Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	3
37	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	3
Items ranked at 1		
29	Behave responsibly outside of the workplace, including being a positive role model for society	
16	View themselves and other professionals as appointed leaders in society	
<u>Statistical significance</u>		
Significant distinguishing statements not captured above		
26	Be able to work independently and confidently according to their own initiative	9
31	Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills	7
35	Demonstrate a strict adherence to policy and procedural guidance at all times	5
41	Be able to manage stressful situations with objectivity rather than emotion	9
Significant distinguishers: P<0.05 P<0.01		
<u>Additional statements of interest (abductive)</u>		
7	Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	2
10	Recognise that having positive relationships with colleagues is part of being good at your job	8
15	Not be resentful of work-related scrutiny and challenge	6

19	Take pride in their profession, and defend it to others	4
23	Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	10
24	Be perceived as demonstrating the required professional standards by those around them	8
26	Be able to work independently and confidently according to their own initiative	9
28	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	2
33	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	11
39	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	8
<u>Comments</u>		
Comments of significant loaders by statements listed above		
1	Or they wouldn't be a manager as this is what they're there for	
2	Half agree, half not Inappropriate jargon, 'layman's terms'	
3	Customer would expect this or won't take any notice of them (managers)	
4	Being professional is about how well you are able to articulate yourself, and how this comes across to colleagues and customers. Clear communication is necessary in any work place. A leader/manager and all levels of staff should be able to articulate clearly what expectations are of their team mates and ways in which people are required to work together.	
5	It is important that managers work <i>with</i> you	
6	Argumentative	
7	To make assumptions on someone based on appearance (such as visible tattoos or piercings or unnatural hair colour) is discrimination and this is not something any true professional would do. As with all the statements this is subjective and probably has a lot to do with context however, I would like to think that there is and should be more acceptance of peoples individual styles choices and how this should not stand as measure to asses an individual competence at their job. Shouldn't matter, as long as it is appropriate to the task Not fair Just the way things are; nothing to do with how you behave Doesn't matter to professionalism; how you do the job does	

8	<p>Trampled on</p> <p>Cannot, but may be 'forgiving' depending on experience</p> <p>Yes, in an ideal world, but it unlikely in reality</p>
9	<p>In the role is all that is needed; you can relax at other times</p> <p>In a work situation, yes – out of work is different</p>
10	Separate; important but not professionalism
11	Not specific to professionalism
12	<p>Professionalism is determined by the perceptions of others - if they don't feel that they can respect and rely on you then you have failed at being professional.</p> <p>trust is crucial in the workplace</p> <p>Want to be approachable but going too far becomes personal</p>
13	Tact & diplomacy
14	Depends on the job
15	<p>it is ok to disagree</p> <p>Similar to 23 & 41 – acting like a normal adult</p>
16	<p>A very junior member of staff has a responsibility to be professional, but they will not be a leader. This doesn't make their professionalism any less important or valued.</p> <p>leaders should be community led not led by stature.</p> <p>They shouldn't see themselves as above everyone else. They are equal</p> <p>Professional are experts in their field that gives them no moral or implied right to be a leader in the workplace or society.\nTo view themselves as \"leaders\" and weakens their true professionalism.</p> <p>Similar to 29 – expectation to sacrifice humanness</p> <p>Mother says 'they're supposed to be a doctor...' regarding gossip about behaviour, but this is unfair relating to behaviour outside of work</p>
17	Definite yes
18	<p>All professionals must be accountable to their peers and if appropriate their professional body.</p> <p>Ambivalent</p> <p>Still need to be accountable, this is the reason for general managers</p> <p>Certain managers may exercise judgement in the wrong way so should be accountable</p>
19	<p>Have it, but it is not necessary to be professional specifically</p> <p>Or they wouldn't be a manager as this is what they're there for</p>

20	<p>We can only improve if we accept we can and indeed do make mistakes, all professionals do, but good\" professional leans and improve.</p> <p>Very important, many don't</p>
21	<p>Not necessary</p> <p>Reminds me of the army, I hate the army</p>
22	<p>Professionalism is not related to the job you do or how senior/junior you are; it is whether or not you behave with dignity in the workplace, and leave your personality and personal life at home.</p> <p>Be respected for quality of character, not status</p>
23	<p>Similar to 15 & 41 – acting like a normal adult</p>
25	<p>Professional need to be challenged by their peers & customers to improve / validate decisions and explain why wrong decisions were with hindsight incorrect but bases on professional knowledge and judgement.</p> <p>Being professional</p>
26	<p>Being professional also means being reliable whilst taking on your own initiative</p> <p>Or they wouldn't be a manager as this is what they're there for</p>
27	<p>mutual respect is key in delivering any activity.</p> <p>Important for the job, but could be seen as unprofessional depending on method</p>
28	<p>Not fair</p> <p>Depends on the level of sacrifice</p>
29	<p>Personal and professional are separate. You might not want to work with a professional if you find out something about their private life that you don't like, but it is possible to be professional in the workplace and be irresponsible out of work - the two do not necessarily have to be related.</p> <p>Any professional can behave however they want outside the workplace if they are off duty. They don't necessarily have to be a good role model .</p> <p>Your behaviour out of work is your own - and as long at it was not illegal or discriminating then your own life out of work should not be used to judge your professionalism - all professional people need to relax.</p> <p>Not fair. Similar to 16 – expectation to sacrifice humanness. Follows closely behind 28& 37 in being not fair</p> <p>Yes, in an ideal world, but it unlikely in reality</p>
30	<p>Ambivalent</p> <p>Many don't see what they're doing wrong</p>

31	<p>Being professional is about being good at a wide range of things that all contribute to how you come across in the work place, and how well you do your job.</p> <p>Depends on the job</p> <p>Unrealistic</p>
32	<p>Not professionalism, just being a good person; you can be professional without this</p> <p>Linked to objectivity</p>
33	<p>They need to be good at what they do and be able to show why they are good at what they do</p> <p>I appreciate colleagues who use their skills in an effective and efficient manner. This builds confidence in the team that the person can be trusted with 'X' task and that they are the right person for the job. Someone that's applying themselves and knows their strengths can serve as an inspiration, especially if they can share or pass on their knowledge to junior colleagues.</p> <p>Important, but not necessarily unprofessional if not – ambivalent though</p>
34	<p>Ambivalent</p> <p>Related to self awareness</p>
35	<p>Shouldn't break rules, but may not get anywhere</p> <p>Gives clear outlines for all staff, it is no good having one rule for one</p>
36	<p>Ambivalent</p>
37	<p>Mental health issues are hard to self-recognise. Important, but not professionalism</p> <p>Not fair</p>
38	<p>Ambivalent</p> <p>Definite yes</p>
40	<p>Constant dominance = no compliments and pointing out flaws</p>
41	<p>I don't think it's possible for anyone to encounter stressful situations without showing a degree of emotion. I agree that people should not react in an overly emotional fashion, but sometimes a level of emotion in response to a situation can be a positive thing.</p> <p>Similar to 23 & 15 – acting like a normal adult</p> <p>Emotions = arguments</p>

Overall comments of significant loaders

- Professional and personal are totally separate, and your work and home personas may be completely different.
- Professionalism is unrelated to your role - it is just as important for junior and senior staff and for those who work with customers and those who don't.
- What you think of your own professionalism is only half the story - it is how you are perceived by others that is often the more important measure in the workplace.
- I see professionals as people who behave professionally in work, but i believe anyone can behave however they want out of work and still be professional in work.
- Work and personal life can be quite seperated.
- I thought about how I would want to be perceived, and how I would perceive managers.
- I noticed the two main themes were about personal ability in the workplace and also personality.
- Pending on the context of the person's job role I may have answered these differently so it was important to keep a fixed example in my mind as to what exactly I was answering the the questions against
- Professionals is a mind set.
- One aspect you did not explore is that professional never go on strike / withdraw their labour - Teachers strikes saw many people change their mind set that teaching is no longer viewed as a profession it's a vocation. What you have not asked is what people view as a profession as that would imform the research.
- General conduct; “don’t be an asshole”
- Professionalism = “adulging”
- Agree with some generally, but not specific to professionalism e.g. you can be professional without
- Agree end - Reflection of professional conduct; behaviours required to be professional
- Disagree end - Unrelated to professionalism. Call too much on sacrificing humanness for professionalism, and this is not necessary to be professional
- Regarding the neutral statements, these were very difficult to sort
- It was difficult to commit to the sort without feeling like it reflects on me i.e., it is important to be loyal, but this is not required for professionalism
- Positive relationships with colleagues are part of doing a job well, but you can be professional without them
- The agree pile are all things that are just ‘normal’
- It was difficult to choose just 2 to strongly agree with
- 10 and 12 are very similar
- Agree end - Same reasoning as 7 – more about how you do the job
- More about individual personalities
- Some managers think that they’re above accountability
- Agree end has a self-awareness theme re conducting self, objectivity

Factor Interpretation Crib Sheet

Summary		
<p>Factor 3 has an eigenvalue of 1.75 and explains 9% of the study variance. 4 participants were significantly associated with this factor. They were 2 females and 1 male, with 1 participant declining to provide their gender. They had an average age of 41.75 years. Participants were drawn from a range of occupational sectors with 2 participants categorising themselves as working in education, 1 in the fire and rescue service, and 1 in the security services. 2 participants shared the same job title of PhD student, alongside 1 Fire Chief and 1 Scenes of Crime Officer. The mean number of years participants had been within their current role was 8.66 years (range 3-5 years based on 3 provided responses), and in the workplace in total was 21.75 years (range 12-35 years). In terms of education levels completed, 2 had completed level 7 education (Masters degree or equivalent), 1 had completed level 6 education (first degree or equivalent), and 1 had completed level 5 education (Diploma of Higher Education or equivalent).</p>		
Items ranked at 12		
30	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	
1	Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	
<u>Abductive significance</u>		
Items ranked higher by F1 than by any other factor (excludes all ties)		
1	Take the perspectives of others, regardless of status or discipline, to ensure that their conduct is appropriate and respectful	12
2	Use language which is tailored to be comfortable and accessible to the person/people being spoken to, but without including slang	9
21	Embody characteristics such as honour, duty, justice, and courage	10
27	Build rapport with customers, in a way that encourages mutual respect and partnership	10
30	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	12
37	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	11
40	Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	6
Items ranked lower by F1 than by any other factor (excludes all ties)		
5	Work collaboratively with colleagues, rather than competitively	4
6	Not come across as uncaring or callous	4
19	Take pride in their profession, and defend it to others	3

23	Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	2
26	Be able to work independently and confidently according to their own initiative	3
28	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	1
31	Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills	6
35	Demonstrate a strict adherence to policy and procedural guidance at all times	2
38	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	4
Items ranked at 1		
7	Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	
28	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	
Statistical significance		
Significant distinguishing statements not captured above		
4	Be able to express themselves clearly, and in a focused and articulate way	9
8	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	7
25	Not be defensive or easily upset in response to workplace events or change, or when being challenged	7
29	Behave responsibly outside of the workplace, including being a positive role model for society	7
32	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	9
Significant distinguishers: P<0.05 P<0.01		
Additional statements of interest (abductive)		
9	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	10
11	Not be dishonest, but it is OK if there are times when they do not tell the whole story	3
12	Be someone that demonstrates loyalty, someone that a colleague can rely on	8
13	Have a keen awareness of where humour can become inappropriate	8
15	Not be resentful of work-related scrutiny and challenge	5
16	View themselves and other professionals as appointed leaders in society	2

17	Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs	5
18	Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	4
20	Acknowledge, accept, and learn from their own mistakes	9
22	Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	11
24	Be perceived as demonstrating the required professional standards by those around them	8
33	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	11
34	Be aware of the limits of their own competence, inform others about them, and remain within them	7
36	Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	6
39	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	8
<u>Comments</u>		
Comments of significant loaders by statements listed above		
1	This is for me about accountability and modelling appropriate behaviours. Drawing on the perspectives of all stakeholders is important in that sense and will ensure professional objectives are met	
7	This feels rooted in an old school of thought, where professionalism was defined by specific types of appearance. Visible tattoos and piercings are expressions of individuality, and being able to maintain these appearances in a professional workplace can contribute to de-stigmatising people who make alternative lifestyle choices. the differences can be a value	
9	Professionalism to me isn't just something you switch on or off as a situation requires, it is something that you are.	
16	I don't think that professionals should think of themselves as \"appointed leaders\" or \"better\" than others in society. They should think of themselves as equals to the people that are their customers, just as someone with a particular skillset that that person needs. After all, that person is likely to have skills that you don't and might need.	
17	often under-served and diverse or unmet needs are situations that needs more time and resources and are neglected cases	
18	Unfortunately, sometimes professionalism requires accountability to a hierarchy, and not everyone can be a maverick. Rather, I believe it's important to understand the limitations of your role, and behave in accordance to them without overstepping boundaries or taking on too much responsibility. This seems to suggest lack of accountability. A professional should have a moral and ethical code. Accountability is part of that, particularly if considering professionals in terms of professions which often define ethics in terms of codes of practice / behaviours.	

20	Part of being professional is to recognise and correct mistakes, if you can't do that, you can't improve and improvement is professional, acting like a sulky child because you did something wrong is not.
22	This for me is about living the experience of being a professional. Often people can talk the talk, but delivery and consistency of behaviour is important in a professional.
28	Self-sacrifice, for me, sounds punitive. I think a professional should have regard for themselves this may suggest acting unethically. This is particularly as demonstrating professionalism is also about leadership, irrespective of hierarchical level and modelling of appropriate behaviours is part of that.
29	a correct standard style of life outside the workplace can be sufficient
30	Learning is a lifetime process, and true professionals recognise opportunities for self development and growth. Rather than endorsing that they have all of the skills they need, they remain open to educating themselves and participating in formational activities within their workplace.
33	This sentence feels like the bottom line of professionalism - at the end of the day, you must be able to utilise your skills to get the job done. The way in which people go about this task shapes the types of professionalism that other embody, some more, some less, but this set of skills feels absolutely essential.
35	Adherence to policy and procedures is good, to an extent, after all they've probably been put in place for a reason. However blind adherence is not good. Policies and procedures that are wrong or detrimental, or for whatever reason no longer serve the company as they did before should be challenged so they can be changed or improved upon. Some things are put in place for very valid reasons, but don't work as stated in the real world situations - they need to be adapted to take both sides into account.
37	serving the most difficult cases to improve the overall abilities
Overall comments of significant loaders	
<ul style="list-style-type: none"> • When completing the task, I reflected on how important personal characteristics felt to me when I was ranking the categories, but at the end of the day, the most important professionalism theme was being able to get the job done effectively. • Next important were being a model citizen, good co-worker and overall good person, but these ranked secondmost to doing what one was hired to do effectively. • Accountability, behaviours, consideration of other perspectives, importance (or not) of appearance - in its broadest sense. • I was thinking about how I would like or expect others to act to appear professional and also how I would want to act to appear professional myself. 	

Factor Interpretation Crib Sheet

Summary		
<p>Factor 4 has an eigenvalue of 1.64 and explains 8% of the study variance. 3 participants were significantly associated with this factor. They were 1 female and 2 males. They had an average age of 42.66 years. Participants were drawn from a range of occupational sectors with 1 participant self-categorising as working in financial services, 1 in dentistry, and 1 in leadership/management. No participants shared the same job title, which were Actuary, Dentistry, and Company Director. The mean number of years participants had been within their current role was 13.66 years (range 2-27 years), and in the workplace in total was 23 years (range 12-30 years). In terms of education levels completed, 1 had completed level 8 education (doctorate or equivalent), 1 had completed level 6 education (first degree or equivalent), and 1 had completed level 4 education (Certificate of Higher Education or equivalent).</p>		
Items ranked at 12		
32	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	
9	Have a sense of authenticity and credibility – professionalism reflects the whole person, and is not a role to be played in certain situations only	
<u>Abductive significance</u>		
Items ranked higher by F1 than by any other factor (excludes all ties)		
7	Not make unconventional choices about their personal appearance (e.g. visible tattoos or piercings, unnatural hair colour, etc.)	4
8	Live by values such as dignity and respect, compassion, forgiveness, and unconditional positive regard for everyone	10
17	Demonstrate a commitment to prioritising the under-served and those with diverse or unmet needs	7
19	Take pride in their profession, and defend it to others	6
28	Centre all of their activities on the needs of the customer, even if this requires self-sacrifice	7
29	Behave responsibly outside of the workplace, including being a positive role model for society	8
31	Have a well-rounded set of skills – they must demonstrate both technical expertise and inter-personal skills	11
32	Have a strong ethical compass and be guided by sound moral values, regardless of the situation or those involved	12
34	Be aware of the limits of their own competence, inform others about them, and remain within them	10
35	Demonstrate a strict adherence to policy and procedural guidance at all times	9
36	Recognise that what the customer wants is not always what the customer needs, and act in their best interests accordingly	7

42	Welcome a sense that they are accountable to others	8
Items ranked lower by F1 than by any other factor (excludes all ties)		
4	Be able to express themselves clearly, and in a focused and articulate way	5
10	Recognise that having positive relationships with colleagues is part of being good at your job	3
13	Have a keen awareness of where humour can become inappropriate	4
14	Be appropriately assertive and able to influence decisions	2
15	Not be resentful of work-related scrutiny and challenge	3
18	Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	1
25	Not be defensive or easily upset in response to workplace events or change, or when being challenged	3
33	Be able to apply their knowledge and skills in a way that is efficient and effective at getting the job done	2
39	Be flexible enough to tolerate, embrace, and work effectively with complexity and ambiguity	4
41	Be able to manage stressful situations with objectivity rather than emotion	5
Items ranked at 1		
18	Be free to exercise their expertise and judgement, without answering to a hierarchy of managers	
16	View themselves and other professionals as appointed leaders in society	
<u>Statistical significance</u>		
Significant distinguishing statements not captured above		
37	Take care of their own physical and mental wellbeing, including having the insight to recognise how this might impact on their work	6
Significant distinguishers: P<0.05 P<0.01		
<u>Additional statements of interest (abductive)</u>		
3	Be dressed in a way appropriate to the task they are performing – casual clothing is OK for some tasks, but not others	5
5	Work collaboratively with colleagues, rather than competitively	8
12	Be someone that demonstrates loyalty, someone that a colleague can rely on	6
20	Acknowledge, accept, and learn from their own mistakes	11

21	Embody characteristics such as honour, duty, justice, and courage	7
22	Be defined by their character, rather than their occupation or job title; professionalism is less about what you do, and more about how you do it	9
23	Keep their personal and professional lives - personal issues or experiences should not influence actions or attitudes in the workplace	10
26	Be able to work independently and confidently according to their own initiative	5
30	Actively increase their competence by taking relevant learning opportunities, including self-improvement and formal training	11
38	Recognise their responsibility to support colleagues, even where this may require uncomfortable conversations with peers	7
40	Not have a nature dominated by traits such as assertiveness, competitiveness, or dominance	4
<u>Comments</u>		
Comments of significant loaders		
3	Perceived as important for some e.g. patients, so student appearance is key. Context matters. However, practicality is also relevant as may need to wear trainers is running around on call, or no tie laces for infection control purposes. There are gender differences e.g. short and tie is pretty much a uniform for men, but it is more flexible and there is more of a range for women.	
4	Like 14 – both good things, but not necessarily professionalism; can be weak in these areas without being unprofessional. Someone just doing their job well. Linked to 14/10/26/33/5	
5	Someone just doing their job well. Linked to 14/4/26/10/33.	
7	Example of a recruited student with tattooed arms which were not seen at interview due to long sleeves. Has changed over time and is still changing. Strict with students but a nurse with blue hair is employed; students do not view her as professional. All about patients having confidence and trust in you – depends on the patient’s demographics. Means nothing in terms of professionalism. Ability to perform is unrelated to fashion choices.	
8	Nice idea but too strong terms; unrealistic. ‘Forgiveness’ is a sticking point/out of place/unsure of – same in 21 for ‘courage’.	
10	Generic rather than professionalism itself e.g. If you worked in a non-profession, this would still be expected therefore it is not professionalism. Someone just doing their job well. Linked to 14/4/26/33/5.	
11	More in the past but now have a duty to disclose e.g. mistakes. May be different in medicine. Relates to defensive medicine and litigation. Odd statement, more grey than the others. Not dishonest but can lie a bit – if just the first part of the statement, would be similar to 32 and it would be right under +5 but the second bit does not = professionalism. But can still only tell patients what they will understand. Professionalism requires honesty.	
12	Similar to 18/16/21 – in the middle as true in a way but old fashioned	
13	Big issue. Husband is giving an after dinner speech soon, but must be careful as it will be video recorded, and standards have changed as to expectations.	

14	<p>Is this actually professionalism? Like 4 – both good things, but not necessarily professionalism; can be weak in these areas without being unprofessional.</p> <p>Don't have to be assertive to be a professional. Only missing the word 'appropriately'. Being more assertive does not = professionalism.</p>
15	Can still resent it, but 'put up with it'.
16	<p>Old fashioned view; personally, 22 is more important.</p> <p>Similar to 18/12/21 – in the middle as true in a way but old fashioned.</p> <p>Very old fashioned, not today. Difference between a professional (old view) and professionalism (new view).</p> <p>Big statement. Almost better than everyone else – a job doesn't make you greater than non-professionals.</p>
18	<p>Would like but doesn't happen anymore. Especially in medicine, but is now all about targets e.g. there may be side effects but you still have to prescribe it.</p> <p>Similar to 12/16/21 – in the middle as true in a way but old fashioned.</p>
20	<p>Extremely difficult for students.</p> <p>Personal experience – problems generally arise when not using this e.g. stuck in their ways.</p>
21	<p>Original, old professionalism. Not now formally, but still true.</p> <p>Similar to 18/12/16 – in the middle as true in a way but old fashioned. 'Courage' is a sticking point/out of place/unsure of – same in 8 for 'forgiveness'.</p>
22	<p>Surprised by interpretations of professionalism – some students are impressed by what someone wears, for example, others by interactions.</p> <p>Not just having a title, it's how you act. Not all dentists are professional.</p>
23	Difficult to achieve but 'should' so agree.
26	Someone just doing their job well. Linked to 4/14/10/33/5.
28	Depends on how extreme; patients come first but how much = perception.
29	Have to or "big trouble". Can lose their entire career for dental students.
32	<p>If ethics are right, everything else will follow.</p> <p>Sense of doing the right thing, playing by the rules at all times despite the situation.</p>
33	Someone just doing their job well. Linked to 14/4/26/10/5.
35	<p>GDC says yes, but 'no' leads to advances in the field. Sometimes you need to push the boundaries to make things better.</p> <p>Feel that this should be higher but the inclusion of the word 'strict' is an issue as only guidelines; can be wrong to stick to them.</p> <p>Sense of doing the right thing, playing by the rules at all times despite the situation.</p>

37	<p>Yes, as would be struck of by GDC. Links to 38.</p> <p>“Physical” – could be wedded to your job and so can’t spend time on self, but still be professional.</p> <p>More physical aspect disagreed with. You can be in top shape but not be professional, in my experience of colleagues.</p>
38	Links to 37.
39	Would need this to be a good dentist, and students struggle with it. Same for all areas; everyone must do this.
40	<p>The old way to professionalism; shouldn’t now.</p> <p>Don’t want it, but there are people with those traits but others say they’re professional.</p>
41	Helps with stress. Removes hot headedness. Learning what works and what is acceptable – has changed over time. The change itself causes more unprofessional behaviour.
<p>Overall comments of significant loaders</p> <ul style="list-style-type: none"> • Disagree end - generally those that are irrelevant or important but not professionalism – good phrases • Just doing a good job vs a professional doing a job • Difference between old view and new view e.g. clothes, leaders in society, how you act not your job title, etc. • Used to be more male dominated e.g. assertiveness – no longer the same • Self improvement and admitting mistakes more important now than in the past – the amount of change is the key revelation for me • Some remain the same though e.g. ethics • Surprised punctuality, using phone, turning up, etc. were not included, as would use these indicators to judge students • Statements relating to conduct outside of the workplace, e.g. ethics – good but not part of professionalism • Outside work has nothing to do with in work, completely distinct from home life; different rules • Don’t want to be defined by job title outside of work • Understand that others may view the statements differently due to their profession e.g. appearance for barristers, societies’ view • Getting the job done where colleagues, customers, and everyone gets what they want from it so all are happy and you make money • Balance profit with customer service, etc. • Interesting to complete • Can see an amalgamation of personal experiences; nostalgic process • Agree end themes relate to a sense of duty and operating within guidelines, but more how you convey it in actions, mannerisms, or interactions • Disagree end is about individual characteristics such as how you show emotion or dress – these don’t mean you’re less professional • Regarding the emotional stuff, how you behave – you can have heated conversations and get wound up/get mad. This isn’t unprofessional, it just shows that you really care so it should be detached from professionalism 	

Appendix G: The Model of Interpersonal Professionalism

Model of Interpersonal Professionalism

Our new model of interpersonal professionalism suggests that professionalism is a dynamic, interpersonal phenomena that cannot be separated from contextual factors, both internal and external. The model considers professionalism as both a *produced* and *received* behaviour, acknowledging that it is beyond the intentions of a professional, it also encompasses the way that those intentions and behaviours are perceived by the other party. Regardless of whether a professional is engaged in an occupation that is regulated, clients and consumers receive professional and unprofessional conduct as an interpersonal process. As a result, following the 'rules' of a given occupation may be insufficient to achieve professionalism if the other party's expectations are not fulfilled.

To better explain the model, consider a two-person interaction between a professional, referred to as the *producer* and a client, referred to as the *receiver*. Beginning with the *producer*, Figure 1 list various factors that affect their approach to an interaction in a professional context. These factors provide an *internal context* to their conduct.

Figure 1. Factors affecting the producer's interpersonal approach.

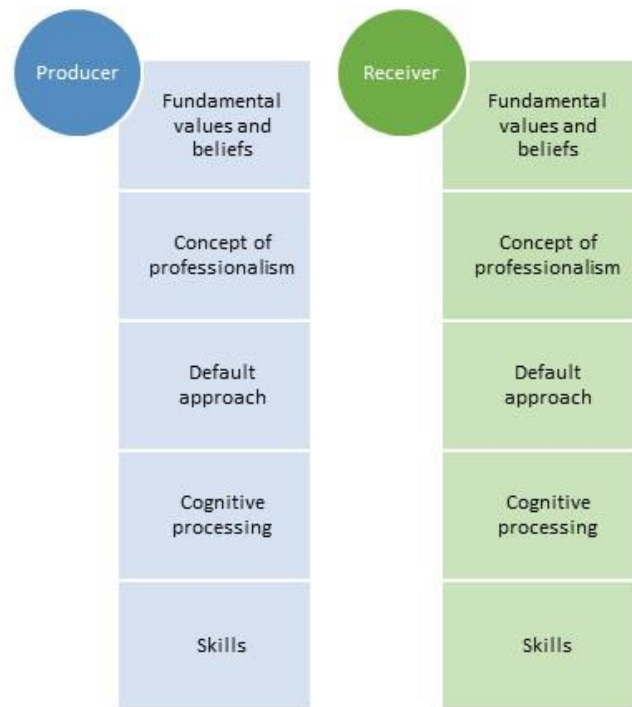


- The *producer's* **fundamental values and beliefs** affect all of the other internal contextual factors. These values and beliefs govern the way they see the world, what they most want from it, and how they view their own role within it.
- These values also affect the way that the *producer* views professionalism. Their **concept of professionalism** might be similar to those of other people, but it could also be very different.
- When faced with unfamiliar situations, the *producer* will use their values and concept of professionalism as a basis for their **default approach**. This approach may be flexible or rigid in nature, depending the *producer's* cognitive processing tendencies. The *producer* might categorise new information in ways that reinforces their default view making it more resistant to change or be likely to learn from experiences and improve the effectiveness of their default.
- The ability to adjust the default approach relates to the *producer's* range of **skills** and abilities. They must be capable of accurately identifying new interpersonal information in a range of forms such as verbal or non-verbal, emotional or behavioural feedback. They must also have the ability to learn from new information and to use this rapidly and effectively to plan, execute, and self-regulate their ongoing behaviour.

Traditional models of professionalism focus less on these internal contextual factors and more on the external context provided by the culture associated with a profession. A cornerstone of such models is the assumption that the factors listed above require little consideration, as upon developing a strong professional identity associated with the formal and informal operational and cultural, rules of an occupation, professional conduct will follow intuitively as a result.

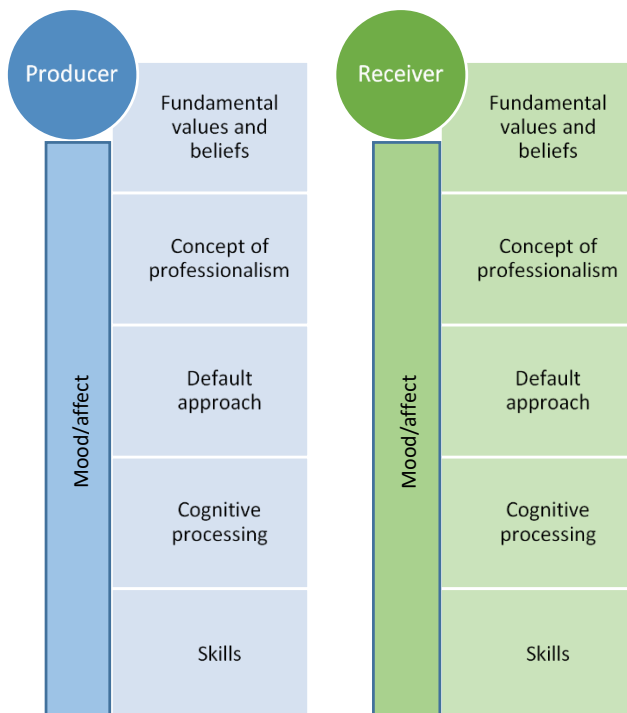
However, the model of interpersonal professionalism recognises that the same factors affecting the *producer's* professionalism also apply to the way that the *receiver* perceives it, as shown in Figure 2. These factors result in a dynamic between the *producer* and *receiver* that forms the basis of our model. Specifically, the level of success with which the *producer* is perceived as professional hinges upon their ability to tailor their approach to the internal context of the *receiver*. This ability to do this requires the *producer* to be able to perceive, identify, and respond effectively to the external cues and feedback that a *receiver* provides.

Figure 1. Factors affecting the producer's and receiver's interpersonal approaches.



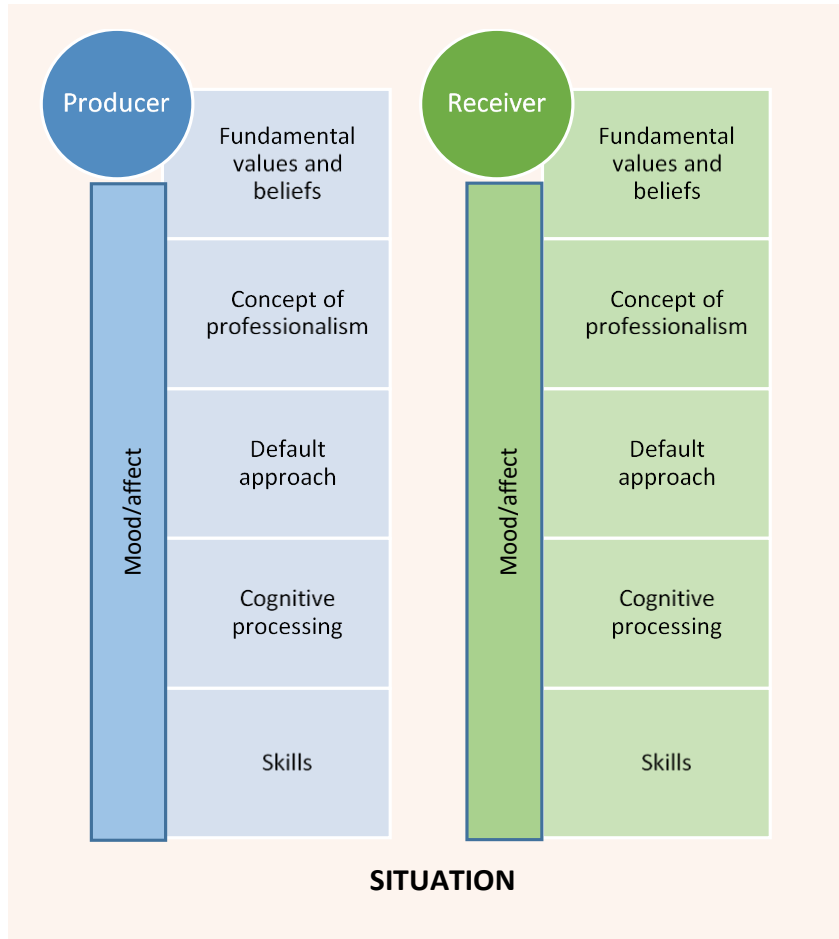
In addition to this, there is also the additional layer of internal context provided by the changeable emotional state of each individual, as shown in Figure 3. The mood of both the *producer* and *receiver* on the day that they meet will affect all of the other internal contextual factors, and so in essence provides a 'lens' through which each of the other factors must operate, a lens that is subject to change on a minute-by-minute basis.

Figure 2. The role of emotion.



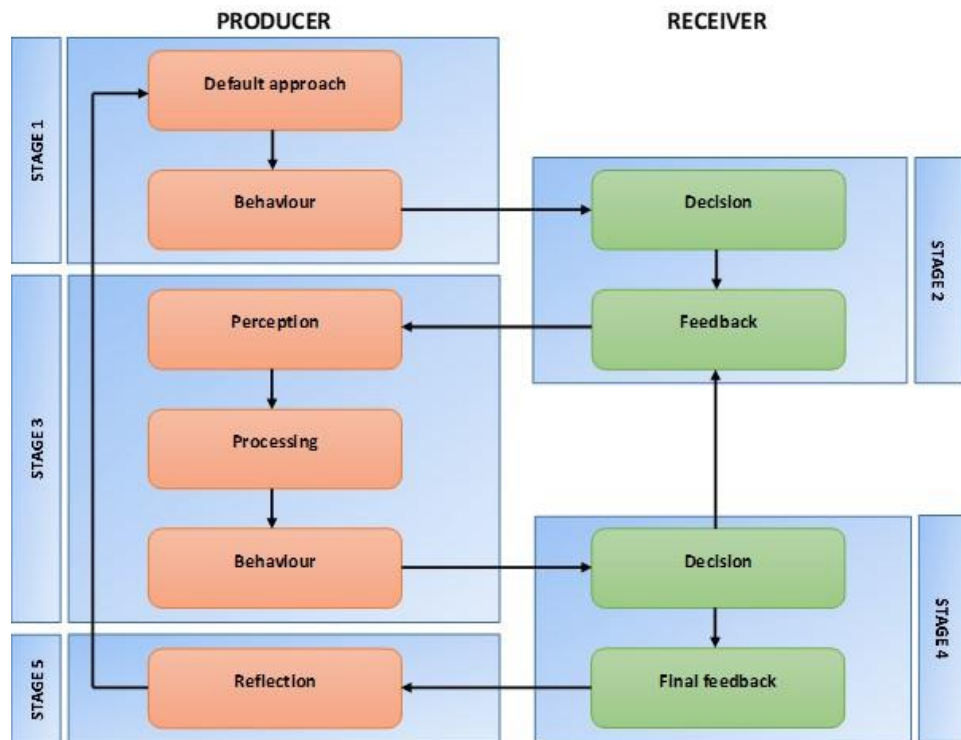
Finally, there is also the role of the external situational contextual factors of the interaction. Situational factors, such as the purpose of the interaction, the consequences of it for each individual, and whether anyone else will be observing the exchange, interact with all of the internal factors previously described, as shown in Figure 4. At times internal factors might have more influence over each person's behaviour, but at others the situational factors might take the advantage. Constant fluctuations in internal and external contextual factors account for a potentially infinite number of behavioural responses from both the *producer*

Figure 1. External contextual factors.



The implications of this model for developing, assessing, and measuring professionalism can be considered under a number of stages characterising a typical two-person professional interaction, as shown in Figure 5.

Figure 1. *The stages of a professional interpersonal encounter.*



Stage 1: Default approach

The *producer* enters an unfamiliar situation whereby they encounter a new colleague or client, the *receiver*. In the absence of information about the *receiver's* internal context, the *producer* sticks closely to their default approach. However, what that default looks like and how closely they stick to it will be affected by the interaction of internal and external contextual factors. For example;

- What are the *producer's* motivations? What outcome holds the most subjective value?
- How rigid is their concept of professionalism? Do they struggle to stray from it, even if it has been unsuccessful in the past?
- What are their cognitive tendencies? Are they likely to recall positive outcomes from using their default approach more than negative, and therefore persist with it?
- Does the *producer* have the skills to alter their default approach? Do they have the skills to learn from previous experiences and regulate future behaviour to make changes?
- What is the *producer's* mood like today?
- How familiar is the situation? Is the encounter something they undertake on a daily basis, even though the other person is unknown to them?

Stage 2: Initial decision

The *receiver* also uses their default approach upon entering the interaction, but this is likely to have a different emphasis to the *producer*. The initial decision they make as to whether the *producer* is professional will be a product of the following interacting factors:

- What are the *receiver's* motivations? What outcome holds the most subjective value?
- How rigid is their concept of professionalism? Do they struggle to look favourably on behaviours that contravene it?
- What are the *receiver's* cognitive tendencies? What are their expectations based on previous experiences?
- Does the *receiver* have the skills to alter their behaviour if they need to, such as to tell the *producer* that they are unhappy? Is it important to them to provide this feedback?
- What is the *receiver's* mood today?
- Is the situation familiar to them? Is the encounter something familiar and comforting, or likely to provoke anxiety?

Stage 3: Perception of feedback information

As the encounter progresses, the *producer* must use their skills to manipulate their behaviour in way that will achieve the desired outcome. Their ability and desire to do this is affected by internal and external factors:

- How skilled is the *receiver* in expressing their initial feedback?
- Does the *producer* have the skills to accurately identify and use this feedback? Can they rapidly plan and regulate their behaviour successfully based on this new information?
- How does the *producer* tend to process feedback cognitively? Do they tend to learn from it, or ignore the more negative aspects?
- Have the *producer's* motivations changed since entering the encounter? Are they still prioritising the same values, or has something about the situation or *receiver* changed them?
- Has the *producer's* mood changed?

Stage 4: Final decision

The *receiver* will continue to make judgements about the *producer's* professionalism and provide feedback regarding those judgements on an ongoing basis as the encounter continues:

- How skilled is the *receiver* in expressing their feedback clearly and are they motivated to do so?
- Have the *receiver's* motivations changed since entering the encounter? Are they prioritising the same values, or has something about the situation or *producer* changed them?
- Has the *receiver's* mood changed? If the encounter is not going well, has a shift in their mood begun to interfere with their ability to deliver feedback?

Stage 5: Reflection

Upon conclusion of the encounter, the *producer* has the opportunity to reflect on and learn from it, but the likelihood of their doing so, and the effects of this, are subject to the following considerations:

- How expressive and clear was the *receiver* in their feedback?
- Does the *producer* have the skills to accurately identify that feedback on a post hoc basis? Can they learn from it in a way that might change their default approach?
- What biases might be affecting their post hoc processing? Are the recalled details altered in ways that allow them to view their own behaviour as in line with their underlying values, for example?
- Now that the encounter has ended, have the *producer's* motivations changed? For example, is their goal now to justify their own behaviour, learn from the experience and improve, or create a professionally acceptable version of the encounter for formal records or challenge?
- Has the *producer's* mood changed since the encounter concluded, and is this affecting their ability to accurately recall the details of the encounter?

Summary

Our model of interpersonal professionalism suggests that regardless of the field a professional works in, their professionalism may be based more on psychological and interpersonal factors than rules specific to the occupational sector. As a result, the development of professionalism may benefit from a shift in focus from it being a solely *produced* behaviour, to one that is based on the dynamic and social interaction of multiple individuals, each of their internal contextual factors, and the external contextual factors of the situation.

In light of this, please consider the comments that you or another stakeholder might provide on the model during this session. Please consider its strengths and weaknesses in the context of your professional experience, and where improvements might be made. The comments you discuss may not be your own personal opinion, but feedback you would anticipate from a colleague, service user, or other stakeholder.