Reliability of Cervical Vertebrae Maturation (CVM) Staging Method using Full versus Cropped Lateral Cephalograms.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctorate of Dental Sciences

By

Julia Therese Mangan

August 2017

Acknowledgements

I would like to take this opportunity to express my gratitude to my supervisor Dr. Jayne Harrison for her constructive advice, unwavering support and endless encouragement in the preparation of this thesis. Her intellect, patience and constant willingness to help, including spending the time to read my thesis has been invaluable. I have been very fortunate to have had the opportunity to work alongside such an inspiring mentor. If I become only half the orthodontist that you are, it will be one of my greatest accomplishments.

I would like to extend my thanks to Dr. Burnside for his statistical support and advice. I am also very grateful for all the observers that participated in this study. Without their enthusiasm and efforts, this thesis would not have been possible. I would like to thank my colleagues for their support throughout this process.

I am forever thankful for my wonderful family and friends, for their support, understanding and constant encouragement that made it possible for me to complete this thesis.

Finally, I would like to thank my Mother for instilling in me a strong work ethic and the belief in myself that I could achieve anything I set my heart to. You have always encouraged me to pursue my dreams and have been selfless in your love and support of me. I stand where I am today because of you and all the sacrifices you have made for me over the years.

Table of Contents

Abstract

Reliability of Cervical Vertebrae Maturation (CVM) Staging Method using Full versus Cropped Lateral Cephalograms.

MANGAN J.T*, HARRISON J.E., BURNSIDE G. (School of Dentistry, University of Liverpool)

Objectives: Staging of cervical vertebral maturation (CVMS) using full lateral cephalograms has been found reliable. However, visibility of the dentition may influence staging so using cropped images may reduce such bias. The primary objective of this study was to determine whether CVMS reliability was improved by using cropped images. The secondary objective was to determine whether the level of orthodontic experience affected CVM reliability.

Design/setting: Two-phase reliability study, with a 3-month wash out period, of the modified CVMS method was undertaken over 2 consecutive audit meetings of Mersey and North Wales Audit Group during 2016.

Methods: A group of 22 orthodontic clinicians were trained to use the modified CVMS method for the assessment of mandibular growth. 144 consecutive images, taken at Liverpool University Dental Hospital, in full or cropped format were independently assessed. The images were displayed on individual computer monitors, to replicate the clinical setting, in two different random orders. Intra- and inter-observer agreement were evaluated using weighted kappa, (K^W) and Fleiss-Nee-Landis kappa, respectively.

Results: Overall intra-observer agreements for cropped and full images were 'almost perfect' (quadratic K^W 0.84, mean agreement 97%). The overall inter-observer agreement for cropped images was 'moderate' with values of 0.46 (95% CI 0.45-0.47) for phase 1 and 0.43 (95% CI 0.43-0.44) for phase 2, mean agreement of 87%). The overall inter-observer reliability for full images was 'fair-moderate' with values of 0.44 (95% CI 0.44-0.45) for phase 1 and 0.40 (95% CI 0.39-0.41) for phase 2, with mean agreement 86%. Intra- and inter-observer observer reliability of orthodontic trainees for cropped images were 0.76-0.95 and 0.50, respectively. Intra- and inter-observer observer reliability of orthodontic trainees for full images were 0.75-0.92 and 0.46, respectively. Intra- and inter-observer observer reliability of orthodontic specialist for cropped images were 0.65-0.89 and 0.44, respectively. Intra- and inter-observer observer reliability of orthodontic specialist for full images were 0.75-0.90 and 0.43, respectively.

Conclusions: The intra-observer agreement values for classifying the vertebral stages with the CVM method were 'almost perfect'. The inter-observer agreement for classifying the vertebral stages with the CVM method were 'fair to moderate'. The use of cropped images did not improve CVMS reliability. Clinical experience did not improve CVM staging reliability.

Chapter 1: Introduction

Optimal orthodontic results require the optimal timing of treatment. There are two key times when a child's growth can have a significant influence on the orthodontic treatment plan. Firstly, dentofacial orthopaedics can be utilised for the early correction of dento-skeletal discrepancies.[1-4] The accelerated growth period treatment is considered most effective for treating Class II malocclusions as it takes advantage of the differential growth potential between the maxilla and mandible and may be responsible for inducing additional growth.[5] The second situation arises when remaining mandibular growth may have an impact on treatment planning for orthognatic surgery or placement of implant for patients with hypodontia.[6, 7] Complete cessation of growth must be determined prior to finalising the treatment plan in order to avoid a relapse of the presenting malocclusion.[8]

The orthodontist is therefore required to evaluate the developmental status of children and adolescents and assess the remaining growth potential in order to formulate a treatment plan best suited to treat the individual's malocclusion and plan an appropriate retention strategy.[8] A typical growth pattern follows an orderly sequential pattern and is characterised by a growth rate that decreases from infancy before accelerating to maximum growth at puberty and a rapid deceleration indicating the near completion of growth.[9] Growth, being a dynamic and complex process, means that there is a great variability in its onset, duration and intensity.[10] This variability is strongly related to the individual's genetic, ethnic and environmental background.[11] This wide individual variability means that it is a challenge for the orthodontist to identify the period of maximum growth. Numerous methods have been proposed to aid the prediction of the adolescent growth spurt with research suggesting that chronological age and dental development do not necessary correlate well with maturational age.[12, 13] The development of secondary sex characteristics is a reliable indicator of the growth spurt however, this reflects a deceleration of growth and are therefore of limited value.[14] The ongoing development of biomarkers for the identification of the peak growth spurt is an exciting prospect, however, it is still in its infancy and not readily available clinically.[15-20]

Assessment of skeletal maturation is an effective method to predict an individual's stage of growth and can be readily assessed from hand-wrist radiographs and lateral cephalogram radiographs.[3, 21, 22] The determination of skeletal age using the hand-wrist radiograph has been extensively studied and has been found to be reliable and valid[21, 23] however, its routine use is precluded due to the need for an additional radiographic exposure for the patient[24] and the need for considerable skill in its interpretation.[25] The morphological changes that the cervical vertebral bodies undergo during development has been investigated with the modified cervical vertebrae maturation staging method found to be as valid as the hand-wrist radiographs in assessing skeletal age.[26] The advantages of the CVM staging method is that it eliminates the need for additional radiographic exposure for orthodontic patients since the vertebrae are readily viewed from the lateral cephalogram which is often taken as a pre-treatment record.[27] Also, the lateral cephalogram concerns an area of anatomy with which the clinician is familiar with and it presents a definitive stage that coincides with the peak in mandibular growth.[28] However, for any new evaluation tool to be implemented for clinical diagnostic purposes it must be reproducible. Many reliability studies have been undertaken over recent years in order to establish the value of the CVM method in the clinical diagnosis of the individual's level of skeletal maturity.[4, 22, 25, 29-41] However, controversy still remains around the reliability of this method, with much of the research reporting very different outcomes. From the large body of published research on CVM reliability, an explanation of the variability in the results can be identified by the broad methodologies utilised by the different authors. A recent study found that the modified CVM method was a reliable diagnostic tool when applied by a group of orthodontic specialist and orthodontic trainees using full lateral cephalometric images.[42]

However, controversy still exists as to whether cropped images should be used so to avoid the dentition influencing the assessment of skeletal developmental age.[33, 41] The purpose of the current study is therefore to establish whether the reliability of the modified CVM staging method is affected by the use of cropped lateral cephalometric images compared with full lateral cephalometric images.

Chapter 2: Literature review

2.1 Introduction

Growth and maturation, although closely related are in fact separate events. Growth is a continuous lifelong process that lacks uniformity having periods of both acceleration and deceleration based on complex endocrine regulation[9] and metabolic mechanisms. Maturation encompasses both growth and development and can be described as a series of successive transformations eventually culminating in the attainment of adult stature.[43] Growth prediction must consider the *'maturation level'* and the *'rate of maturation'*. The *maturation level* refers to the status of an individual relative to the full completion of growth. The *rate of maturation* refers to the time it takes a child to achieve key maturation events i.e. a child be described as delayed, average or advanced in attaining these events when compared to a group of his or her peers.[44]

The time of onset, direction, intensity and duration of the adolescent growth spurt seen in the craniofacial complex is of great importance to the orthodontist. The clinical importance of identifying and matching periods of accelerated growth to the timing of dentofacial orthopaedics has been highlighted in the treatment protocols of a wide variety of dento-skeletal disharmonies.[3, 4, 29] Treatment results may be influenced by changes in growth rate.[45] In orthodontic treatment planning, a knowledge of facial growth velocity and the percentage of remaining facial growth is very important for the effective prescription and use of growth modification interventions.[46] Initiating orthodontic treatment at the appropriate time may enhance treatment effects in the correction of skeletal discrepancies in all three planes of space; sagittal, transverse and vertical[47] and may result in a more favourable outcome for the patient.[28]

It has been advocated that the early intervention and utilisation of the differential growth potential, observed between the maxilla and mandible at puberty, helps to improve the prognosis of successful treatment of malocclusions with skeletal discrepancies.[1, 5, 48, 49] However, if the timing of the intervention is miscalculated then this may lead to an extended total treatment time and possible compliance issues. Treatment protocols for Class III malocclusions, aimed at enhancing maxillary growth,[50, 51] utilise the time period prior to the peak pubertal growth spurt while treatment aimed at restraining mandibular growth demonstrate greater effects when the pubertal growth spurt is included in the treatment interval.[28] In the treatment of Class II malocclusions, utilising functional appliance therapy, the objective is to induce supplementary lengthening of the mandible by stimulating increased growth at the condylar cartilage whilst restraining maxillary growth.[2] Success of this functional therapy strongly depends on the biological responsiveness of the condylar cartilage, which in turn depends on the growth rate of the mandible.[1, 5] Sagittal condylar growth, in patients treated with the Herbst appliance at the peak in pubertal growth, was twice that observed in patients treated 3 years before or 3 years after the peak.[52] Consideration of further dentofacial growth following completion of orthodontic treatment is important in the prediction of possible post treatment relapse[12] and aids determination of the most appropriate retention regime for the individual to maintain stability of the correction.[8] For certain patients, for example, patients with a moderate Class III malocclusion where camouflage or orthognathic surgery are being considered, or those patients planned for orthognathic surgery and/or implants, it may be best to delay orthodontic treatment until growth has been completed in order to achieve more stable results in the long-term.[6, 7]

These factors highlight the importance of accurately assessing the stage of development for the individual so that treatment effects can be maximised in the shortest time duration thus improving its efficiency while also promoting long-term stability.

2.2 Craniofacial development and growth

Development is a continuous process that begins with differentiation and ends with maturation.[45] Once differentiation is complete, the process of growth leads to an increase in size and mass of tissues and organs through cellular activities.[53] The craniofacial complex is formed both endochondrally and intramembranously; with the cranial base, nasal septum and mandible following an endochondral ossification process while the maxilla and cranial vault follow an intramembranous ossification process.[10]

Development and growth is a dynamic and complex process with each person expressing their own individual growth pattern.[10] An individual's typical growth curve is comprised of a series of differing rates of growth. Both body and facial growth rates follows a pattern of deceleration following birth, a small period of increased growth rate at 6-8 years of age, a pre-pubertal deceleration, a maximum growth acceleration at puberty, before decelerating rapidly until growth is complete in adulthood *(Figure 2.1).* While the sequence of growth is predictable, its precise timing is not.

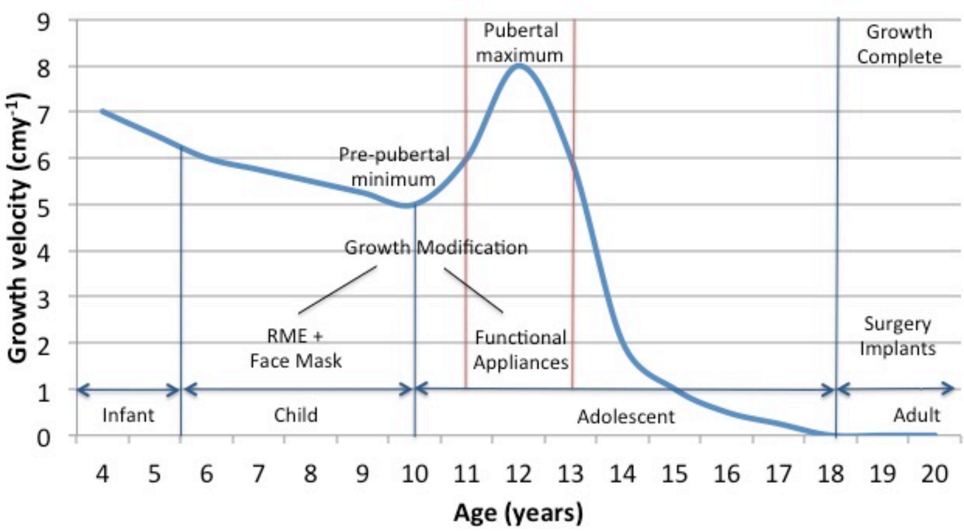


Figure 2.1: Growth velocity curve (growth per unit of time) for skeletal growth and timing of various orthodontic treatment modalities. Courtesy of Dr. Jayne Harrison, Orthodontic department, Liverpool University Dental Hospital.

Although much information exists with regard to craniofacial growth, many areas of our knowledge and understanding remain insufficient.[46] This uncertainty has resulted in many treatment modalities being proposed for the management of malocclusions with skeletal manifestations.[54] Over the decades, many theories pertaining to craniofacial growth have been proposed.[20-25] Brash proposed the *remodelling theory*, which emphasised the role of differential deposition and resorption of bone.[55] Weinmann and Sicher proposed an alternative hypothesis; the *sutural growth theory,* which postulated that the connective tissue and cartilaginous joints of the craniofacial skeleton are the principal locations where genetically regulated and primary growth of bone occurs.[55, 56] Scott presented the *nasal septum theory,* which concluded that the nasal septum is responsible for craniofacial skeletal growth.[55] He proposed that during the late pre-natal/early post-natal period to approximately four years of age, the anterior-inferior growth of the nasal septal cartilage moves the midface downward and forward. The cranial base synchondroses, which are comparable to epiphyseal growth plates, continue the process of craniofacial growth with the spheno-occiptal synchrondroses fusing at 14 years. The theory also proposed that the mandibular condylar cartilage behaves similarly to the cranial base and nasal septal cartilages, and directly determines the growth of the mandible downward and forward, increasing its total length and height.[57, 58] Moss proposed *the functional matrix theory*, where craniofacial growth occurs in response to physiological functional needs and in which genetics do not contribute significantly to the process.[59] Petrovic proposed the *servosystem theory* based on comprehensive in-vitro and in-vivo experiments.[60] This hypothesis postulated that the genetically regulated growth of primary cartilages determines growth and provides a constantly changing reference input which is mediated by dental occlusion causing the mandible to respond by adaptive muscular and condylar growth.[55]

2.3 Prediction of growth and treatment timing

2.3.1 Introduction

The major issue confronting the orthodontist is that each individual child grows at different times, in different directions and at different rates.[61] The typical craniofacial growth pattern, from birth to adulthood, follows an orderly, predictable and sequential pattern but is characterised by a wide individual variability in both the amount and rate of growth. This is innately related to genetic, gender and ethnic differences.[11] This wide variability can account for the observation that the same orthodontic treatment does not necessarily evoke the same response in all individuals.[56] The considerable variation in the growth rate and its timing is not wholly confined to the craniofacial complex and exists in all body systems. *Figure 2.2* represents postnatal systematic growth patterns and indicates that growth of the craniofacial complex, neurocranium and neural tissues are precocious in their growth and development, completing approximately 80% of total growth by 6-8 years of age. General body tissues show an S-shaped curve with a definite slowing of the growth rate during childhood and then acceleration at puberty. Growth of the mandible is not linear throughout its development[13] and both the maxilla and the mandible follow the somatic growth curve[62, 63] with approximately half of their total growth having been completed by age 8-10.[64] Therefore, the maxilla and the mandible have a considerable potential for further growth making it possible to have a significant treatment impact during the pubertal growth period.[45]

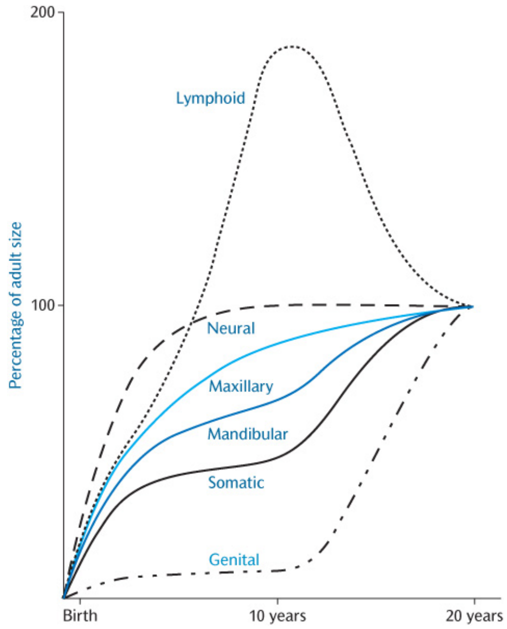


Figure 2.2: Scammon's curve of systemic growth. Growth of each structure is expressed as a percentage of the total gain between birth and 20 years. Size at 20 equals 100% of the vertical scale.[65]

Most of the data available regarding growth curves and trends have been taken from longitudinal studies of large groups of individuals. These have facilitated the generation of statistical data, such that, much of the clinical measurements and scientific reports from which they are generated, are based on averages *(Figure 2.1)*. Although these averages are useful and serve a guide to the clinician, predication of an individual's peak pubertal growth is more problematic. It is observed that there is as much chance for the individual patient to follow the mean pattern as it is to differ from it.[66] Obtaining an accurate method of assessing an individual's stage of growth and development, along with the prediction of remaining growth, has proven difficult. It would appear logical to place an individual on the growth curve based on their chronological age and from that assess the amount of remaining growth. However, given the wide individual variation in the timing of the pubertal growth spurt this renders chronological age a poor indicator.[12, 13, 67] Clinicians, therefore, need a method to determine the physiological development of the individual patient to reduce variability in treatment results.

The pubertal growth spurt has been defined as the peak in growth rate of standing height[14, 21, 68, 69] or mandibular growth.[3, 67] In order to determine the time of the peak growth period in facial dimensions for an individual, several maturity indicators have been investigated.[11, 70] The assessment of biologic age is confined to two main approaches; the clinical evaluation of various physiological indicators including longitudinal records of standing height, weight, dental development, development of secondary sex characteristics and the assessment of skeletal maturation based on radiographs[71] which was found to have the closest relationship.[11] The use of insulin-like growth factor I (IGF-I) to detect the pubertal growth spurt has also been described and correlated to cervical skeletal maturity, however, it has yet to achieve clinical acceptability and applicability in orthodontic clinical practice.[15, 72]

2.3.2 Chronological age

Traditionally, the time for commencing orthodontic treatment was determined by the patient's chronological age.[73] However, variability in the chronological age is seen when pubertal growth spurt occurs meaning that the onset of peak skeletal maturation or skeletal growth velocity in adolescences cannot be estimated accurately from chronological age.[12, 35, 70]

The relationship between chronologic age and skeletal age, with regard to craniofacial growth in humans, has been extensively investigated and the findings suggest that for the majority of subjects, chronologic and skeletal age do not coincide at the peak craniofacial growth period.[74] There is variation in the chronological age at which the onset adolescent growth spurt occurs for both genders ranging from 8-14 years suggesting that chronologic age has limited value in the appraisal of skeletal maturation.[47] It was demonstrated that on average, there was a four-year range in the onset of the adolescent growth period in males with a mean of 12.79 years, while the female peak growth varied by a five-year range in the onset with a mean of 10.41

years.[75] Due to this large inter-individual variation in the chronological age at which the peak pubertal growth occurs, it is recommended that a biological indicator be used when evaluating individual skeletal maturity.[70]

An individual with a skeletal age that varies one or more years from their chronological age may be categorised as having a delayed, average or advanced rate of maturation.[75] Significant differences have been noted between chronologic age and skeletal age of the spurt onset in advanced and delayed subjects but not in the normally maturing individuals[76] and this agreed with previous studies.[63, 75] In addition, it was found that facial growth did not cease in any male before the skeletal age of 18.[75]

The low correlation between chronologic age and skeletal maturity indicates that the initiation of orthodontic treatment, based solely on the evaluation of chronologic age, may coincide with periods of highly varying growth rates in both genders, because of the wide variation of timing of the pubertal growth spurt.[13] This also means that chronologic age cannot be accurately used to assess individual skeletal maturation.[3] However, recent publications refute this assumption and conclude that chronologic age is a better predictor of mandibular peak growth than skeletal maturation[71, 77-79] with vertebral and hand-wrist measures of skeletal age offering no significant advantage relative to chronological age in the identification of peak mandibular growth rate.[77] These findings, however, must be interpreted with caution because the samples used were historical and the onset of puberty is affected by several factors including ethnicity, genetics, nutrition and socioeconomic status.[80]

Nevertheless, the majority of evidence would suggest that while chronologic age may serve as a guide to treatment planning, it is not a valid indictor of skeletal maturation due to significant individual variations. As a consequence, using chronological age as a tool for evaluating skeletal growth and maturation, with regards to orthodontic treatment planning and commencement of orthodontic treatment, is of little clinical value. This has prompted investigation into the evaluation of the biologic age of an individual as a more accurate measure of skeletal maturity. The biologic age is determined by assessing the degree of maturation of different tissue systems that can be estimated by considering somatic, dental, sexual and skeletal maturation.[28]

2.4 Physiological age indicators

2.4.1 Somatic growth and height

When height measurements are plotted against chronological age, a growth chart is produced which shows a steady increasing curve that continues to increase from birth to adulthood until maximum height is achieved. While the growth of the individual can be followed longitudinally it does not indicate the velocity of growth or when the peak in statural height may occur. By plotting the height change against age, the rate of growth can be assessed and the peak growth spurt identified by the sharpest point on the curve *(Figure 2.3).*[81, 82] The growth rate is highest in the first year post-partum, followed by a rapid deceleration until 5-6 years of age. After this stage, a slowly decelerating phase exists which may be briefly interpreted by the juvenile growth spurt around 6-8 years. An accelerated phase of increased growth occurs between the ages of 10-16 years. This is associated with puberty and is often referred to as the 'circumpubertal growth spurt', which, may extend over a three-year period. In the first two years, an increase to the peak height velocity is observed and in the third year, it decreases to the level just above that found before puberty.[82] On average, the pubertal spurt begins at 10 years of age in females and 12.1 years of age in males and ends at 14.8 years in females and 17.1 years in males. The peak height velocity (PHV) occurs on average two years after the onset of this spurt (12 years in females and 14.1 years in males) with growth terminating at 17.5 years in females and 19.2 years in males.[83] A recent study, utilising longitudinal data from the Broadbent-Bolton growth study, corroborate the ages of onset and peak in statural growth for boys, however, the girls were accelerated by 9 months with statural onset at 9.3 years and statural peak at 10.9 years.[77] It was speculated that these differences may be due to secular variations among the different population samples used and the different times the data were collected.

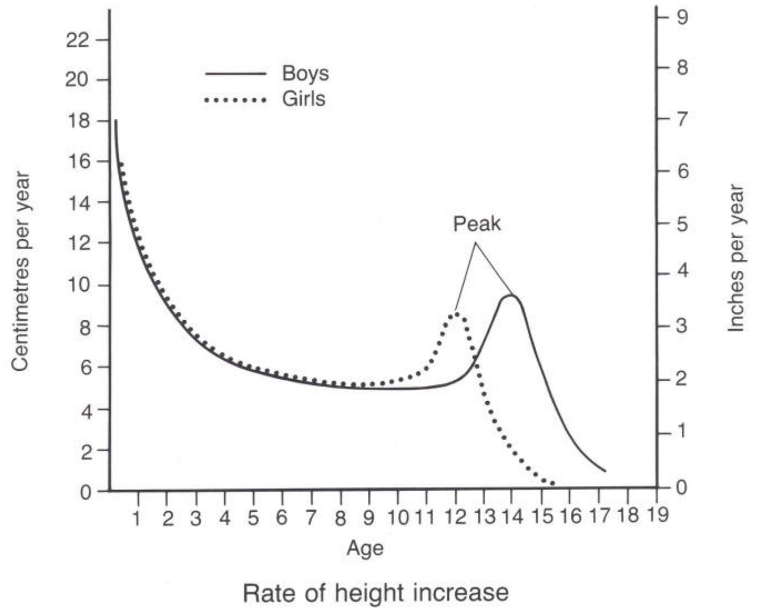


Figure 2.3: Average height velocity growth curve throughout stages of life from infancy to adulthood for average Male and Female Children[84]

Sullivan reported a method of predicting the individual's onset of pubertal growth spurt by the measuring standing height.[82] While this remains a common method for assessing skeletal maturity and growth potential it requires the standardised measurement of standing height, using a stadiometer, over a period of time prior to commencement of orthodontic treatment. After the second reading and every reading thereafter, height measurements are translated into height velocity values and plotted on a growth chart in order to determine whether the individual is in the accelerative phase of the pubertal growth spurt. Using the individual measurements of boys and girls from the Harpenden growth study, Tanner fitted logistic curves over the height velocity charts facilitating the clinician to determine which growth chart the patient was following and from which predictions could be made.[81, 85] This method is no longer used as the Harpenden sample is historic and the data no longer applicable to the current population. The validity of this method is further diminished as the girls in the sample were in state care and many had emotional problems, which may have affected the pubertal onset and growth. The growth charts currently used have a supplemental guide for assessing pubertal stage using Tanner's stage of puberty, however, they require a patient interview regarding the development of secondary sex characteristics. This may be an inappropriate line of questioning for many patients seeking orthodontic treatment and one that the clinician, parent and patient would find uncomfortable in the orthodontic setting.

Facial growth is closely related to somatic growth and accelerates greatly during adolescence.[86] The facial growth spurt which coincides with the skeletal growth spurt, is of a shorter duration and occurs earlier in females than in males.[11] A large body of evidence has shown a positive relationship between the peak facial growth and the peak height velocity, (PHV), although a disagreement on the sequence of these two events remains.[62, 63, 75, 76, 87] Several studies have found that the peak facial growth coincides with the peak growth rate in statural height[75, 76, 87] without a gender difference.[88] In contrast, other research has reported a difference in the timing between peak height velocity and peak facial growth rate, suggesting that the maximum facial growth lags behind PHV by variable amounts of time.[62, 63, 89] Moore *et al*. [90] found that the peak velocities of mandibular length, anterior facial height and total posterior face height in males occurred after the peak velocities for standing height while the velocities for standing height, mandibular length and anterior facial height in girls were coincident. The correlation between changes in facial dimensions and statural growth was investigated and revealed that facial dimensions, with the exception of total posterior face height in girls aged 10-11 years, were only weakly correlated.[90] A recent study demonstrated, that in both sexes, the peak in stature had a shorter duration and tended to occur a few months before that of the face and mandible.[77] A wide variation exists in the timing of peak facial growth and much of the conflicting evidence may be the result of the different methods used to measure mandibular length and small sample sizes. In fact, information concerning changes in standing height and facial dimensions are more important than the actual lengths of the distances themselves.[91] The growth velocity of standing height is a good indicator of total posterior facial height and to a lesser extent anterior facial height.[91] A longitudinal study investigating craniofacial growth observed that peak mandibular growth coincided with peak statural height in 93.5% of individuals.[3] However, Mitani and Sato[61] studied the lateral cephalograms, hand-wrist radiographs and height records of a sample of thirty-three Japanese girls in order to compare mandibular growth during puberty with the growth characteristics of the hyoid bone, cervical vertebrae, hand-wrist radiographs and standing height and concluded that there was large variability in the timing and magnitude of the peak pubertal growth spurt between all the variables, with mandibular growth showing the most unpredictable variability. The diagnostic reliability of the peak standing height in the identification of the mandibular growth peak has been reported with a variable diagnostic accuracy (0.61-0.95).[68] Each individual has their own pattern of growth velocity that is specific and unique and displays a nonconformity to the pattern exhibited by the general population.[74]

Peak growth velocity in standing height is a valid representation of the overall rate of skeletal growth however, it offers little predictive value of future growth or percentage of total growth remaining.[12] Longitudinal standardised height assessments are required to identify accurately the onset of the accelerated phase of growth but it may only be possible to detect the maximum growth event when the velocity graph takes a downward turn. From a clinical perspective, the recording of standing height may be useful tool when used in conjunction with other radiographical indicators, particularly as it non-invasive. While this type of data would greatly benefit the orthodontist prior to considering a surgical approach in the treatment of a skeletal discrepancy by identifying the end of the growth period, it would be of limited value in those cases where active growth is beneficial to achieve the optimum treatment aims.

2.4.2 Dental development

While many researchers have attempted to establish whether there is a relationship between skeletal maturation and the maturity of the permanent dentition, the evidence remains inconclusive.[92] Teeth vary in their development[93] and dental maturity can be determined by either investigating the stage of tooth eruption or the stage of tooth calcification.[94, 95] Dental eruption is more readily influenced by various local[96] and environmental factors than the calcification process.[97] Therefore, tooth formation is the more reliable criterion[98] facilitating the assessment of multiple stages in the tooth's development covering a wider range of maturity and is expected to produce higher associations than the age of eruption. There has been a considerable amount of research conducted to investigate dental development as a potential predictor of skeletal maturation.[47, 73, 92, 95, 97, 99-101] Those that advocate dental development indictors, highlight the advantage of identifying various dental development stages, through the use of a routine panoramic or intra-oral periapical radiographs. This offers a practical method for the assessment of physiologic maturity without resorting to further radiographic exposures,[97] thus allowing the clinician to practice the 'As Low As Reasonably Achievable' (ALARA) principle.[98] It has been reported by some authors that there is a low correlation between dental and skeletal maturation[92, 100] while other research reports a high correlation.[73, 95, 97-99] These inconsistencies may be due to the different methods used to assess skeletal and dental maturity.[95] No correlation has been shown between stage of dental development and mandibular growth[47] and only weak correlations with statural height have been demonstrated.[13, 94]

A low correlation between dental emergence and pubertal growth indicates that dental development is of little value as a predictor of pubertal growth.[13, 83] Dental development was found to have a limited value in predicting puberty as eruption to the occlusal level could occur several years before or after the maximum pubertal skeletal growth.[94] Males have a more advanced dental development than girls[13, 94, 95] while elsewhere it was reported that females have an earlier tooth eruption than males with a evident relationship between maturational status and tooth formation as adolescence approaches.[92] In addition to ethnic, environmental and gender differences, dental anomalies may also affect dental maturity. Patients with hypodontia were found to have a mean delay in dental development of 1.51 ± 1.37 years.[102]

Several studies have related dental maturity to skeletal maturation for specific teeth with a high correlation found between the mandibular canine and skeletal maturity indictors [73, 96, 97] while others identified the second molar as showing the highest correlation.[95, 98, 99] It was suggested that the second molar tooth offers an advantage over other teeth as its development tends to continue over a longer period and until a later age,[98] with apex closure generally extending up to the age of 16 years in the average child.[99] Sierra[73] highlighted the low degree of correlation between dental and skeletal maturity due to research methods using ossification centres that exhibited a wide variation in their onset. By evaluating the correlations between the developmental stages of ossification centres that showed the least variability in their onset and the calcification of the permanent mandibular canine, it showed a significant correlation between skeletal maturity and maturity of the dentition. The eruption of the maxillary permanent canine showed great inter-individual variability with regard to the age of emergence when compared with skeletal maturity.[100] It must be noted that studies that have examined the mandibular canines[97] in relation to dental maturity may exhibit some limitations as both the root formation and the apex closure of mandibular canines are usually complete by age thirteen however the majority of children exhibit active growth up to the age of 16-17 years. Assessment of dental maturity using third molars[103] is flawed as these teeth are the most common missing teeth in the permanent dentition, making them unreliable for age assessment.[98] Other studies have investigated the skeletal maturation and the different phases of the dentition; the early mixed, the intermediate mixed, the late mixed and the early permanent phases. Both Perinetti et al.[101] and Franchi et al.[47] found that the early mixed dentition was a useful indicator for the pre-pubertal stage of skeletal maturity but the late mixed and the early permanent dentition were of little value in providing precise identification of the pubertal growth spurt. Identification of this pre-pubertal stage may be beneficial for the commencement of growth modification in Class III malocclusions.[48] On review of the literature, when compared to a physiological indicator of skeletal maturity, dental development stage performs poorly in the evaluation of the onset of the pubertal growth spurt.[104] Skeletal maturation becomes a more important predictor for future mandibular growth. However, evaluation of dental development may be used as an adjunct for the assessment of skeletal growth.[104]

2.4.3 Secondary sex characteristics

The timing of the pubertal onset is highly variable due to gender, heredity, ethnicity, environmental influences and secular trends.[14, 80, 94] Sexual maturation and secondary sex characteristics correlate well with other biologic indictors such as skeletal age and statural height.[14] The development of secondary sexual characteristics includes the appearance of pubic hair, facial hair, voice changes, testes and penile growth, thelarche (the onset of female breast development) and menarche. While the appearance of secondary sexual characteristics are useful assessments of growth and development, a physical examination of the patient is not appropriate in the orthodontic setting and for this reason the majority of investigators have limited the assessment of pubertal development to menarche in girls and voice changes in boys.[14, 83, 94]

Menarche is a relatively late marker of female puberty. It most often occurs after the maximum increment of mandibular growth and while mandibular growth may continue after menarche the average amount of mandibular growth is greater before menarche than after it.[14] In late maturing girls, menarche may occur before peak mandibular growth.[14] A close association was found between the age of PHV and the age of menarche.[81, 94] The appearance of first menses is indicative of the PHV having been attained 1.1 years previously and a deceleration of growth.[105] However, while menarche is a highly reliable it is not an absolute indicator that the PHV has been reached.[83] This is refuted by Bjork and Helm,[94] who suggest that a history of menarche is a reliable indication that the pubertal growth spurt has been achieved and therefore is of little value in identifying the PHV for girls. Therefore, menarche may indicate to the orthodontist that the opportunity to treat skeletal discrepancies with particular treatment modalities has been missed. The onset of thelarche is reported to be as variable as chronologic age and is therefore not suitable to identify the stage of maturity.[81]

For boys, voice changes from the pre-pubertal to male voice occurs throughout the period of the circumpubertal growth spurt. Attainment of the male voice is a reliable indicator that PHV has been attained or surpassed[13] and growth is decelerating.[105] Whilst it may be a beneficial indicator that no significant further growth is likely to occur, it has limited use in predicting peak growth. Genital development precedes the peak growth spurt and is a reliable indicator for predicting the impending growth spurt.[81] The development of pubic hair is closely related to PHV but does not correlate to the onset of the pubertal spurt and is therefore of limited value.[81] While the development of secondary sexual characteristics is a useful clinical tool, it is restricted to the adolescent period.

Maturation indicators, that require a physical examination of secondary sexual characteristics are not applicable in clinical orthodontics and, the questioning of patients on the appearance of sexual maturity characteristics, particularly during this sensitive stage of their development, may cause embarrassment. Indeed, the parent and the patient may not appreciate the significance or the appropriateness of such as invasive line of questioning. For these reasons, such information is not routinely obtained during the orthodontic patient history. As such, it may be deemed inappropriate for an orthodontist to enquire about features of sexual development in an attempt to predict growth.

2.4.4 Biomarkers

Biomarkers have recently been proposed as a new method of assessing individual skeletal maturity. These have the advantage of avoiding radiation exposure and being related to the patient's own physiology. The usefulness of biomarkers, such as insulin-like growth factor I (IGF-I)[15], alkaline phosphatase (ALP)[19, 20] and total protein content from gingival crevicular fluid (GCF),[106] to predict skeletal maturity have been investigated. These biomarkers, with the exception of GCF total protein content, reported increased levels during the pubertal growth spurt when compared to the pre- and post-pubertal levels.[15, 19] Recent studies have shown that IGF-I can be used to predict the annual growth rate of the mandible and total anterior facial height when used in conjunction with cervical stage, skeletal classification and gender.[16, 17] The invasive nature of collecting blood spots for the assessment of IGF-I makes markers from GCF of particular interest. However, optimal gingival conditions, with the absence of plaque accumulations, is necessary as levels of GCF ALP activity can be altered by local tissue inflammation.[18]

2.5 Skeletal Age Maturation

Biologic age, skeletal age, bone age and skeletal maturation are all synonymous terms that can be used interchangeably to describe the stages of maturation of an individual. The only indicator of development available from birth to maturity is skeletal age.[107] Skeletal maturation refers to the degree of ossification in a bone.[29] During growth, every bone undergoes a sequence of changes that are relatively consistent for that bone in every individual and these changes can be seen radiographically. Maturation change of the bone begins as a primary centre of ossification and ends with fusion of the epiphyses to the main body of the bone. The underlying principle is that osseous changes seen in the bones indicate more general skeletal changes[21] and this is a widely used approach to predict timing of pubertal growth, to estimate growth velocity and to estimate the amount of growth remaining.[12] Skeletal size and maturation can vary independently from each other. Most areas of the skeleton have been examined to evaluate maturity including the hand-wrist, knee, shoulder, elbow, foot and the iliac bone.[26] The two methods for the assessment of skeletal age for the prediction of growth are the assessment of the bones in the hand and wrist from a hand-wrist radiograph and the cervical vertebrae from a lateral cephalogram. These radiographs may be used to indicate an ossification event or a discrete stage of skeletal maturation. When using radiographical indicators of the growth phase that are based on sequential discrete stages, it is important to distinguish between stages and ossification events. Stages are specific periods in the development of a bone described in that particular staging method, while an ossification event occurs when a given stage matures into the next.[69, 108] As such, because an ossification event occurs at a midpoint between two consecutive stages, its identification requires serial radiographs and this limitation has meant that this approach has lost favour clinically. The staging method used in both hand-wrist and cervical vertebral staging indices is disadvantaged by the fact that a single stage may have various durations[21, 109] making the prediction of the onset of the peak pubertal period less reliable.

2.5.1 Hand-wrist radiographs

Historically, the standard method of evaluating skeletal maturity has been to use radiographs of the hand-wrist area as it was preferred to other anatomical sites due to the ease of positioning and the quantity of bones present.[12, 107] It facilitated the comparison of bones to various published atlases or skeletal maturation indices. Various techniques for the assessment of skeletal age have been developed since the 1930s with Todd (citated Cox)[107] and then Greulich and Pyle, 1959[110] producing atlases. Later, in the 1980s, the Tanner-Whitehouse method[111] and the Fishman Method[21] of Prediction were developed.

2.5.1.1 The Greulich and Pyle method

In 1959, Greulich and Pyle published the "Radiographic atlas of the hand and wrist", an atlas of skeletal maturity indicators of the hand and wrist.[110] This methodology included a series of images of the left hand-wrist covering the range of variability around different chronological ages. The median of the range was referred to as the 'reference standard'. Two separate standards were developed to allow maturity estimations for males from 0-19 years and females from 0-18 years. A comparison of the development of hand-wrist bones of an individual with the standard reference facilitated assignment of a corresponding skeletal age. The reference standards were derived from a higher socio-economic Caucasian, North American population during the 1930s and 1940s and therefore this historical sample may not be applicable to more present day, ethnically diverse, populations.[107] It is believed that this method remains the most frequently used method to assess skeletal age worldwide.[107] A modified version of this technique may be used, where the overall appearance of a given radiograph is compared with the reference radiographs and the nearest match selected. While this approach is considerably faster, it may be less accurate.[112]

2.5.1.2 The Tanner-Whitehouse method

The Tanner-Whitehouse method[111] was developed from 3000 hand wrist radiographs, (HWR), of healthy children from the Harpenden Growth Study, a longitudinal study which began in the 1940s, of boys and girls living in a children's home outside London, UK. This bone-by-bone system relies on the systematic evaluation of the maturity of 20 bones in the hand and wrist. Each of the 20 bones are scored from 1 to 8 and then added together. The final tallied score, called the 'maturity score', can range from 0 (totally immature) to 1000 (mature) and is then converted to a skeletal age using the respective tables for males and females. The advantage of this method, over the atlas system, is that only one series of values are required because differences between the sexes are accounted for in the scoring system. Also, the maturity score is weighted therefore it facilitates statistical analysis.[107] It must be remembered that the Tanner-Whitehouse method was derived from children from socially deprived conditions, which may have a bearing on the findings, and also this method attempts to assign discrete values to maturation even though this is a continuous process.

A small study compared the two methods[113] and suggested that there was close agreement between them. However, the data was analysed inappropriately using regression analysis, and a subsequent study found that both methods produced different values for bone age, which would be significant in clinical practice.[112]

Both the Greulich and Pyle method and the Tanner-Whitehouse method have been criticised in the literature as it is difficult to set reference standards due to the differential rate of maturation in the different bones in individuals of the same population or in different populations.[12, 69] As a result, several standards for different populations have been published. However, reference standards are not available for each population or ethnicity and changing secular trends equates to these standards becoming outdated. Therefore, staging of skeletal maturity by describing specific ossification events on a HWR may be more valid as it is independent of differences among populations, secular trends and availability of references standards.[12, 114]

2.5.1.3 The Fishman method

Fishman developed a clinically based method of evaluating skeletal maturation based on hand-wrist films.[21] The Fishman Method of Prediction (FMP) was based on longitudinal and cross-sectional data. The longitudinal data, derived from the Denver Child Research Council, were utilised to gather information on absolute growth and growth velocities while the cross-sectional data, derived from patients attending Fishman's private practice and the Eastman Dental Centre, reduced secular error in the chronology of onset and progression of growth during adolescence. The system of Skeletal Maturation Assessment (SMA) uses four stages of bone maturation which are found at six anatomical sites located on the thumb, third finger, fifth finger and radius (*Figure 2.4*).[21]

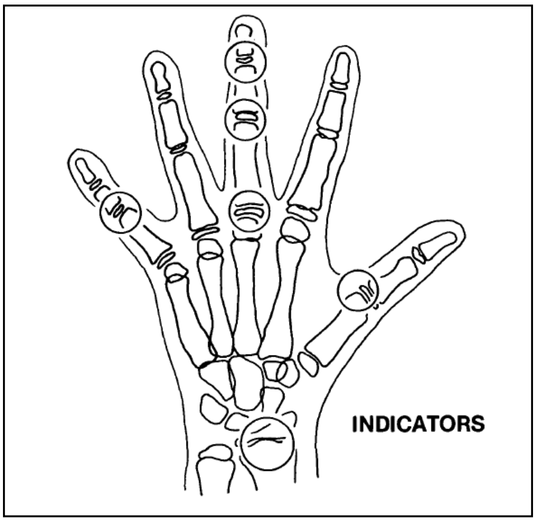


Figure 2.4: Sites of skeletal maturity indicators[21]

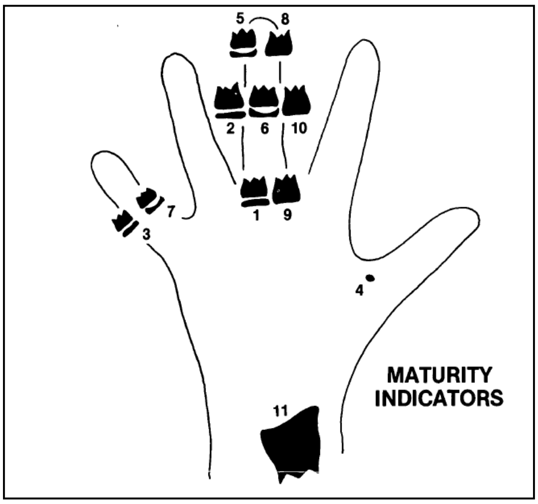


Figure 2.5: Eleven skeletal maturity indicators[21]

Eleven discrete skeletal maturation indicators (SMI) are found at these six sites and coincided with the entire period of adolescent development (*Figure 2.5*).[21]

The sequence of the four ossification stages progress from epiphyseal widening, to ossification of the adductor sesamoid of the thumb, to 'capping' of the selected epiphysis over their diaphyses and ending with fusion of selected epiphyses and diaphyses *(Figure 2.6*).[21]

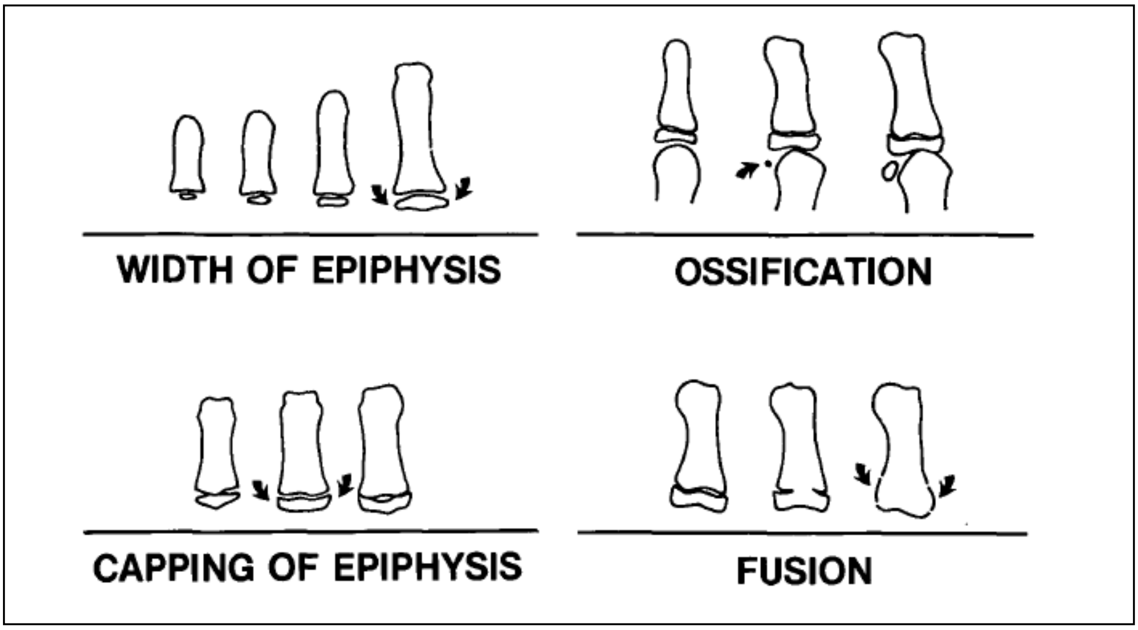


Figure 2.6: Radiographic identification of skeletal maturity indicators, A-Epiphysis equal in width to diaphysis, B-Appearance of adductor sesamoid of the thumb, C-Capping of epiphysis, D-Fusion of epiphysis.[21]

This eleven-stage sequence of maturation was found to be stable with only three deviations detected from over two thousand observations. These 11 stages are divided into four categories including ephiphyseal widening (SMI 1-3), ossification of the sesamoid of the thumb (SMI 4), capping of the third and fifth finger epiphyses over their diaphyses (SMI 5–7), and finally fusion the third finger epiphyses and diaphysis and radius (SMI 8–11) (*Table 2.1*).[21]

A	**Width of epiphysis as wide as diaphysis**	1. Third finger-proximal phalanx 2. Third finger-middle phalanx 3. Fifth Finger-middle phalanx
B	**Ossification**	4. Adductor sesamoid of thumb
C	**Capping of epiphysis**	5. Third finger-distal phalanx 6. Third finger-middle phalanx 7. Fifth Finger-middle phalanx
D	**Fusion of epiphysis and diaphysis**	8. Third finger-distal phalanx 9. Third finger-middle phalanx 10. Fifth Finger-middle phalanx 11. Radius

Table 2.1: Individual SMI in chronological order.[21]

- SMI 1 and SMI 2 occur before the standing height and mandibular growth peak i.e. pre-pubertal period.
- SMI 5 and SMI 6 generally coincide with standing height and mandibular growth peaks i.e. pubertal spurt.
- SMI 8 and SMI 9 occur after the standing height and mandibular growth peaks i.e. post-pubertal.

Once the SMI was determined, the analysis facilitated an estimate of relative growth velocity, completed growth[115] and percentage of growth remaining. Such an approach is both valid and reliable and is more useful method than an analysis which provides a skeletal age, as it minimises environmental influences, racial and gender differences.[12] The ossification of the adductor sesamoid was found to be highly correlated with peak height velocity and the start of the adolescent growth spurt.[94] A study consisting of two surveys; a cross sectional and longitudinal study, was carried out to relate the ossification of the sesamoid bone to the increases in statural height of adolescents and concluded that the maximum velocity in height occurs after the commencement of sesamoid ossification.[116] In their studies, evaluating skeletal stages of the hand and wrist as indicators of the pubertal growth spurt, Hagg and

Taranger[13, 117] determined that ossification of the sesamoid indicated the peak and the end of the pubertal spurt but not the beginning. There is significant correlation between skeletal maturity and craniofacial growth[12] with ossification of the sesamoid preceding the peak mandibular growth velocity.[118] In a systematic review conducted to evaluate the validity of skeletal maturation to predict facial growth, it was concluded that overall horizontal and vertical facial growth velocities are related to the SMI determined by hand-wrist radiographs but individual mandibular and maxillary growth velocities were less robustly correlated for overall facial velocity.[12]

While skeletal age assessment utilising the hand-wrist anatomical area has been found to be reliable and valid, its routine use in clinical orthodontics is no longer as frequently carried out due to a number of factors. There are a number of limitations regarding the use of the HWR for the prediction of skeletal maturity requiring considerable skill and anatomical knowledge of the hand-wrist area. The ossification sequence and timing of skeletal maturity within this anatomical area shows polymorphism and sexual dimorphism which may limit its predictive use.[69] Ossification events in the hand and wrist indicate the peak and the end of the pubertal growth spurt, however it does not indicate the onset of the pubertal growth spurt.[77] Moreover, it exposes the patient to additional radiation exposure. This has important implications for patient and staff safety as well as costs. As clinicians we have a responsibility to keep radiation exposures to minimum levels and achieve diagnostic information while keeping levels of radiation 'As Low As Reasonably Achievable' (ALARA) to minimise possible adverse health risks.[24] Current radiological policies and practices are based on the assumption that a degree of adverse risk does exist, and this risk must be clearly outweighed by benefits.[119]

2.5.2 Cervical vertebrae

The need for serial hand-wrist radiographs to estimate the PHV preclude their routine use in most clinical situations[69] due to ethical issues.[34] The attention of orthodontists has been directed to the cervical vertebrae visible on a lateral cephalogram in order to estimate a patient's maturation and growth potential. This anatomical area is ideal for orthodontists because lateral cephalograms are routinely taken as part of pre-orthodontic treatment records as it is a diagnostic tool from which the skeletal morphology and directional growth patterns can be assessed. Therefore the assessment of skeletal age and the remaining growth potential can be evaluated without exposing the patient to additional radiation.

2.5.2.1 Morphology of the cervical vertebrae

The cervical vertebrae are comprised of the first seven vertebrae in the spinal column and they follow an anterior convexity. The function of the cervical vertebrae, together with their intervertebral discs, is to support the head. The morphology of the first cervical (CV1) and second vertebrae (CV2) are atypical in their morphology and different from the remaining five vertebrae (CV3-CV7). CV1, called the atlas, has a ring shape but has no body or spinous process and its function is to support the skull by articulating with the occiput (the atlanto-occipital joint) and inferiorly with the axis (the atlanto-axial joint). The atlanto-axial joint is responsible for about half of all cervical rotation while the atlanto-occipital joint is responsible for half of the flexion and extension.

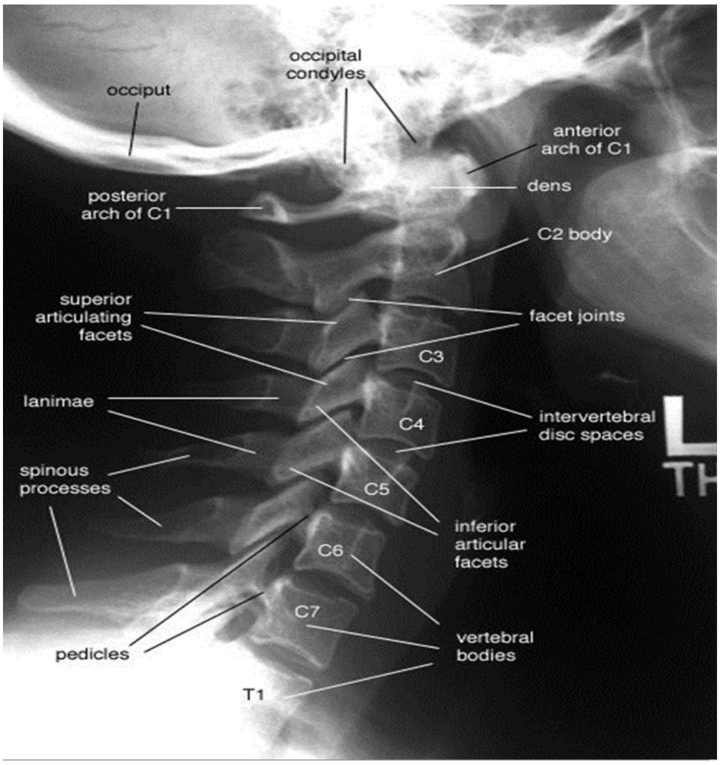


Figure 2.7: Radiograph depicting the cervical vertebrae anatomy[120]

Fused remnants of the atlas body have become part of CV2, where they are called the odontoid process, or dens. The odontoid process is held in tight proximity to the posterior aspect of the anterior arch of the atlas by the transverse ligament, which stabilises the atlanto-axial joint. The apical, alar, and transverse ligaments, allow spinal column rotation, and provide further stabilisation. The atlas is made up of a thick anterior arch, a thin posterior arch, two prominent lateral masses, and two transverse processes. The transverse foramen, through which the vertebral artery passes, is enclosed by the transverse process. The axis is composed of a vertebral body, heavy pedicles, laminae, and transverse processes, which serve as attachment points for muscles. The axis articulates with the atlas via its superior articular facets, which are convex and face upward and outward (*Figure* 2.7).[121] Growth of the cervical vertebrae can be measured from the lateral cephalogram and from the age of 2 years, the morphology of the first, second, and third cervical vertebrae is established.[122]

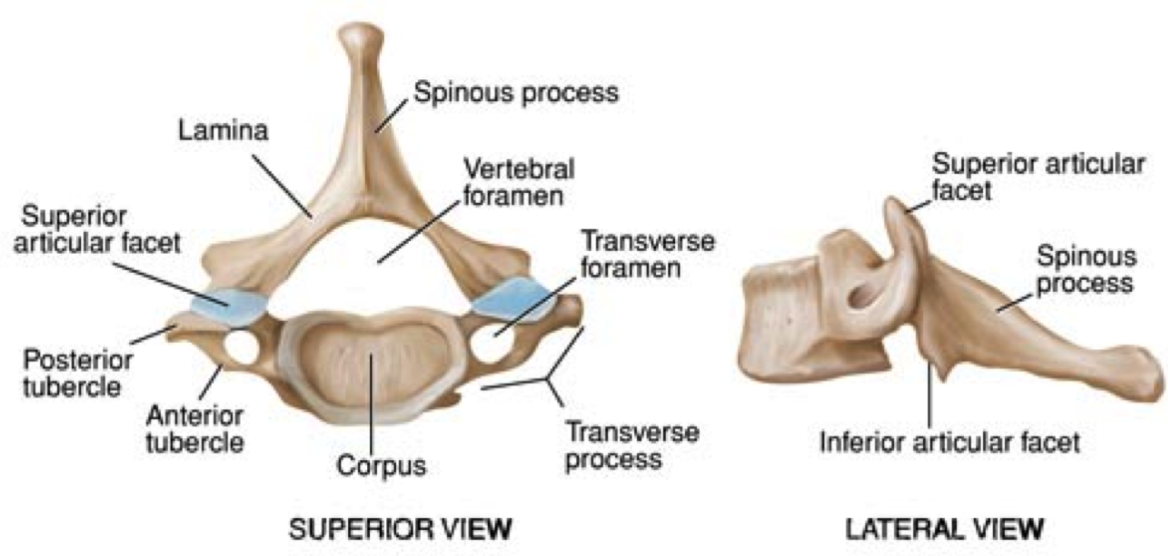


Figure 2.8: Morphology of the cervical vertebrae[123]

The typical cervical vertebrae (CV3-CV6) have superior articular processes that face upwards and backwards, an uncus at each side of the upper surface of the body, a triangular vertebral foramen and a bifid spinous process (*Figure* 2.8). The spinous process of CV7 is usually non-bifid and bulbous at its end.

The vertebral column originates from mesenchyme[124] and its development differs between individual vertebrae. The cervical vertebrae ossify from three primary centres, one in the body and one in each lateral vertebral arch[122] at week 7-8 inauterine.[124] At birth, the typical vertebra consists of three bony parts, united by hyaline cartilage and begin to fuse during the first year of life. Fusion of the body with the vertebral arches occurs between 5 to 8 years. During puberty, five secondary ossification centres develop: one in the spinous process; one in the tip of each transverse process; and two called anular epiphyses, one in the superior and one in the inferior edge of the centrum. The atlas usually ossifies from three centres, one in each lateral mass and one in the anterior arch that fuse between 3-4 years of age. The axis is ossified from five primary centres and two secondary centres; two for the arch, which fuse posteriorly between 3-4 years, one for the body and another two for the lower two thirds of the odontoid process that fuse between 3-7 years; one for the apex of the odontoid process that appears between 2-6 years and then fuses with the main mass of the dens at 11-12 years and another one for the epiphyseal plate on the under surface of the body of the bone.[124] After endochrondral ossification is complete periosteal apposition and remodelling continues until the final vertebral form is established in adulthood.[29] Differing rates of growth for the anterior, posterior and middle portions of the vertebrae result in age related size and shape changes that are consistent throughout the growth period.[26] During the adolescent growth spurt, the body of the vertebrae grow in a vertical direction on the anterior and posterior lower body of the body. Maturational changes are evident and predictable from birth to puberty,[125] with these morphological changes evident from radiographic examination. A set of standards for cervical vertebral maturation were created and correlated to the hand-wrist radiograph and were found to be as statistically and clinically reliable in assessing skeletal age as the hand-wrist technique.[26] Later it was demonstrated that both the length and height of the cervical vertebrae increased with age and was significantly correlated with statural height, making the developing cervical vertebrae suitable to assess biological maturity.[126] Visual examination and tracing of the lateral cephalogram are the best method of assessing and recording the morphology of the cervical vertebrae[26] however, superimposition of cervical structures may limit the interpretation of vertebral morphologic changes.[127]

2.5.2.2 Cervical Vertebral Maturation

Many studies have investigated the relationship between hand-wrist radiographs and cervical vertebral maturation[22, 23, 29, 31, 32, 34, 35, 37-39, 128-130] with a high correlation found between the two methods and with identifying the peak pubertal growth. CVM therefore has the potential to provide information for the optimum timing of growth modification and for when growth has ceased. It also has the benefit of reducing the radiation exposure to the patient. In fact, given the small discriminatory ability between these two methods in the detection of peak pubertal growth, subjecting a child/adolescent to an additional radiograph may appear unethical.[23] It has been suggested that cervical vertebral maturation (CVM) should replace the skeletal maturation evaluation from hand-wrist radiographs for the evaluation of skeletal maturity.[34] However, a recent study evaluated the effective doses of a lateral cephalogram radiograph with and without thyroid shield and compared the differences with the radiation dose of a hand-wrist radiograph and concluded the effective dose of a standard HWR and a lateral cephalogram with a thyroid shield was lower than that of a lateral cephalogram without the thyroid shield. They suggested that given the sensitivity of the thyroid to radiation exposure, the beam should either be collimated to exclude the thyroid or a thyroid shield used.[119] This would however, forgo the possibility of determining the skeletal age using the CVM method making a hand-wrist radiograph necessary in cases were skeletal maturation requires assessment. Low correlations between chronological age and both the CVM method and the HWR confirmed that chronological age was not suitable to measure skeletal maturity[35] however, recent evidence suggests that the HWR method has a lower error than mean chronological age or the CVM method in the identification of both peak growth in statural height and mandibular growth.[77] These research findings, when combined with the expressed concerns regarding the vulnerability of the thyroid tissue to radiation has led to a debate and re-evaluation of the suitability of CVM staging method.[131]

2.5.3 Development of the Cervical Vertebral Maturation indices

2.5.3.1 Lamparski 1972

The aim of Lamparski's research was to determine whether the maturation changes of the cervical vertebrae, as seen on a standard lateral cephalometric radiograph, could be used to evaluate the skeletal age of a patient. By developing the work of previous investigators he described changes in the size and shape of the cervical vertebrae and created a set of maturational standards.[26] Lateral cephalograms of 72

Caucasian females and 69 Caucasian males aged between 10-15 years were selected from the records of the Michigan Growth Study and used to determine the standards. This age range was selected as it represented the age of the majority of orthodontic patients and is the period during which most maturational changes occur. The standards selected were based on a group of lateral cephalometric images of individuals whose chronologic and skeletal ages were within 6 months from the age under study and were arranged in a sequence of maturity as based on vertebral development and separated for each gender. A series of six standards, corresponding to maturational stages for each gender, were created for each age from 10-15 years. No statistical significant differences were found between the maturation stage determined from the hand-wrist film and that determined from the cervical vertebrae and it was concluded that the cervical vertebrae could be used to predict skeletal age.[26] The maturity indicators that were considered to be valid were the development of concavities in the lower border of the vertebrae and an increased vertical height of the vertebral bodies. These maturity markers were the same for both males and females but females matured earlier than males. Ossification of the vertebral ring, the sagittal diameter of the vertebral canal and the overall vertebral development were not found to be usable features to describe maturity.[26]

In summary, Lamparski developed a maturational standard from studying the morphological changes of CV2-CV6 from which he identified six stages of maturation in children aged between 10-15 years. This age limit restricts the application of the maturational standards as it is well documented that circumpubertal growth continues beyond this time period, particularly in males.[77] In addition, this classification system is not sufficiently accurate in males and can only be used for females.[32] Some years later, Lamparski's cervical vertebral maturation (CVM) method was correlated to changes in mandibular length.[67] Annual lateral cephalograms of 13 Caucasian girls from 9-15 years of age were evaluated and a statistically significant increase in mandibular length, corpus length, and ramus height in association with specific maturation stages in the cervical vertebrae were found. It was concluded that CVM stages 1 to 3 occurred before peak velocity in mandibular growth and stages 4 to 6 occurred after the peak mandibular growth velocity. A subsequent study, using a modified Lamparski's classification method, found that the peak mandibular growth velocity occurred between Lamparski's stage 3 and 4 and concluded that it was valid for the evaluation of skeletal maturity and identification of the pubertal peak in craniofacial growth making it a very useful diagnostic tool in orthodontics.[3]

2.5.3.2 Hassel and Farman 1995[29]

Hassel and Farman[29] further expanded Lamparski's work by focusing on three entities; the presence or absence of curvature in the inferior borders of CV2, CV3 and CV4, the shape of the vertebral bodies and visualisation of the intervertebral spacing. These areas were selected as they could be readily visualised from the lateral cephalogram when a protective thyroid collar was worn. Lateral cephalogram radiographs and hand-wrist radiographs from a sample of 220 subjects, aged between the 8 to 18 years from the Bolton-Brush Growth Centre were evaluated. The study combined the observations of the changes in the hand-wrist and the changes in the cervical vertebrae during skeletal maturation. From the key entities of interest, six categories of cervical vertebral maturation were defined (*Figure* 2.9 and *Table* 2.2).

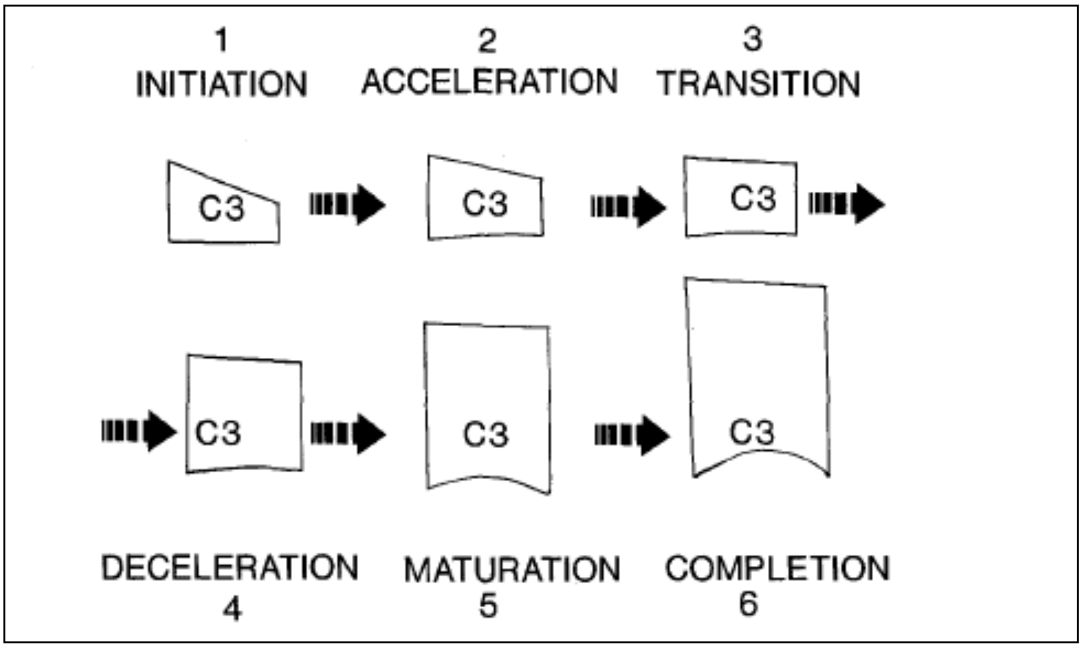


Figure 2.9: Cervical vertebral maturation indicators using CV3 as a guide.[29]

CATEGORY	STAGE	DESCRIPTION
1	INITIATION	During this stage the inferior borders of CV2, CV3 and CV4 are flat, the vertebral bodies are wedge shape with the superior vertebral borders tapered from the posterior to anterior wall. Adolescent growth is just beginning and 80-100% of adolescent growth is still expected.
2	ACCELERATION	During this stage concavities being to develop in the inferior borders of CV2 and CV3 while the inferior border of CV4 remain flat. The cervical vertebral bodies of CV3 and CV4 are nearly rectangular in shape. An acceleration of growth is beginning at this stage with 65-85% of adolescent growth remaining.
3	TRANSITION	Distinct concavities are seen in the inferior borders of CV2 and CV3 while a concavity in the inferior border of CV4 is developing. The cervical bodies of CV3 and CV4 are rectangular in shape. Adolescent growth is accelerating with 25-65% of growth expected.
4	DECELERATION	Distinct concavities are seen in the inferior borders of CV2, CV3 and CV4. The cervical bodies of CV3 and CV4 are a more square shaped appearance. Adolescent growth decelerates dramatically at this stage with 10-25% of adolescent growth expected
5	MATURATION	More accentuated concavities are seen in the inferior borders of CV2, CV3 and CV4. The vertebral bodies of CV3 and CV4 are nearly square to square in shape. Final maturation of the cervical vertebrae occurs in this stage with 5-10% of adolescent growth expected.
6	COMPLETION	Deep concavities are seen in the inferior borders of CV2, CV3 and CV4. The bodies of CV3 and CV4 are square or greater in a vertical dimension than in a horizontal dimension. Growth is considered to be complete at this stage with little or no adolescent growth expected.

Table 2.2: Six categories of cervical vertebral maturation.[29]

In summary, Hassel and Farman assessed the amount of growth in each of the six categories as a percentage of the adolescent growth expected and concluded that the greatest amount of expected growth was found between categories 2 and 3 with a wide range of 65-85% and 25-65% respectively. This study confirmed Lamparski's original findings, correlating cervical vertebral maturation with the Fishman's SMI method of assessment of skeletal maturity and concluded that the Hassel and Farman method of CVM assessment was both valid and reliable. Further studies then went on to corroborate the effectiveness of the Hassel and Farman method as a maturational indictor.[22, 31, 32, 128] The Hassel and Farman's index was evaluated in relation to increases in body height and it was concluded that the cervical vertebral analysis had a comparable high reliability and validity as the hand-wrist bone analysis in the assessment of individual skeletal maturity.[132] It was reported that this method was appropriate to estimate the maturation stage in both sexes,[32] and provided a more detailed description of each stage utilising a limited number of vertebral bodies on which to perform the staging process. However, there are some issues with the study design that limit the reliability of the results. The authors reported a high inter- and intra-observer agreement, using Pearson's correlation, which is an inappropriate statistical analysis, as it measures association rather than agreement.[41] Further limitations of Hassel and Farman's method include its reliance on cephalometric tracings of the vertebrae, which can introduce errors.[133] The authors described distinct stages of maturation for a continuously changing area of anatomy where in reality, each maturation stage blends into the next, making it sometimes difficult to differentiate borderline cases. The research recommends that the cervical vertebral maturation method be used as an additional tool to aid determination of the growth potential of the adolescent patient and to augment other observations by the orthodontist and advised that one diagnostic test should not be too heavily relied on.[36]

2.5.3.2 Franchi, Baccetti and McNamara 2000

The Lamparski method[26] was modified by Franchi *et al*.[3] to make it applicable to both sexes and easier to use. They described six developmental stages from examining the morphology of the 2nd to 6th cervical vertebrae (*Table 2.3* and *Figure 2.10).*

STAGE	Cvs	DESCRIPTION
1	Cvs1	The inferior borders of all cervical vertebrae are flat. The superior borders are tapered from posterior to anterior.
2	Cvs2	A concavity develops in the inferior border of the second vertebra. The anterior vertical height of the bodies increases.
3	Cvs3	A concavity develops in the inferior border of the third vertebra.
4	Cvs4	A concavity develops in the inferior border of the fourth vertebra. Concavities in the lower borders of the fifth and of the sixth vertebrae are beginning to form. The bodies of all cervical vertebrae are rectangular in shape.
5	Cvs5	Concavities are well defined in the lower borders of the bodies of all six cervical vertebrae. The bodies are nearly square in shape and the spaces between the bodies are reduced.
6	Cvs6	All concavities have deepened. The vertebral bodies are now higher than they are wide.

Table 2.3: A description of the six cervical vertebrae maturational stages.[3]

Franchi *et al*.[3] was a retrospective study that evaluated longitudinal records of 24 subjects (15 males and 9 females) from the University of Michigan Elementary and Secondary School Growth Study to assess the validity of the cervical vertebral maturation method for the evaluation of mandibular skeletal maturity and changes in body height. They concluded that this modified staging index detected the greatest increment in statural height, which was found to occur between stage 3 and stage 4 in all males and 87% females and corresponded to the greatest changes in mandibular length. They also reported a high inter- and intra-reliability of 98.6% and 100% respectively, however traced cervical vertebrae were used and the raters had a 'research level' of knowledge, which may have inflated the reliability.[27, 33, 41] Franchi and his team advocated this staging method for the prediction of the pubertal peak in mandibular growth, with the peak spurt yet to occur if either CVM stage 1 or stage 2 was recorded in the individual patient.

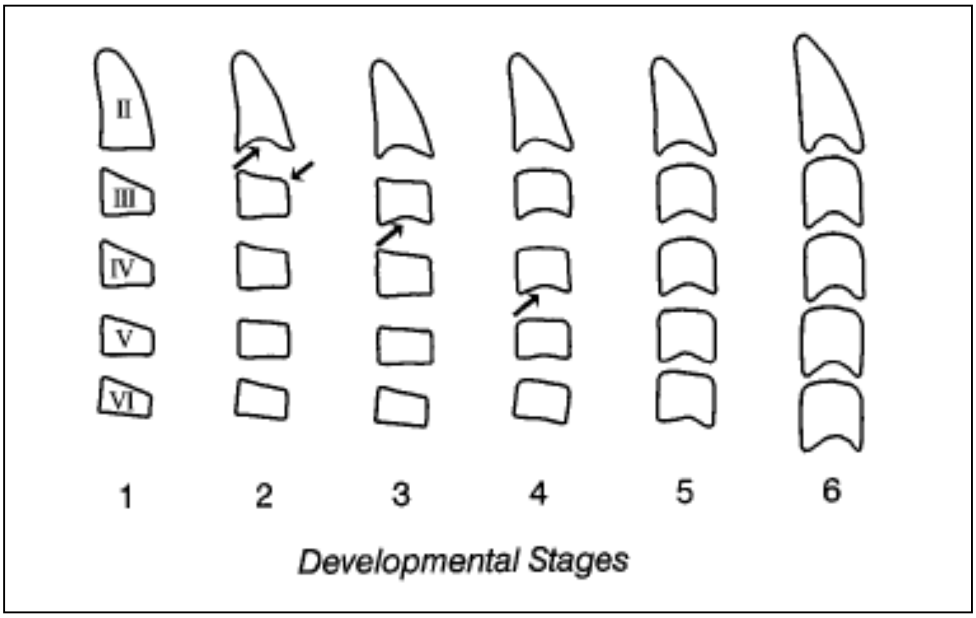


Figure 2.10: The six stages in cervical vertebral maturation described by Franchi *et al.*[3]

While this study provides strong evidence to support the validity of the CVM method for evaluation of skeletal maturity and the onset of pubertal peak in craniofacial growth, its limitations, due to the small sample size, intimate relationship of the authors to the development of the index, possible bias associated with using a retrospective historical sample and the ambiguity with the sample selection process, must be considered.

2.5.3.3 Baccetti, Franchi and McNamara 2002[4]

Two years later Baccetti *et al.*[4] published an improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. Several improvements were made including a reversion back to the assessment of three cervical vertebrae (CV2-CV4) as with the Hassel and Farman method instead of five cervical vertebrae (CV2-CV6) described in the Franchi *et al.*[3] method. They reduced the number of maturational stages from six to five and modified the definitions of the stages, based on comparative assessment of between stage changes (*Figure* 2.11). This had the advantage of being able to assess changes from a single lateral cephalogram, even when a protective thyroid collar is worn.

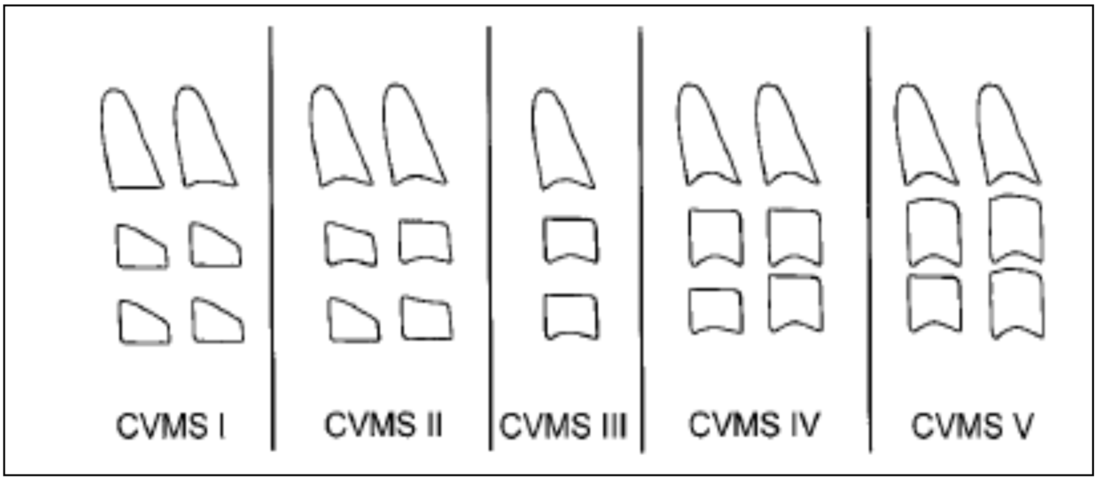


Figure 2.11: The newly improved CVM method[4] (five developmental stages, CVMS I through CVM VI). Different combinations of morphological features in the bodies of C2, C3, and C4 are presented for the new method.

These modifications were based on the longitudinal records of 30 subjects (18 males and 12 females) again, from the Michigan Growth Study. The maximum increase in total mandibular length (Condylion-Gnathion) between two consecutive cephalograms was used to define the peak in mandibular growth at puberty and this was used to represent the peak pubertal growth in each of the individual subjects. These two consecutive cephalograms comprising of the maximum mandibular growth, together with two earlier consecutive cephalograms and two later consecutive cephalograms formed a total sample of six images per subject. The morphology of the second, third, and fourth cervical vertebrae bodies in the six consecutive images were examined using both visual analysis and cephalometric evaluations using digitising software. Five different maturational stages were defined (CVMS I through to CVMS V) with the peak in mandibular growth occurring between CVMS II and CVMS III. This differed from Franchi *et al*. [3] who indicated that peak mandibular growth occurred between stage 3 and 4 in the six stage method. The authors concluded that this staging method is particularly useful when skeletal maturity has to be appraised on a single cephalogram *(Table 2.4)* with a high inter-observer reliability of 96.7%. However, only two examiners, who were familiar with the assessment method and participated in the selection of images assessed the reliability, which may account for the high percentage agreement.

CVM Stage	Definition
CVMS I	The lower borders of all the three vertebrae are flat, with the possible exception of a concavity at the lower border of C2 in almost half of the cases. The bodies of both C3 and C4 are trapezoid in shape (the superior border of the vertebral body is tapered from posterior to anterior). The peak in mandibular growth will occur not earlier than one year after this stage.
CVMS II	Concavities at the lower borders of both C2 and C3 are present. The bodies of C3 and C4 may be either trapezoid or rectangular horizontal in shape. The peak in mandibular growth will occur within one year after this stage.
CVMS III	Concavities at the lower borders of C2, C3, and C4 now are present. The bodies of both C3 and C4 are rectangular horizontal in shape. The peak in mandibular growth has occurred within one or two years before this stage.
CVMS IV	The concavities at the lower borders of C2, C3, and C4 still are present. At least one of the bodies of C3 and C4 is squared in shape. If not squared, the body of the other cervical vertebra still is rectangular horizontal. The peak in mandibular growth has occurred not later than one year before this stage.
CVMS V	The concavities at the lower borders of C2, C3, and C4 still are evident. At least one of the bodies of C3 and C4 is rectangular vertical in shape. If not rectangular vertical, the body of the other cervical vertebra is squared. The peak in mandibular growth has occurred not later than two years before this stage.

Table 2.4: Definitions of new improved CVM stages.[4]

2.5.3.4 Baccetti, Franchi and McNamara 2005[28]

Baccetti outlined the characteristics of an "ideal" biologic indicator of individual mandibular skeletal maturity, stating it should have at least five features and include:

- An efficacy in detecting the peak in mandibular growth. The method should present with a definite stage or phase that coincides with the peak in mandibular growth in the majority of subjects.
- No indication for additional x-ray exposure.
- An easy method of recording.
- Consistency in the interpretation of the data. The inter-examiner error in the appraisal of the defined stages or phases should be as low as possible.
- Usefulness for the anticipation of the occurrence of peak growth with the method presenting a definable stage or phase that occurs before the peak in mandibular growth in the majority of subjects.[28]

In 2005, Baccetti *et al.* introduced a modified version of their 2002 CVM staging method for the detection of peak mandibular growth.[28] Using the same longitudinal sample[4] they reverted back to a six-stage maturational index citing that it is a more valid index for the appraisal of mandibular skeletal maturity in the individual patient and a far more practical method for the clinician to apply (*Figure 2.12*). They used a description of each maturational stage, which did not rely on the previous stage definition.

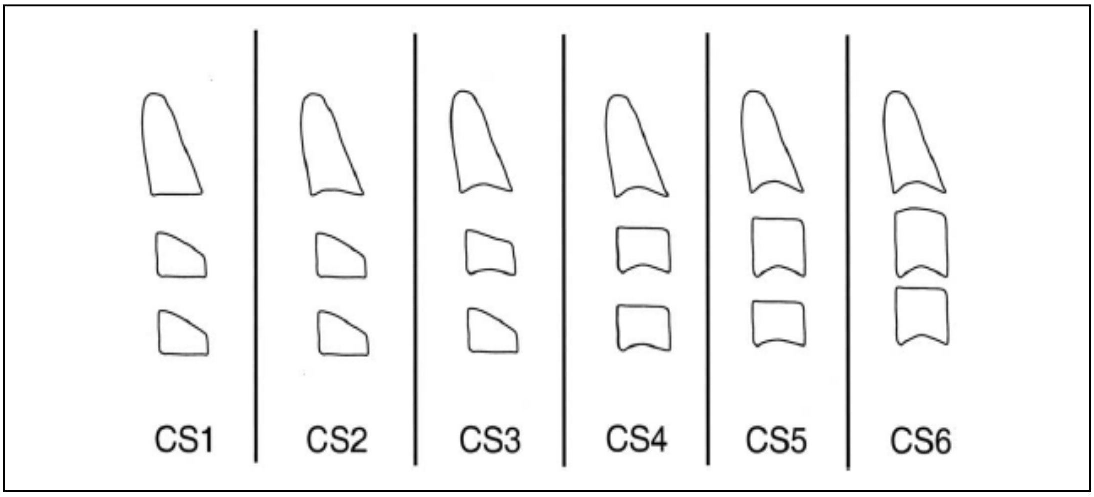


Figure 2.12: Schematic representation of the stages of the cervical vertebrae according to the newly modified method.[22]

Cervical Stage (CS) 1 and 2 were classified as pre-peak stages, with the peak in mandibular growth occurring between CS3 and CS4. However, the exact timing of peak mandibular growth remains vague, having occurred during the year after CS3 and one to two years before CS4.[109] CS6 indicates that the peak in growth had occurred at least two years earlier (*Table* 2.5).

Cervical Stage	Description
Cervical stage 1 (CS1)	The lower borders of all the three vertebrae (C2-C4) are flat. The bodies of both C3 and C4 are trapezoid in shape (the superior border of the vertebral body is tapered from posterior to anterior). The peak in mandibular growth will occur on average 2 years after this stage.
Cervical stage 2 (CS2)	A concavity is present at the lower border of C2 (in four of five cases, with the remaining subjects still showing a cervical stage 1). The bodies of both C3 and C4 are still trapezoid in shape. The peak in mandibular growth will occur on average 1 year after this stage.
Cervical stage 3 (CS3)	Concavities at the lower borders of both C2 and C3 are present. The bodies of C3 and C4 may be either trapezoid or rectangular horizontal in shape. The peak in mandibular growth will occur during the year after this stage.
Cervical stage 4 (CS4)	Concavities at the lower borders of C2, C3, and C4 now are present. The bodies of both C3 and C4 are rectangular horizontal in shape. The peak in mandibular growth has occurred within 1 or 2 years before this stage.
Cervical stage 5 (CS5)	The concavities at the lower borders of C2, C3, and C4 still are present. At least one of the bodies of C3 and C4 is squared in shape. If not squared, the body of the other cervical vertebra still is rectangular horizontal. The peak in mandibular growth has ended at least 1 year before this stage.
Cervical stage 6 (CS6)	The concavities at the lower borders of C2, C3, and C4 still are evident. At least one of the bodies of C3 and C4 is rectangular vertical in shape. If not rectangular vertical, the body of the other cervical vertebra is squared. The peak in mandibular growth has ended at least 2 years before this stage.

Table 2.5: Stages of the cervical vertebral maturation.[28]

Recent studies have evaluated the appropriateness of this staging method in identifying the peak mandibular growth. Ball *et al.*[109] used the six stage CVM method of Baccetti *et al*.[28] to assess a longitudinal sample from the Burlington Growth Centre and found that the peak mandibular growth occurred during CS3 for approximately 16% of the subjects, with 69% of subjects obtaining a peak in their mandibular growth at CS4. Their findings reported that progression from one CS stage to another did not occur annually as presented by Lamparski[26] but in fact, the length of time spent in each CS stage varied from 1.5 years to 4.2 years, depending on the particular stage. This finding may not have been possible to detect in previous small studies and narrow longitudinal records on which previous studies were based.[3, 4, 26, 28] However, it must be emphasised that two different landmarks were used for assessing total mandibular length with Franchi *et al.*[3] and Baccetti *et al*.[4, 28] using condylion (Co) and Ball et al. [109] using articulare (Ar). The anatomical landmark condylion is a more accurate measure of mandibular growth,[46] as it takes into account condylar growth/remodelling but, it is notoriously difficult to identify on the lateral cephalogram.[134] Articulare on the other hand is more easily identified than condylion on a lateral cephalogram[134] however, it is not a landmark which pertains exclusively to the mandible and it does not show full mandibular length. In addition, articulare does not undergo the same positional change following a change in the amount or direction of growth in the mandible. Finally, measuring total mandibular length using articulare introduces the possibility of a false interpretation of an increase in mandibular length, if the condyles are positioned anteriorly within the glenoid fossa.

While the limitations of using small sample sizes for the development and modification of the cervical vertebral staging method have been previously outlined, the use of longitudinal data, rather than cross-sectional data for the development of the staging index, is more beneficial. Modifications made to the index were established according to changes in the annual rate of increase in total mandibular length instead of comparison with the hand-wrist skeletal age.[36] Many studies have been carried out to assess the validity of the CVM method, with the majority of these studies using convenience samples.[3, 4, 22, 29, 31, 135] These cross sectional samples present a number of limitations to the quality of evidence, as the information they provide is relatively insensitive to individual variability. Longitudinal research has been advocated for the study of craniofacial growth as it provides a much clearer understanding of the use of cervical vertebrae as an indicator of skeletal maturation.[136] A recent systematic review and meta-analysis concluded that the CVM method of skeletal maturation was valid and recommended that it should replace the HWR in predicting pubertal growth spurt.[121] The CVM method has been correlated with mandibular growth using longitudinal data[3, 67, 68, 77, 78, 109] with a number of studies stating that they would not recommend the sole use of the CVM method for the determination of skeletal age in treatment planning.[33, 40, 41, 131]

2.6 Reliability of the CVM method

When assessing whether a new index is useful to a clinician, it is important that its interpretation is not a product of chance and instead is a measure of precision. This means that in order for any new evaluation tool to be implemented for clinical diagnostic purposes it must be reliable. Reliability implies that a measurement should be the same if repeated by the same or different observer.[131] Many reliability studies have been undertaken, over recent years, in order to establish the value of the CVMM in the clinical diagnosis of the individual's level of skeletal maturity.[3, 4, 22, 23, 25, 29, 31-38, 40-42, 78, 109, 130, 131, 137-141] However, controversy still remains around the reliability of the CVM assessment method, with much of the research reporting very different outcomes.[29, 33, 42, 137] From the large body of published research on CVMM reliability, an explanation of the variability in the results can be identified by the broad methodologies utilised by the different authors. The methodologies and reported findings of the main CVM reliability studies carried out to date are summarised in *Table 2.6*. Studies pertaining to CVMM's high reliability[3, 100] have been criticised for using observers with a research level of knowledge.[130] In addition, some observers have been involved in the selection and preparation of the images used in the study.[3, 4, 23, 29, 31, 35, 39, 42, 78, 109, 139] These factors are believed to inflate the reported reliability of the CVM method. Several reliability studies have used observers with a wide range of clinical orthodontic experience from dental undergraduates to associate professors with the reported reproducibility of CVMM not improved by the level of clinical experience.[40]

The number of raters used to assess the intra-observer reliability varied greatly ranging from a single rater[3, 23, 29, 34, 35, 39, 70, 78, 109, 130, 140] to multiple raters of two,[31, 139] five,[131, 138] ten,[41, 137, 141, 142] eleven,[36] thirteen,[25] twenty[42] and thirty.[40] The number of raters used for inter-observer reliability also varied greatly in all studies and ranged from two[3, 4, 22, 29, 31, 35, 39, 139] three,[34] four,[140] five,[131, 138] ten,[41, 137, 141, 142] eleven,[36] thirteen,[25] twenty,[42] and thirty[40] observers. None of the authors explained or justified the number of raters used.

The sample source of the images used in the rating process range from using small longitudinal historical samples familiar to the researcher(s)[3, 4, 29, 36, 109, 131] to contemporary samples.[25, 42, 138, 140, 141] The number of sample images used in these studies varied greatly from fifteen[25] to six-hundred[70] images with only a small number of studies basing their sample size on the statistical analysis used.[42, 109] Small sample sizes and too few observers may reduce the generalisability of the results.[42]

Several studies have reported their findings using inappropriate statistical methods in their analyses[29, 32, 130] while other studies failed to provide the appropriate training in the CVMM[33] or failed to report whether training was provided.[31, 34, 35, 39, 70, 130, 138, 139] Standardised CVMM training prior to execution of the reliability study has been reported to increase reliability.[25, 42, 137, 140]

Many different CVM staging methods were used in these reliability studies and this may in part explain the large variability in the results. Several studies used traced outlines of the cervical vertebrae[2-4,23, 32, 78, 109, 130, 137] and this may have introduced bias in the staging results affecting reliability outcomes. Of particular interest is that much of the research to date has been limited to using either cropped lateral cephalograms, illustrating CV2-CV4[29, 130] or CV1-CV4[33, 41, 138] or full lateral cephalograms.[25, 34, 36, 42, 140] It has been argued that the use of the full lateral cephalogram images reflects everyday clinical practice while other authors suggest that it is preferable to use a cropped image in order to eliminate any possible influences from additional indicators of maturity, such as the stage of dental development, that might bias the observer.[41] The wide variation in the methodologies used in CVM reliability studies to date is reflected in the large variation in both intra- and inter-observer reliability found.

Recently, a systematic review on the reliability of the CVM method to predict the pubertal growth spurt highlighted the low level of evidence and the numerous methodological flaws of the published studies.[27] They suggested that a more robust testing of the staging index is advocated in order for it to be endorsed as a clinically applicable biological diagnostic tool.[27]

Reliability study			Participants				Intervention							Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Hassel & Farman[29]	Hassel & Farman[29]	Longitudinal Broadbent growth study	20 9 m 11 f	2 independent evaluators (A + B)	1	Expert Named author	✓	-	-	-	NR	NR	✓	Pearson correlation r^2	0.85 A 0.90 B	1.0
Kucukkeles *et al.*[38]	Hassel & Farman[29]	Cross-sectional Turkish population	20	3 independent observers	3 independent observers	NR	✓	-	-	-	×	✓	✓	Pearson correlation r^2	<u>A+B</u> 0.87 <u>A+C</u> 0.87 <u>B+C</u> 0.88	<u>A</u> 0.81 <u>B</u> 0.96 <u>C</u> 0.74
Franchi *et al.*[3]	Franchi *et al.*[3]	Longitudinal Michigan Growth Study	50	2	1	Expert Named authors	✓	-	-	-	NR	NR	✓	Percentage agreement	98.6	100

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Chang *et al*[37]	Modified Lamparski[26]	Cross-sectional Taiwanese population Randomly selected	25	2	1	Named authors	✓	-	-	-	NR	NR	✓	Wilcoxon sign rank	W=0.5 df=24 p=1.0 24/25 agree	W=1.5 df=24 p=0.50 24/25 agree
San Román *et al.*[32]	Lamparski[26] Hassel & Farman[29]	Cross-sectional Spanish Caucasian	50	-	NR	NR	✓	-	-	-	NR	NR	✓	Pearson correlation r^2	-	0.96-0.99
Baccetti *et al.*[4]	Baccetti *et al.*[4]	Longitudinal Michigan Growth Study	180 30x6 consecutive LC	2	NR	Expert Named authors	✓	-	-	-	NR	NR	✓	Percentage agreement	96.7%	-
Uysal *et al.*[31]	Hassel & Farman[29]	Turkish population	30	2	2	Named authors	✓	-	-	-	NR	NR	NR	Spearman Brown	0.987	0.955

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Flores-Mir *et al.*[130]	Baccetti *et al.*[4]	Canadian population	10 triplicated	-	1	Research	✓	-	-	-	×	✓	✓	Intra-class correlation coefficient	-	0.889 (0.723-0.968)
Özer *et al.*[139]	Küçükkeles[38]	Turkish population	150	2	2	Named authors	✓	-	-	-	NR	NR	NR	Percentage agreement	98%	99.3%
Baccetti *et al.*[70]	Baccetti *et al.*[28]	Cross-sectional Italian population	600	-	1	Expert	✓	-	-	-	NR	NR	NR	Percentage agreement Kappa	-	94.8% 0.90
Alkhal *et al.*[35]	Baccetti *et al.*[28]	Cross-sectional Southern Chinese population	25	2	1	Principle investigator and 'other' investigator	✓	-	-	-	NR	NR	NR	Kappa	0.846	24/25 reported the same

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Soegiharto *et al.*[23]	Baccetti *et al.*[4]	Cross-sectional 200 Indonesian 100 White	300	-	1	Principle investigator	✓	-	-	-	NR	NR	✓	Cohen Kappa	-	0.85-0.97 Indonesia 0.94-0.95 White
Lai *et al.*[34]	Baccetti *et al.*[28]	Cross-sectional Taiwanese population	30	3	1	NR	✓	-	-	-	✓	-	NR	Percentage agreement Spearman's rank correlation coefficient	90-93.3% 0.96-0.98	90%
Wong *et al.*[39]	Baccetti *et al.*[28]	Same as **Alkhal** *et al.*[35]	25	2	1	Named author *Alkhal* & 'another' orthodontist	✓	-	-	-	NR	NR	×	-	23/25 reported the same	23/25 reported the same

Reliability study				Participants			Intervention								Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Gabriel *et al.*[33]	Baccetti *et al.*[28]	Longitudinal growth records of untreated subjects	90 30 & 30 pairs	10	10	Private & experienced practice orthodontists Independent of research	-	✓	-	2	×	✓	×	Kendall coefficient of concordance Weighted Kappa	0.72-0.76	 0.36-0.79
Jaqueira *et al.*[140]	Hassel and Farman[29] Baccetti *et al.*[4] Seedat and Forsberg[143]	Brazilian population Private orthodontic practice	23	4	1 (radiologist)	1X Radiologist and 3X Orthodontists with various years of clinical experience	-	✓	-	NR	✓	×	×	Weighted Kappa ***Hassel and Farman***[29] ***Baccetti et al.***[4] ***Seedat and Forsberg***[143]	0.64	 0.64-0.76 0.73-0.75 0.54-0.66

Reliability study			Participants				Intervention							Outcome		
Author	*CVM Method*	*Image Sample*	*Image Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Nestman *et al.*[41]	Baccetti *et al.*[28]	Same as **Gabriel** *et al.*[33]	30 15 white boys & girls	10	10	Private & experienced practice orthodontists Independent of research	-	✓	-	2	×	✓	×	Kendall's W Coefficient Percentage agreement	0.45 -	- 44-62%
Ball *et al.*[109]	Baccetti *et al.*[28]	Male subjects only Longitudinal Burlington Growth Centre	72	-	1	Principle investigator	-	✓	-	1	×	✓	✓	Kappa	-	0.943-1.0
Zhao *et al.*[36]	Baccetti *et al.*[28]	Longitudinal records Peking University	86	11	11	Independent of research	-	✓	-	NR	✓	×	×	Kendall's W k^w % Agree	0.83-0.83 - 39.3-47.7	- 0.53-0.86 40.7-79.1

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Sample*	*Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*Traced*		*Inter*	*Intra*
Ballrick *et al.*[25]	Baccetti *et al.*[4]	University based images	15	13	13	Independent of research Orthodontic residents	-	✓	-	NR	✓	×	×	Kendall's W Weighted Kappa Percentage agreement	0.79-0.87 - 46.7-53.2%	- 0.86 62%
Perinetti *et al.*[137]	Baccetti *et al.*[28]	Pre-treatment images	72 37f, 35m 10 cases/CVM stage 12 reference standards	10	10	6 PG, 2 PD 1 AP, 1 UG 2 different Universities Independent of research	-	✓	-	2	×	✓	✓	Kendall's W Weighted Kappa (linear)	0.9-0.91 -	- 0.52-0.9
Rongo *et al.*[40]	Baccetti *et al.*[28]	Patients of orthodontic department University of Naples	50 25 f, 25 m	30	30	10 Junior 10 PG 10Specialist Independent of research	-	✓	-	1	×	✓	NR	Kendall's W Weighted Kappa (linear) % Agree	0.7-0.81 - 42-46%	- 0.24-0.81 26-78%

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Sample*	*Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*Traced*		*Inter*	*Intra*
Danaei *et al.*[22]	Hassel & Farman[29]	Cross-sectional Subjects with short stature	178 90f,88m	2	NR	MFU radiologist Orthodontic resident Named author	-	✓	-	NR	×	✓	×	Weighted Kappa	0.89	0.96
Predko-Engel *et al.*[141]	Baccetti *et al.*[28]	Patient records Czech Republic	50	10	10	3 experienced Orthodontists 7 inexperienced in CVM assessment Association with research NR	-	✓	-	NR	×	✓	NR	Weighted Kappa	0.28	0.44

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Sample*	*Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*Traced*		*Inter*	*Intra*
Sohrabi *et al.*[138]	Baccetti *et al.*[4]	Pre-treatment images Iranian patients	70	5	5	5 experienced orthodontists Association with research NR	✓	-	-	NR	×	✓	×	Kappa Fleiss kappa	0.48 P1=0.45 P2=0.51	 0.59-0.85
Engel et al.[131]	Baccetti *et al.*[28]	Nijmegen Growth Study	29	5	5	Senior orthodontic residents Independent of research	-	✓	-	NR	×	✓	×	Kappa	0.30	0.36
Gray *et al.*[78]	Baccetti *et al.*[28]	Burlington Growth Study	25	-	1	Named author Principle investigator	-	✓	-	NR	NR	NR	✓	Percentage agreement Kappa Weighted Kappa	- - -	84% 0.80 0.89

Reliability study				Participants			Intervention							Outcome		
Author	*CVM Method*	*Sample*	*Sample n*	*n*		*Experience level*	*Training*				*Images*			*Statistical analysis*	*Reliability*	
				Inter-reliability	*Intra-reliability*		*NR*	*Y*	*N*	*n*	*F*	*C*	*T*		*Inter*	*Intra*
Rainey *et al.*[42]	Baccetti *et al.*[28]	Contemporary sample of pre-treatment images	72	20	20	18 raters independent of research No previous CVM experience 2 named authors involved in study design 9 specialist 11 orthodontic trainees	-	✓	-	2	✓	×	×	Percentage agreement Weighted Kappa	88% 0.68	89% 0.70

Table 2.6: Summary of pertinent CVM Method reliability studies.

Key: n; number, NR; not reported, Y; yes, N; no, F; full lateral cephalogram, C; cropped lateral cephalogram, T; traced, m; males, f; females, PG; Postgraduates, PD; Post Doctorate, AP; Assistant Professor, UG; Undergraduate, Junior; <1 year orthodontic experience, Postgraduates; 2-4 years orthodontic experience, Specialists; >7 years orthodontic experience, MFU; maxillofacial unit

2.7 Background and justification of methodology

2.7.1 Sample size

One of the main methodological flaws highlighted in a recent systematic review assessing the reliability of the CVM staging method was the lack of sample size calculations.[27] The determination of an adequate sample size is an important part of a study's methodology and sufficient sample sizes are required to provide both credible and precise results. To date there has been little emphasis on sample size calculations, particularly related to orthodontic studies[144] therefore results of previous reliability studies with insufficient sample sizes must be interpreted with caution[3, 25, 29, 31, 32, 34, 35, 39-41, 78, 130, 138, 140, 141], as it is necessary to have a sufficient sample size in a study to conclude with reasonable confidence that an index is reliable and precise. In any study a degree of compromise between statistical criteria and practicality is required when determining a suitable sample size.[144] A sample size calculation was performed to accurately determine the number of images necessary to test the CVM staging index, reducing selection bias while also increasing the generalisability of the study. This facilitated a balance between the recruitment of a suitable sample size large enough to evaluate the reliability of the modified CVM staging index effectively whilst minimising bias and simultaneously maintaining the number of images to a practical size for the observer to rate comfortably and without the onset of fatigue. The literature indicates that the number of images evaluated for the assessment of CVM reliability is highly variable, ranging from 178[22] to as little as 15 images,[25] with only three studies executing a formal sample size calculation similar to this study.[42, 109, 137]

2.7.2 Reliability

Reliability is a measure of agreement within a rater/observer over time or between raters/observers.

2.7.2.1 Intra-observer agreement

Intra-observer or intra-rater agreement is defined as the agreement between the same observer/rater, using the same scale, to assess the same object at different times.[145]

2.7.2.2 Inter-observer agreement

Inter-observer or inter-rater or agreement is defined as agreement between the different observers/raters, using the same scale to assess the same objects.[145]

In order to assess the reliability of an index it is often necessary to obtain multiple readings, either from the same individual at different time points or from different individuals at the same time point. The degree of agreement, among the various observers, can indicate the consistency or reliability of the index. If agreement among the observers is high, one can be confident that the index can be used reliably while low agreement reflects less confidence in the index.

The primary outcome measure of this study was to assess the intra-observer and inter-observer reliability of CVM stage determination among a group of orthodontic clinicians for both the full and cropped lateral cephalogram images.

Observers are an important source of measurement error.[146] Consequently, reliability studies are conducted in experimental conditions that have been developed and standardised to facilitate the assessment of observer variability.[147] Reliability studies should employ statistical analyses that are appropriate for the type of data used, i.e. qualitative or quantitative, binary, categorical, ordinal or continuous and should account for the fact that observers will sometimes agree or disagree by chance.[148] In order to analyse the extent to which there is agreement, other than that expected by chance, the kappa statistic was used, as the data were categorical.[147-150] Observer agreement, where the data are categorical, can be analysed using kappa and weighted kappa statistics.[149, 150]

2.7.3 The Kappa coefficient

The Kappa coefficient is a simple measure of reliability for nominal scales providing a quantitative measure of the magnitude of agreement between observers.[148] This coefficient is described as a chance-corrected agreement measure[146] and is a commonly used descriptive statistic for summarising agreement between observers on a categorical scale.[151]

However, while Cohen's kappa takes into account disagreement between observers it does not account for the degree of disagreement i.e. all disagreements are treated the same.[150] Further development of the kappa coefficient involving the use of weighting, was devised to overcome this issue.[150] The weighted kappa (k^w) is calculated using a predefined table of weights, which measure the degree of disagreement between observers' ratings. The higher the disagreement the lower the weight assigned. Weighted kappa is used to measure the precision of agreement between observers (inter-observer agreement) and within the same observer (intra-observer) involving categorical and ordinal data.[145]

While percentage agreement may be used to calculate agreement between observers it can be misleading as it includes agreements which can be accounted for by chance and this gives an over-estimation of the level of agreement.[149, 150]

2.7.3.1 Unweighted Kappa

The kappa coefficient can be used to summarise the cross-classification of two nominal variables with identical categories.[152] This coefficient of intra- and inter-observer agreement for nominal scales represents the proportion of agreements after chance agreement is removed.[149] The use of unweighted kappa implicitly assumes that all disagreements are of equal significance and is therefore deemed unsuitable for ordinal data.[150] When the relative seriousness of each kind of disagreement can be specified, a weighted kappa can be employed, with the proportion of weighted agreement corrected for chance.[153]

2.7.3.2 Weighted Kappa

The weighted kappa statistic was developed to reflect disagreements between two observers that are not of equal importance.[150] The use of weights permits the closeness of agreement between categories to be described. For example, disagreements on two distant categories are more heavily penalised than disagreements on neighbouring categories on an ordinal scale. This concept was further extended to study the agreement between many raters.[154] Weighted kappa is the most appropriate statistic to assess observer agreement with ordinal data.[150] As the CVM staging index is a categorical and ordinal scale, the weighted kappa statistic was deemed most suitable to allow credit for complete and partial agreement amongst the raters.[150] While weighted kappa and the intra-class correlation coefficient are mathematically equivalent, the latter assesses observer reliability when data are continuous.[153]

The two most commonly applied weights used for weighted kappa are linear weights and quadratic weights. Linear weightings are proportional to the number of categories apart, whereas quadratic weightings are proportional to the square of the number of categories apart.

2.7.3.3 Linear weighted kappa

Linear weighted kappa can be defined as weights that are proportional to the deviation of individual ratings such as the numbers of categories of disagreement. The linear form of the kappa coefficient presents an advantage over the quadratic version, as it is less sensitive to the number of categories and it is this reason that linear weight kappa is preferred when the number of categories of the ordinal scale is large[155] (*Table* 2.7).

	1	2	3	4	5	6
1	1	0.8	0.6	0.4	0.2	0
2	0.8	1	0.8	0.6	0.4	0.2
3	0.6	0.8	1	0.8	0.6	0.4
4	0.4	0.6	0.8	1	0.8	0.6
5	0.2	0.4	0.6	0.8	1	0.8
6	0	0.2	0.4	0.6	0.8	1

Table 2.7: Linear kappa weighting.

2.7.3.4 Quadratic kappa

Quadratic weighted kappa coefficients are weights that are proportional to the square of the deviation of individual ratings. It has been reported that quadratic weighted kappa can be interpreted as an intra-class correlation coefficient in a two-way analysis of variance setting and is equivalent to the product-moment correlation. As a result its use has been recommended for its practical interpretation[153, 156] While commonly used in clinical practice, the value of quadratic weighted kappa tends to increase as the number of categories increases, with the greatest variation occurring between two to five categories, the range of categories most frequency used in practical applications.[155] Furthermore, this statistic may produce high values even when the level of exact observed agreement is low. It has been concluded that quadratic weighted kappa tends to behave as a measure of association instead of an agreement coefficient and therefore may lead to misleading conclusions[157] (*Table* 2.8).

	1	2	3	4	5	6
1	1	0.96	0.84	0.64	0.36	0
2	0.96	1	0.96	0.84	0.64	0.36
3	0.84	0.96	1	0.96	0.84	0.64
4	0.64	0.84	0.96	1	0.96	0.84
5	0.36	0.64	0.84	0.96	1	0.96
6	0	0.36	0.64	0.84	0.96	1

Table 2.8: Quadratic kappa weighting.

As the value of the weighted kappa coefficient can vary considerably according to the weighting scheme used (linear or quadratic) the reporting of both linear and quadratic weighted kappa coefficients in ordinal agreement studies is recommended as it better describes the shape of the disagreement distribution rather than reporting only one of the two coefficients.[158] However, if a single index of agreement for an ordinal scale is to be used then it is recommended that the linear weighted kappa is used instead of the quadratic weighted kappa.[151]

In this study, both linear and quadratic kappa weightings were determined for intra-reliability by entering data in to *StatsDirect* Statistical software package.

Inter-observer reliability was determined using both the Fleiss-Nee-Landis kappa and the Berry-Mielke statistic. The Fleiss-Nee-Landis kappa coefficient is necessary for measuring agreement when there is more than two observers.[159] The Berry-Mielke Universal R coefficient of agreement is a generalisation of Cohen's kappa to an ordinal measurement scale that is suitable for more than two raters. Using categorical data, R is equivalent to a linearly weighted kappa statistic i.e. R is chance-corrected and appropriate for the measurement of reliability.[160]

2.7.3.5 Interpretation of Kappa

The kappa statistic represents the proportion of agreement corrected for chance or in other words, agreement beyond chance[157] and scaled to vary from -1 to +1. A negative value indicates poorer than chance agreement, zero indicates exactly chance agreement and a positive value of 1 indicates perfect agreement.[161] A range of labels has been devised in order to maintain a consistent nomenclature when describing the relative strength of agreement associated with kappa statistics[147] (*Table* 2.9). Much criticism has been directed towards these divisions due to their arbitrary nature[162] and that the use of different weighting schemes can lead to different conclusions, however they do provide a useful benchmark for interpretation of the kappa coefficient and are a widely accepted scale of reproducibility.[147]

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Table 2.9: Interpretation of Kappa.[147]

While *P* values and confidence intervals can be calculated alongside the kappa values, it is important to note that both these parameters are sensitive to the sample size and with a large enough sample size, any kappa above 0 will become statistically significant.[148]

Chapter 3: Rationale for current study

If the CVM staging index is to be advocated for routine use to aid orthodontic diagnosis and treatment planning, then it must be reliable, attaining sufficient consistency in its measures both for different observers (inter-observer reliability) and for the same observers on different occasions (intra-observer reliability). While a reasonable body of evidence pertaining to the reliability of the CVM staging method exists, major methodological flaws have been highlighted[33] and a systematic review concluding that the evidence on the reliability of CVM staging is of poor quality.[27]

The reliability of the CVM staging index continues to be subjected to conflicting reports in the literature.[27] Many of these studies differ in their methodologies which may partially account for the differing results in CVM staging reliability. One of the great flaws associated with the current body of research on CVM reliability is the failure to report comprehensively on the methodologies used. Many studies have either used traced and/or cropped lateral cephalogram images to include only the pertinent cervical vertebrae, CV2, CV3 and CV4.[3, 23, 29, 32, 37, 38, 130] Using traced cervical vertebrae[23, 29, 32, 130, 137] instead of the actual cropped radiographic image[33, 40, 41, 138, 141] may introduce a bias potentially inflating the reliability of CVM staging index and detracting away from how assessment of CVM would be undertaken in clinical practice.[33] However, using a cropped, untraced image removes the potential influence that visibility of the dental development may offer and bias it may introduce. Other studies have investigated CVM staging reliability using unaltered, full lateral cephalogram images and these have been found to be relaible.[25, 34, 36, 42, 140] These investigators argue that manipulating the lateral cephalogram image can reduce its resemblance from a normal clinical environment.

While previous research has focused on the reliability of the CVM staging index using either full or cropped or traced images, no study, to my knowledge, has examined if the reliability is affected by using both full or cropped images when assessed by trainee or specialist orthodontists. As there remains a gap in our knowledge base, regarding whether the cropping of the lateral cephalogram image affects the reliability of the CVM staging method, it was thought to be useful to determine whether cropping the image improved the reliability of the CVM assessment or not.

The purpose for this study was therefore to:

1. Address the methodology flaws of previous research,
2. Determine the reliability of the improved method of CVM staging[28] when assessing cropped images and
3. Determine whether cropping the images improves the reliability of CVM staging when compared to staging images of full lateral cephalogram.

Chapter 4: Diagnostic Quality of Lateral Cephalograms for Cervical Vertebral Maturation (CVM) Staging

4.1 Background

The timing of orthodontic treatment can have a significant effect on the outcome of dentofacial orthopaedic treatment that is designed to utilise remaining growth to aid correction of an underlying dentoskeletal discrepancy.[28] Orthodontic treatment that is aiming to modify growth ideally should commence before the onset or early during the pubertal spurt to take advantage of this accelerated growth period. Therefore the orthodontist is required to evaluate the developmental status of the adolescent and assess remaining growth in order to formulate a treatment plan best suited to treat the individual's malocclusion and plan an appropriate retention strategy. In addition, when undertaking combined orthodontic/orthognathic treatment or placing implants to replace congenitally missing teeth, it is necessary to know when growth has slowed and/or has ceased.

There are many advantages of applying CVM staging in order to assess the stage of growth and development of orthodontic patients including:

- Elimination of the need for additional radiographic exposure since the vertebrae are readily viewed from the lateral cephalogram taken as a pre-treatment record,[27]
- It concerns an area of anatomy with which the clinician is familiar,
- It is reliable[42] and
- It presents definitive stages that coincide with growth.[28]

The CVM staging method therefore, permits the orthodontist to make more informed decisions regarding the timing of treatment and treatment modalities. However, in order to use this staging method the second, third and fourth cervical vertebrae need to be visible on the lateral cephalogram image.

4.2 Aims

The aims of this audit were to determine:

1. If the second, third and fourth cervical vertebrae were routinely visible on lateral cephalograms taken at the Liverpool University Dental Hospital and
2. That the lateral caphalograms were of a sufficient diagnostic quality to enable the staging of skeletal maturation using the modified CVM staging method.[28]

4.3 Design

Lateral cephalograms requested by Specialty Registrars (St1), Senior Post-CCST Specialty Registrars (St4) and a post-graduate student for patients in their respective cohorts and taken between October 2014 and January 2015, were assessed retrospectively for the inclusion of the second, third and fourth cervical vertebrae and the diagnostic quality to enable use of the CVM staging method (*Figure 4.1*).[28]

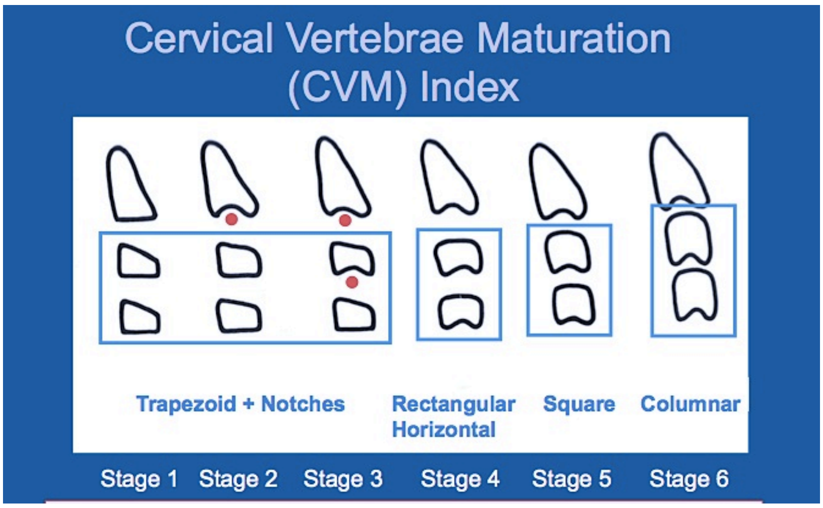


Figure 4.1: CVM Index[28]

4.4 Standards

In order for lateral cephalograms to be useful for applying the CVM method, they should have the second, third and fourth cervical vertebrae (CV2-CV4) clearly visible and be of sufficient diagnostic quality to allow application of cervical vertebrae maturation staging index.[27, 29]

Based on previous audit cycles, a standard was set that 90% of lateral cephalograms would display the bodies of CV2, CV3 and CV4 and be of sufficient diagnostic quality to allow CVM staging.[42]

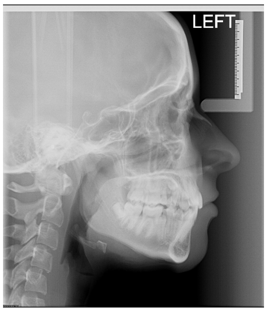


Fig. 4.2.1 CV2-CV4 visible

Fig.4.2.2 CV3 cropped

Fig.4.2.3 CV4 cropped

4.5 Results

One hundred and twenty six lateral cephalograms taken between October 2014 and January 2015 were analysed.

- All the lateral cephalograms displayed the body of the second cervical vertebrae.
- One image did not display the third cervical vertebrae fully.
- In nine images, the body of the fourth cervical vertebrae was cropped (*Figure 4.2.1-3*).
- Overall, 93% of the lateral cephalogram images were of sufficient diagnostic quality to enable CVM staging (*Table 4.1*).

Total images analysed	2nd Cervical Body		3rd Cervical Body		4th Cervical Body		Achieved CVM Staging	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**
126	126	100	125	99	117	93	117	93

Table 4.1: Results of the audit cycle 2014-2015 (*n=number of images, %=percentage of images)*

4.6 Discussion

The outcome of the 2014-2015 audit cycle indicated that 93% of the lateral cephalograms met the agreed standard. As this was the fourth cycle investigating the diagnostic quality of lateral cephalogram images taken in the Liverpool University Dental Hospital the previous audit results are summarised in *Table 4.2*.

Cycle	Year	Image number	% CV2 visible	% CV3 visible	% CV4 visible	% Achieved CVM Staging	Standard met
1	2010-2011	264	100	97	83	83	No
2	2011-2012	134	100	100	93	93	Yes
3	2013-2014	80	100	96	85	85	No
4	2014-2015	126	100	99	93	93	Yes

Table 4.2: Results of the four audit cycles between 2010 and 2015

An overview of the four audit cycles indicates that CV2 was routinely displayed on all lateral cephalogram images taken at LUDH. It was identified that CV3 and/or CV4 were not always visible on these images. Following the results of the first audit cycle an investigation was carried out to ascertain the reasons why CV3 and CV4 were not fully visible on all images. A discussion with the consultant radiologist found that incorrect patient positioning in the cephalostat resulted in the cropping of CV3 and/or CV4. When patients were positioned in the cephalostat with their neck hyperextended and not in natural head position this may result in the cervical bodies of CV4 and CV3 being cropped from the radiographic field. Therefore, incorrect patient positioning was a significant factor in the accuracy of determining the CVM stage with an upward or downward head inclination adversely affecting the ability to evaluate the CVM stage.[163]

Training in correct patient head positioning was provided to the radiographers and after an adequate wash-out period of six months, a 2nd cycle in 2011-2012, was carried out. This cycle showed 93% of images met the standard, emphasising the positive effect of the training received.

The audit cycle was repeated 21 months later in 2013-2014 and the proportion of lateral cephalogram images with CV2-CV4 clearly visible fell to 85%. The 90% target may not have been achieved in this cycle due to the sample size being significantly lower and therefore not representative. However, another potential reason for this variation was highlighted at a meeting with the radiology department where it was suggested that this decline was possibly due to a recent high staff turnover. As Liverpool University Dental Hospital is a teaching hospital, a number of trainee radiographers rotate through the department on a continuous basis. This may have resulted in many of the present radiographers not having received training in correct patient positioning. As this issue is commonly encountered in all teaching hospitals, training about patient positioning in the cephalostat was implemented as an on-going educational programme. Subsequently, poster prompts were displayed in the radiography department reminding staff to position the patient in the natural head position within the cephalostat, to ensure visibility of CV2-CV4.

The fourth cycle was carried out in 2014-2015, after a wash out period of 6 months and the effect of the poster prompts resulted in 93% of the lateral cephalogram images having met the standard of 90% (*Figure 4.3*).

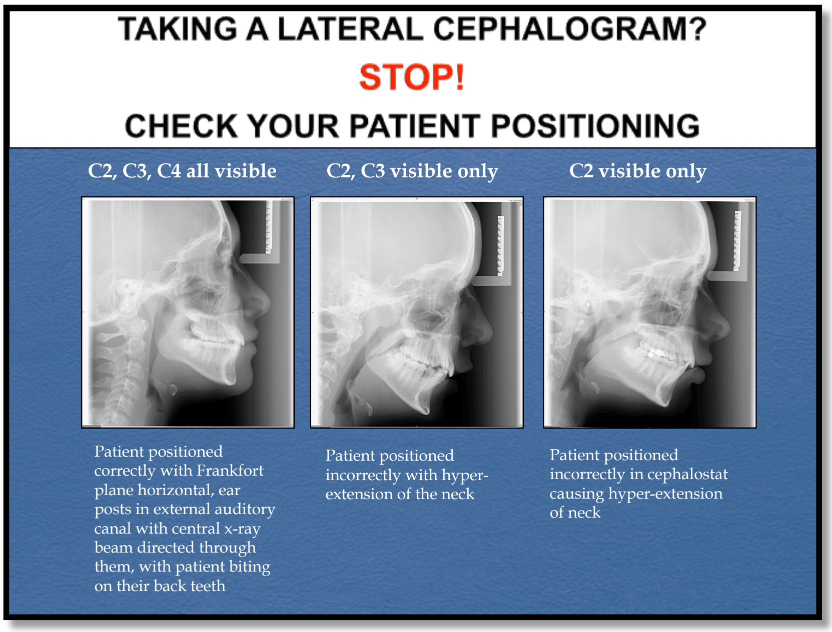


Figure 4.3: The poster prompts displayed in the radiology department, LUDH

4.7 Conclusions

This four-stage audit cycle has been invaluable in identifying the key issues responsible for not obtaining full visibility of CV2 to CV4 that are required for CVM staging. The target, that 90% of lateral cephalograms had the 2nd, 3rd and 4th cervical vertebrae clearly visible, was met in the second and fourth cycle but not in the first and third cycles. After the issue of incorrect patient positioning was identified, the required changes were implemented. This included the provision of training to the radiographers in correct patient positioning in the cephalostat. This led to the 90% target being achieved in the second audit. The target however, was not maintained and after the 3rd audit cycle results were obtained, measures were taken to ensure the ongoing training of radiographers in the radiology department at LUDH through the development of poster prompts, acting as aide memoirs. These implementations have resulted in the success of this 4th audit cycle, with 93% of the lateral cephalograms achieving the standard.

While frequent monitoring is advisable, particularly in incidences of high staff turnover, the poster prompts have proven to be an effective means of reminding the radiographer of the correct head position of the patient.

4.8 Recommendations

It is recommended to place a poster, in a visible location, next to the cephalostat illustrating the ideal head position to ensure that CV2, 3 and 4 are included on the lateral cephalogram, acting as an useful 'aide memoir'.

It is also recommended that continuous training and education is provided to radiographers taking lateral cephalograms regarding correct patient positioning.

It is recommended to re-audit the diagnostic quality of lateral cephalograms in 2 years.

Chapter 5: Study Aims and Outcome measures

5.1 Aims

5.1.1 Primary

1. To determine the reliability of CVM stage determination, using both full and cropped lateral cephalometric radiographs, amongst trainee orthodontists and specialist orthodontists.
2. To determine whether cropped images affect the reliability of the CVM staging method.

5.1.2 Secondary

1. To determine the precision of the assessments made
    a. By individual observers
    b. Of individual images.

5.2 Outcome Measures

5.2.1 Primary outcome measures

The intra–observer and inter-observer reliability of CVM stage determination among a group of orthodontic clinicians for the full and cropped lateral cephalogram images.

Chapter 6: Null Hypothesis

6.1: Null Hypothesis

There is no difference in the reliability of staging lateral cephalograms using full images of lateral cephalogram radiographs compared with using cropped images of lateral cephalogram radiographs, against the alternative hypothesis of a difference.

Chapter 7: Methods and materials

7.1 Study Design

7.1.1 Study Type

Reliability studies provide an important contribution to an assessment of the usefulness of an index by providing information about the amount of inherent error to any score or measurement.[158] In the assessment of research validity, it is important to ascertain whether the interpretation of a particular index is a product of chance or a measure of its precision. Reliability and agreement are not fixed properties of an index but are, in fact a product of interactions between the various sources of variability; the index used, the observers, the sample characteristics, the measurement setting and the statistical approach. It is for these reasons that study results are only interpretable when the measurement setting is sufficiently described and the statistical method is fully explained.[145]

This study was conducted as a two-phase reliability study of the modified CVM staging method as described by Baccetti, Franchi and McNamara.[28] This CVM diagnostic index was applied by a group of orthodontic specialists and orthodontists in training, to a consecutive sample of digital lateral cephalogram radiographs each of which were viewed as a full image and a cropped image.

7.2 Study Setting

The study was executed at two consecutive meetings of the Mersey and North Wales Audit Group between January and July 2016. The rationale for choosing the audit meetings as the forum for carrying out this study was that it provided a convenient setting where all orthodontic consultants, junior and senior trainee orthodontists gathered quarterly and where the required audio, visual and computer equipment for CVM training and image projection were readily available.

7.3 The Sample

7.3.1 Recruited Participants as Observers

At the preceding audit meeting of the Mersey and North Wales Audit Group all orthodontic specialists and orthodontic trainees in attendance were presented with an outline of the study's protocol and invited to participate as observers in this two-phase reliability study.

It was accepted that each recruited observer possessed different cognitive, visual and perceptual capabilities and were an important source of measurement error.[147] In order for the study sample to be generalisable to a population of observers, samples of more than ten observers from several different institutions are required.[164] The aim of this study was to recruit a minimum of 20 observers. This was based on a previous reliability study with a similar number of observers.[42]

All potential participants received an information leaflet prior to the distribution of a consent form (Appendix 1 and 2). A signed consent form from the recruited observer was required in order to indicate their voluntary participation in the study and allow their answers to be used as data for the research project. A total of 25 observers were recruited, with 22 observers completing both phase 1 and phase 2 of the study.

The authors of this study (JEH and JM) participated in the CVM rating process for calibration and to ensure a range of 'knowledge' base and experience.

7.3.2 Lateral Cephalogram Image Sample

7.3.2.1 Sample Size

The sample size calculation for this study was based on a previous CVM reliability study by Rainey *et al*.[42]

Weighted kappa, k^w is the most appropriate statistic to apply for the assessment of observer agreement when data are categorical and measured on an ordinal scale. The valid use of weighted kappa, k^w, the chance-corrected agreement measure, requires a sample size of $N \geq 2k^2$.[165] Therefore the minimum number of lateral cephalograms in the image sample required for the valid use of weighted kappa, k^w, was approximated by the sample size equation, $2k^2$, where k is the number of categories in the rating scale. The modified CVM rating scale used in this study has 6 different categories, ranging from CS1 to CS6.[28] This gave a minimum sample size requirement of $2(6^2)$, or, 72 radiographs. As the aim was to determine the reliability of CVM stage determination, using both full and cropped lateral cephalometric radiographs, 72 full radiographic images and 72 manipulated or cropped radiographic images were deemed to be required using this formula, $2k^2$. Therefore the total sample size requirement for this reliability study was determined to be 144 lateral cephalogram images.

7.3.2.2 Sample Frame

The sample frame used for this study included patients who had been previously assessed by an Orthodontic consultant and deemed appropriate for treatment in a hospital department under the National Health Service (NHS). The patients were identified from the patient databases of first year (St1) and post-CCST (St4) Orthodontic Specialty Registrars (StR). The sample of images was generated from patients who had had a lateral cephalogram taken as part of their pre-treatment records and who had commenced treatment from the non-surgical orthodontic patient waiting list between October 2014 and January 2015. Patient case notes were accessed for the purpose of determining their eligibility and to ensure that they satisfied the study's inclusion and exclusion criteria. Once a lateral cephalogram was deemed suitable, based on the defined criteria, it was identified and exported from the Royal Liverpool and Broadgreen University Hospitals NHS Trust Picture Archiving and Communication System (PACS) to an anonymised image database and given a unique numerical identification code and stored on a password protected NHS computer. The image sample generated was compiled from the first 72 consecutive lateral cephalograms fulfilling the inclusion/exclusion criteria and no further image collection was performed.

7.3.2.3 Sample magnification

As the same cephalostat located in the Liverpool University Dental Hospital was used to take all the lateral cephalograms used in this study, no correction for magnification was necessary.

7.3.2.4 Cropping

Cropped images of the consecutive sample of full lateral cephalograms were generated using *PowerPoint*™. The full lateral cephalogram image was cropped to include only the 2^{nd}, 3^{rd} and 4^{th} cervical vertebrae thus eliminating any additional information such as the development stage of the dentition that might bias the observer about the cervical maturation stage. Once generated, each cropped image was then enhanced to the same standarised size. In order to exclude the introduction of tracing error and to mimic clinical conditions the cropped format was not traced. Each image was saved in full and cropped format. The final image sample comprised of 72 lateral cephalogram images in the full format and 72 images, derived from the full lateral cephalograms, in the cropped format.

7.3.2.5 Randomisation Process

In order to limit selection and performance bias and to ensure random assignment of the 144 full and cropped images, a web-based randomising programmer, *RANDOM.ORG*, was used on two separate occasions to generate random numbers between 1 and 144. This system was utilised for the random ordering the 144 images for both phase 1 and phase 2 of the reliability study.

7.3.2.6 Presentation of Images

The sample of full and cropped lateral cephalogram images was saved as a series of high-resolution images in a *JPEG* format to maintain the original radiographic quality and then presented in a random order using a *PowerPoint* ™ presentation. With the advent of digital radiography, the clinician now routinely views patients' images on an individual computer monitor positioned at the chairside. The aim of this study was to replicate the clinical setting by uploading the *PowerPoint*™ presentation on individual computer monitors for the participants to view each image.

7.3.2.7 Timing and Breaks
Each image was displayed for 20 seconds. In order to minimise assessment variability due to operator fatigue, a rest period of 5 seconds was given after every six images. Two ten-minute breaks were incorporated into the timed presentation after image number 48 and 96.[146]

7.4 Radiographic exposure
All patients underwent radiographic exposure as part of their orthodontic treatment and in line with normal clinical practice. There was no additional radiographic exposure. The Liverpool University Dental Hospital Consultant Radiologist, Mr. Paul Nixon, confirmed the radiographs for the purpose of the study.

7.5 Inclusion/Exclusion criteria
7.5.1 Inclusion criteria
Lateral cephalogram images were included if the patient:

- Was below the age of 18 years, irrespective of gender.
- Was in good general health with an absence of endocrine disorders or nutritional deficiencies.
- Had no previous orthodontic treatment.
- Had no history of cervical trauma.
- Was due to commence orthodontic treatment with orthodontic St1 and St4 trainees in the academic year 2014-2015 and if,
- The lateral cephalogram had complete visualisation of cervical vertebrae C2, C3, C4.

7.5.2 Exclusion criteria
Lateral cephalogram images were excluded if the patient:

- Was over 18 years of age when the start records were obtained.

- Had poor general health, an endocrine disorder or a nutritional deficiency.
- Had previous orthodontic treatment.
- Had been diagnosed with any congenital clefts of the lip and/or palate, or known or suspected craniofacial syndromes or growth related conditions.
- Had a history of cervical trauma.
- Required orthognathic surgery and if
- The lateral cephalogram radiograph was of an unsuitable quality or did not clearly display cervical vertebrae C2, C3 and C4.

7.6 Reliability study

7.6.1 Orthodontic trainees-Phase 1

Eleven orthodontic trainees were recruited and agreed to participate as observers in this reliability study. The first stage commenced in January 2016. Each observer received training in the visual assessment of CVM staging using a reference standard, as described by Baccetti *et al.,*[28] prior to being asked to assess the CVM stage of each image displayed on the computer monitor. The co-author of the modified CVM index, Dr. James McNamara, provided the CVM training presentation material (Appendix 3). This training format described diagrammatically, the morphological changes associated with the 2nd, 3rd and 4th cervical bodies during the maturation process. This was further demonstrated using a cropped radiographic format in which each CVM stage was related to mandibular growth. A simple method of remembering each CVM stage was introduced to the participants. Training was concluded with a final review of each CVM stage in a radiographic format. Prior to the commencement of the reliability study, a group calibration exercise was executed to allow the observers to apply their knowledge ensuring they were confident with using this staging method and to allow an opportunity to clarify any queries with the research team. Laminated reference material, diagrammatically summarising each morphological change associated with CV2, CV3 and CV4, was provided at each individual workstation for consideration throughout the reliability study (Appendix 4). Immediately following the training presentation and group calibration exercise, the observers, once stationed at their individual computer monitors, were shown the randomised and timed images in both the full and cropped format. Each image was displayed for 20 seconds in which time the observer recorded the CVM stage that they felt best described the image, on the score sheet provided (Appendix 5). Two 10-minute breaks (one after image 48 and the other after image 96) were provided to counteract the effects of fatigue.

7.6.2 Orthodontic trainees-Phase 2

The same eleven orthodontic trainees carried out the second phase of the reliability study three months later in April 2016. This three month period between phase 1 and phase 2 was agreed on for two reasons; firstly, this three month period was set as an appropriate 'washout period' to minimise any memory or carry-over effect and secondly, it coincided with when the next regional audit meeting was due to take place. Re-training in CVM staging and group calibration was provided to the observers in the exact same manner and under the same conditions as in phase 1. The images were presented in a different randomised order from the first phase but all the timings remained the same. Reference material for consideration during the reliability study was again provided to each observer.

7.6.3 Orthodontic specialists-Phase 1

Fourteen orthodontic specialists and consultants were recruited and agreed to participate as observers in the reliability study. They carried out their first stage in April 2016. Training and calibration was executed in the exact same manner as for the trainees. Images were displayed in a timed and randomised order on individual computer monitors. Reference material was provided for consideration throughout the reliability study.

7.6.4 Orthodontic specialists-Phase 2

Eleven of the original fourteen orthodontic specialists recruited attended for the second phase of the reliability study in July 2016. The exact same training and calibration exercise was provided before a different randomised order of images was shown to the observers. The same reference material as before was provided for consideration throughout the timed presentation.

7.7 Statistics

Statistical support was sought from Dr. G. Burnside, Lecturer in medical statistics, University of Liverpool.

For reasons outlined previously, both linear and quadratic weighted kappa were used to determine the intra-observer reliability. The Fleiss-Nee-Landis kappa and the Berry Mielke statistic was used to calculate the inter-observer reliability. Percentage agreement was also calculated.

7.8 Regulatory issues

7.8.1 Ethical approval

Ethical approval was obtained from the London Queen Square Research Ethics Committee with the REC reference 15/LO/160 based on protocol number RD&I 5061; Amendment version 2 and IRAS project identification 174153. This was granted on 9th November 2015 (Appendix 6).

Sponsorship from the Royal Liverpool and Broadgreen University Hospitals NHS Trust was obtained before accepting participants into the study (Appendix 7). Annual progress reports were and a final report of the study will be submitted to the sponsor and the REC within the timelines stipulated by the Trust and REC regulations.

7.8.2 Access to study data and documentation

The Investigators permitted study-related monitoring, audits, REC review and regulatory inspections and direct access to source data and other documents would be provided if requested.

7.8.2.1 Confidentiality and data handling

The Chief Investigator, registered under the Data Protection Act, preserved the confidentiality of all participants taking part in the study and acted as the custodian for the study data. All radiographic images used were anonymised and all data pertaining to participants were designated a unique study identification number. All this data were stored on a password-protected NHS computer in the Investigator's office in a secured room on the fifth floor of the Liverpool University Dental Hospital. All study data were stored and archived in compliance with the Research Governance Framework 2nd Edition 2005 and not limited to the Medicines for Human Use (Clinical Trials) 2004 Act plus its appendices and the Data Protection Act 1998. On conclusion of the study, anonymised data will be stored on a hospital Trust computer, by the Chief Investigator, JEH, and password protected for 5 years.

7.8.2.2 Quality assurance

The Lead Investigator monitored the study to ensure compliance with Good Clinical Practice and scientific integrity. The sponsors maintained the management and oversight of the study.

7.8.3 Indemnity

The Royal Liverpool and Broadgreen University Hospitals NHS Trust provided Indemnity and insurance cover with NHSLA, NHS Indemnity Arrangements for clinical negligence claims in the NHS, which was applied to this study.

7.8.4 Sponsor

The Royal Liverpool and Broadgreen University Hospitals NHS Trust was the main Sponsor for this study.

7.9 Consent

This study utilised anonymised radiographic images only, therefore obtaining patients' consent for use of their medical records was not required. Consent was sought and obtained from the observers who agreed to participate in the reliability study after a full explanation about the study had been given, an information leaflet offered and time allowed for consideration. Signed participant consent was obtained (Appendix 2). The right of the participants to refuse to participate, without giving reasons, was respected. All participants were free to withdraw at any time from the study without giving reasons and without prejudice.

7.10 Publication Policy

The intention is that the results of the study will be reported and disseminated at international conferences and in a peer-reviewed scientific journal. The information also forms part of a research thesis submitted in partial fulfillment of a DDSc at the University of Liverpool.

7.11 Financial Aspects

Funding was obtained from the Orthodontic Department DDSc research fund for the procurement of materials required for the fabrication of patient information leaflets, laminated reference cards, consent forms and study information packs.

CHAPTER 8: Results

8.1 Overall intra-observer reliability results

The overall intra-observer agreement for the application of the modified CVM staging index to cropped images was 'substantial' for linear weighted kappa (0.72) with an overall average agreement of 91%. The overall intra-observer agreement for the application of the modified CVM staging index to cropped images was 'almost perfect' for quadratic weighted kappa (0.84) with an overall average agreement of 97%. See Table 8.1.

The overall intra-observer agreement for the application of the modified CVM staging index to full images was 'substantial' for linear weighted kappa (0.71) with an average agreement of 91%. The overall intra-observer agreement for the application of the modified CVM staging index to full images was 'almost perfect' for quadratic weighted kappa (0.84) with an average agreement of 97%.

The P-value for both linear and quadratic weighted kappa was <0.0001, reflecting a "statistically high significance". However, P-values must be interpreted with caution, as it is sensitive to the sample size, with the literature reporting that with a large sample size, any kappa above 0 will become statistically significant. [148]

IMAGE SAMPLE	K^W	Average K^W	Strength of Agreement	Percentage Agreement
CROPPED	Linear	0.72	Substantial	91%
	Quadratic	0.84	Almost perfect	97%
FULL	Linear	0.71	Substantial	91%
	Quadratic	0.84	Almost perfect	97%

$P<0.0001$

Table 8.1: Intra-observer agreement of both cropped and full mages over Phase 1 and Phase 2

8.1.1 Overall intra-observer reliability for cropped and full images

When the overall intra-observer reliability for both cropped and full images were compared the results suggested that there was a minimal difference in the intra-observer reliability with both having a 'substantial to almost perfect' strength of agreement (Cropped; Linear K^w = 0.72, Quadratic K^w = 0.84. Full; Linear K^w = 0.71, Quadratic K^w = 0.84) and 91-97% percentage agreement.

Therefore, from these results it can be inferred that:

- The modified CVM staging method shows 'substantial to almost perfect' intra-observer agreement using both cropped and full images
- The cropping of the lateral cephalogram does not affect the intra-observer reliability of the modified CVM staging method as applied by our sample of 22 orthodontic trainees and orthodontic specialists and the null hypothesis can be accepted.

8.2 Overall inter-observer reliability results

8.2.1 Cropped images

The overall inter-observer reliability of cropped images over the two phases was found to be 'moderate' with the Fleiss-Nee-Landis kappa (Phase 1=0.46, 95% CI 0.45-0.47; Phase 2=0.43, 95% CI 0.43-0.44 *Table* 8.2).

The Berry-Mielke Universal R coefficient of agreement for the overall inter-observer reliability for cropped images was reported using the Berry-Mielke statistic as 'fair' (Phase 1 = 0.31, Phase 2 = 0.23). The overall inter-observer percentage agreement for the cropped images was 85% for Phase 1 and 89% for Phase 2.

8.2.2 Full images

The overall inter-observer reliability of full images over the two phases are reported as 'fair to moderate' with the Fleiss-Nee-Landis kappa (Phase 1=0.44 95% CI 0.44-0.45; Phase 2=0.40 95% CI 0.39-0.41). The Berry-Mielke Universal R coefficient of agreement for the overall inter-observer reliability for full images in both phases was reported as 'fair' (Phase 1=0.29, Phase 2=0.24). The overall percentage agreement for inter-observer reliability for the full images was 85% for phase 1 and 88% for phase 2.

8.2.3 Overall inter-observer reliability for cropped versus full images

The overall inter-observer reliability for both cropped and full images were very similar with the full images having 'moderate' inter-observer reliability for Phase 1 and a 'fair' inter-observer reliability for Phase 2 compared to a 'moderate' reliability for the cropped images over the two phases (Fleiss-Nee-Landis kappa).

The Berry-Mielke kappa found 'fair' inter-rater reliability for both cropped (Phase 1=0.31, Phase 2 = 0.23) and full images (Phase 1 = 0.29, Phase 2 = 0.24).

Regardless of whether a cropped image or a full image was used, the overall inter-observer reliability was better for Phase 1 compared to Phase 2.

From these results it can be inferred that:

- The modified CVM staging method showed 'moderate' inter-observer reliability when applied by our sample of orthodontic trainees and orthodontic specialists (Phase 1 Fleiss-Nee-Landis kappa).
- The modified CVM staging method showed 'fair' inter-observer reliability over the two phases when applied by our sample of orthodontic trainees and orthodontic specialists (Phase 1 and 2 Berry-Mielke Universal R coefficient of agreement).
- As the kappa value for inter-observer agreement was below the threshold of 0.5 this indicates poor agreement of the modified CVM staging method as applied by our sample of 22 orthodontic trainees and orthodontic specialists.[166]
- The cropping of the lateral cephalogram image did not increase the inter-observer reliability of the modified CVM staging method significantly. The inter-observer reliability for cropped images only appeared to show a slightly better strength of agreement of 'moderate' compared to a 'fair' agreement for full images (Phase 2 Fleiss-Nee-Landis Kappa). The null hypothesis can therefore be accepted.

IMAGE SAMPLE	Statistic	Ph1	95% CI	Ph2	95% CI	Strength of Agreement
CROPPED	Fleiss-Nee-Landis kappa	0.46	0.45-0.47	0.43	0.43-0.44	Moderate
	Berry-Mielke	0.31	-	0.23	-	Fair
	Percentage Agreement	85	-	89	-	-
FULL	Fleiss-Nee-Landis kappa	0.44	0.44-0.45	0.40	0.39-0.41	Fair-Moderate
	Berry-Mielke	0.29	-	0.24	-	Fair
	Percentage Agreement	85	-	88	-	-

P<0.0001

Table 8.2: Overall Inter-observer agreement of all cropped and full images over Phase 1 and Phase 2

8.3 Overall reliability of orthodontic trainees

The data were examined to investigate whether a difference in CVM staging reliability was apparent between the cropped and full images when applied by the orthodontic trainees. Firstly, the intra- and inter-observer reliability of the modified CVM staging method as applied by the 11 orthodontic trainees is reported.

8.3.1 Intra-observer reliability of orthodontic trainees using cropped images

The overall intra-observer agreement for the application of the modified CVM staging index to cropped images using both linear and quadratic weighted kappa indicated a 'substantial to almost perfect' strength of agreement (Linear K^w = 0.67-0.87, Quadratic K^w = 0.76-0.95) with overall percentage agreement of 92% and 98% respectively (*Table* 8.3).

8.3.2 Intra-observer reliability of orthodontic trainees using full images

The overall intra-observer agreement of the orthodontic trainee group, for the application of the modified CVM staging index to full images was 'moderate to almost perfect' for linear weighted kappa (0.58-0.84) with an average agreement of 90% and 'substantial to almost perfect' for quadratic weighted kappa (0.75-0.92) with an average agreement of 97%.

The intra-observer reliability for the trainee group showed very similar levels of agreement for both cropped and full images over the two phases (*Table* 8.3).

8.3.3 Trainee intra-observer reliability for cropped versus full images

The intra-observer reliability of orthodontic trainees for cropped and full images only differed slightly when the linear weighted kappa results were compared. The intra-observer reliability amongst the orthodontic trainees using cropped images reported a slightly higher agreement level in linear weighted kappa of 'substantial to almost perfect' agreement (0.67-0.87) compared to the linear weighted kappa of 'moderate to almost perfect' agreement (0.58-0.84) for full images (*Table* 8.3). However, this difference wasn't statistically significant.

8.3.3 Orthodontic trainee intra-observer reliability for cropped and full images

IMAGE SAMPLE	**K^W**	**Range**	**Strength of Agreement**	**Percentage Agreement**
CROPPED	Linear	0.67-0.87	Substantial-Almost perfect	92%
	Quadratic	0.76-0.95	Substantial-Almost perfect	98%
FULL	Linear	0.58-0.84	Moderate-Almost perfect	90%
	Quadratic	0.75-0.92	Substantial-Almost perfect	97%

P<0.0001

Table 8.3: Intra-observer reliability of cropped and full images for orthodontic trainees over Phase 1 and 2

8.3.4 Inter-observer reliability of orthodontic trainees using cropped images

The inter-observer reliability as assessed using the Fleiss-Nee-Landis kappa indicated ‘moderate’ agreement (Phase 1=0.5 and Phase 2=0.5) while the Berry-Mielke values of 0.37 for Phase 1 and 0.29 for Phase 2 indicate a ‘fair’ agreement.

The overall percentage agreement for Phase 1 and Phase 2 was 90% (Table 8.4).

8.3.5 Inter-observer reliability of orthodontic trainees using full images

The inter-observer reliability as assessed by the Fleiss-Nee-Landis kappa indicated a ‘moderate’ agreement (Phase 1=0.46 and Phase 2=0.43) while the Berry-Mielke values of 0.31 and 0.35 for Phase 1 and 2 respectively indicate a ‘fair’ agreement (*Table* 8.4).

8.3.6 Orthodontic trainee inter-observer reliability for cropped versus full images

The overall inter-observer reliability of the trainees, assessed using the Fleiss-Nee-Landis, the Berry-Mielke Universal R coefficient of agreement and the overall percentage agreement were found to be similar for both cropped and full images (*Table* 8.4).

IMAGE SAMPLE	Statistic	Phase 1	Phase 2	Strength of Agreement
CROPPED	Fleiss-Nee-Landis kappa	0.5	0.5	Moderate
	Berry-Mielke	0.37	0.29	Fair
	Percentage Agreement	90	90	-
FULL	Fleiss-Nee-Landis kappa	0.46	0.43	Moderate
	Berry-Mielke	0.31	0.35	Fair
	Percentage Agreement	90	89	-

$P<0.0001$

Table 8.4: Inter-observer reliability of cropped and full images for orthodontic trainees over Phase 1 and 2

8.4 Overall reliability of orthodontic specialists

The intra-and inter-observer reliability of the modified CVM staging method as applied by the 11 orthodontic specialists is reported below for both cropped and full images.

8.4.1 Intra-observer reliability of orthodontic specialists using cropped images

The overall intra-observer agreement for the application of the modified CVM staging index to cropped images by the orthodontic specialists indicated a 'moderate to substantial' strength of agreement for linear weighted kappa (0.55-0.76) and a 'substantial to almost perfect' agreement for quadratic weighted kappa (0.65-0.89) with overall percentage agreement of 90% and 97% respectively (*Table* 8.5).

8.4.2 Intra-reliability of orthodontic specialists using full images

The overall intra-observer agreement for the application of the modified CVM staging index to full images by the orthodontic specialists was 'substantial' for linear weighted kappa (0.62-0.79) with an average agreement of 91% and 'substantial to almost perfect' for quadratic weighted kappa (0.75-0.90) with an average agreement of 97% (*Table* 8.5).

8.4.3 Orthodontic specialist intra-observer reliability for cropped versus full images

The intra-observer reliability for the orthodontic specialists showed very similar levels of agreement for both cropped and full images over the two phases.

The intra-observer reliability amongst the orthodontic specialists showed a slightly wider range of agreement level with linear weighted kappa values suggesting 'moderate to substantial' agreement (0.55-0.76) using cropped images compared to the linear weighted kappa of 'substantial' agreement (0.62-0.79) for full images.

Quadratic weighted kappa agreement suggested a 'substantial to almost perfect' level of agreement for both cropped and full images when applied by the orthodontic specialist group (*Table* 8.5)

IMAGE SAMPLE	K^W	Range	Strength of Agreement	Percentage Agreement
CROPPED	Linear	0.55-0.76	Moderate-Substantial	90%
	Quadratic	0.65-0.89	Substantial-Almost perfect	97%
FULL	Linear	0.62-0.79	Substantial	91%
	Quadratic	0.75-0.90	Substantial-Almost perfect	97%

P<0.0001

Table 8.5: Intra-observer reliability of cropped and full images for orthodontic specialists over Phase 1 and 2

8.4.4 Inter-observer reliability of orthodontic specialists using cropped images

The inter-observer reliability, as assessed using the Fleiss-Nee-Landis kappa, indicated ‘moderate’ agreement for Phase 1 (0.44) and a ‘fair’ agreement for Phase 2 (0.39) while the Berry-Mielke values indicate a ‘fair’ agreement for Phase 1 (0.25) and a ‘slight’ agreement for Phase 2 (0.17). This suggests that inter-observer reliability of the orthodontic specialist group using the modified CVM staging index on cropped images was better in Phase 1 compared to Phase 2. The percentage agreement for Phase 1 and Phase 2 was 80% and 88% respectively (*Table* 8.6).

8.4.5 Inter-observer reliability of orthodontic specialists using full images

The inter-observer reliability, as assessed by the Fleiss-Nee-Landis kappa, indicated ‘moderate’ agreement for Phase 1 (0.43) and a ‘fair’ agreement for Phase 2 (0.38) while the Berry-Mielke values suggest ‘fair’ agreement for Phase 1 (0.25) and ‘slight’ agreement for Phase 2 (0.17). This suggests that inter--observer reliability of the orthodontic specialist groups using the modified CVM staging index on full images was better in Phase 1 compared to Phase 2 (*Table* 8.6).

The overall inter-observer reliability of the orthodontic specialist group, as assessed using the Fleiss-Nee-Landis, the Berry-Mielke Universal R coefficient of agreement and the overall percentage agreement were found to be similar for both cropped and full images (*Table* 8.6).

Once again, as the kappa values for inter-observer agreement were below the threshold value of 0.5 this indicates poor inter-observer agreement of the modified CVM index in the staging of full images as applied by our sample of 11 orthodontic specialists.

8.4.6 Orthodontic specialist inter-observer reliability for cropped and full images.

IMAGE SAMPLE	Statistic	Phase 1	Phase 2	Strength of Agreement
CROPPED	Fleiss-Nee-Landis kappa	0.44	0.39	Fair-Moderate
	Berry-Mielke	0.25	0.17	Slight-Fair
	Percentage agreement (%)	80	88	-
FULL	Fleiss-Nee-Landis kappa	0.43	0.38	Fair-Moderate
	Berry-Mielke	0.25	0.17	Slight-Fair
	Percentage agreement (%)	80	87	-

P<0.0001

Table 8.6: Inter-observer reliability of cropped and full images for orthodontic specialists over Phase 1 and 2

8.5 Comparison of agreement between observers with different levels of clinical experience

The influence of clinical experience on the reliability of the modified CVM staging method was explored using our data. The data were examined to investigate whether a difference in the CVM staging reliability was apparent between the different levels of observer experience.

8.5.1 Intra-observer reliability of orthodontic trainees compared with specialists using cropped images

It was apparent that when linear weighted kappa was applied, the level of agreement for the orthodontic trainees was found to be ‘substantial’ to ‘almost perfect’ while the intra-examiner agreement of the orthodontic specialists was ‘moderate’ to ‘substantial’. This indicates a slightly better but greater range of intra-observer agreement for the trainee group in comparison to the specialist group. However, when the quadratic weighted kappa values were considered both the orthodontic trainees and the specialists both had ‘substantial’ to ‘almost perfect’ agreement. The study results, therefore, suggest that the level of clinical experience did not affect the reliability of the CVM staging method when applied to cropped images (*Table* 8.7 and *Graph* 8.1)

8.5.2 Intra-observer reliability of the modified CVM staging index using cropped images

GRADE	Rater	Linear K^W	Agreement	95% CI	Quadratic K^W	Agreement	95% CI
Trainee	1	0.78	Substantial	0.69-0.86	0.90	Almost perfect	0.84-0.96
	2	0.77	Substantial	0.69-0.85	0.88	Almost perfect	0.83-0.94
	3	0.81	Almost perfect	0.73-0.90	0.90	Almost perfect	0.84-0.96
	4	0.87	Almost perfect	0.81-0.93	0.95	Almost perfect	0.93-0.98
	5	0.70	Substantial	0.57-0.82	0.76	Substantial	0.59-0.93
	6	0.67	Substantial	0.55-0.79	0.77	Substantial	0.65-0.90
	7	0.71	Substantial	0.60-0.81	0.82	Almost perfect	0.74-0.91
	8	0.72	Substantial	0.62-0.82	0.84	Almost perfect	0.75-0.93
	9	0.83	Almost perfect	0.75-0.91	0.92	Almost perfect	0.88-0.97
	10	0.78	Substantial	0.71-0.85	0.92	Almost perfect	0.89-0.95
	11	0.67	Substantial	0.58-0.76	0.84	Almost perfect	0.78-0.90

GRADE	Rater	Linear K^W	Agreement	95% CI	Quadratic K^W	Agreement	95% CI
Specialist	**1**	0.65	Substantial	0.54-0.71	0.78	Substantial	0.66-0.90
	2	0.73	Substantial	0.63-0.83	0.83	Almost perfect	0.73-0.93
	3	0.75	Substantial	0.66-0.83	0.89	Almost perfect	0.83-0.95
	4	0.68	Substantial	0.56-0.80	0.77	Substantial	0.63-0.91
	5	0.73	Substantial	0.63-0.84	0.83	Almost perfect	0.74-0.92
	6	0.69	Substantial	0.58-0.80	0.81	Almost perfect	0.71-0.91
	7	0.69	Substantial	0.58-0.81	0.80	Substantial	0.68-0.92
	8	0.76	Substantial	0.67-0.84	0.88	Almost perfect	0.83-0.94
	9	0.68	Substantial	0.57-0.79	0.83	Almost perfect	0.73-0.93
	10	0.72	Substantial	0.61-0.83	0.83	Almost perfect	0.73-0.93
	11	0.55	Moderate	0.39-0.70	0.65	Substantial	0.48-0.83

P<0.0001

Table 8.7: Intra-observer reliability of Orthodontic Trainees and Specialists using cropped images Phase 1and Phase 2

8.5.3 Intra-observer reliability of orthodontic trainees and specialists using the modified CVM staging index to stage full images

GRADE	Rater	Linear K^W	Agreement	95% CI	Quadratic K^W	Agreement	95% CI
Trainee	1	0.84	Almost perfect	0.76-0.92	0.92	Almost perfect	0.85-0.98
	2	0.75	Substantial	0.66-0.83	0.89	Almost perfect	0.84-0.94
	3	0.72	Substantial	0.60-0.84	0.80	Substantial	0.68-0.92
	4	0.70	Substantial	0.60-0.80	0.84	Almost perfect	0.76-0.92
	5	0.69	Substantial	0.57-0.81	0.80	Substantial	0.67-0.92
	6	0.62	Substantial	0.50-0.74	0.77	Substantial	0.66-0.87
	7	0.74	Substantial	0.65-0.83	0.88	Almost perfect	0.82-0.94
	8	0.80	Substantial	0.71-0.89	0.90	Almost perfect	0.83-0.96
	9	0.76	Substantial	0.66-0.85	0.87	Almost perfect	0.79-0.94
	10	0.75	Substantial	0.65-0.85	0.85	Almost perfect	0.75-0.96
	11	0.58	Moderate	0.46-0.69	0.75	Substantial	0.64-0.86

GRADE	Rater	Linear K^W	Agreement	95% CI	Quadratic K^W	Agreement	95% CI
Specialist	1	0.64	Substantial	0.53-0.76	0.78	Substantial	0.66-0.90
	2	0.73	Substantial	0.59-0.86	0.75	Substantial	0.53-0.96
	3	0.72	Substantial	0.62-0.83	0.87	Almost perfect	0.80-0.94
	4	0.79	Substantial	0.70-0.88	0.90	Almost perfect	0.85-0.95
	5	0.76	Substantial	0.67-0.84	0.87	Almost perfect	0.80-0.94
	6	0.73	Substantial	0.63-0.83	0.86	Almost perfect	0.79-0.94
	7	0.62	Substantial	0.51-0.73	0.80	Substantial	0.71-0.89
	8	0.71	Substantial	0.61-0.81	0.85	Almost perfect	0.78-0.92
	9	0.66	Substantial	0.54-0.77	0.80	Substantial	0.68-0.92
	10	0.67	Substantial	0.55-0.78	0.81	Almost perfect	0.72-0.91
	11	0.55	Moderate	0.39-0.70	0.65	Substantial	0.48-0.83

P<0.0001

Table 8.8: Intra-observer reliability of Orthodontic Trainees and Specialists using full images Phase 1 and Phase 2

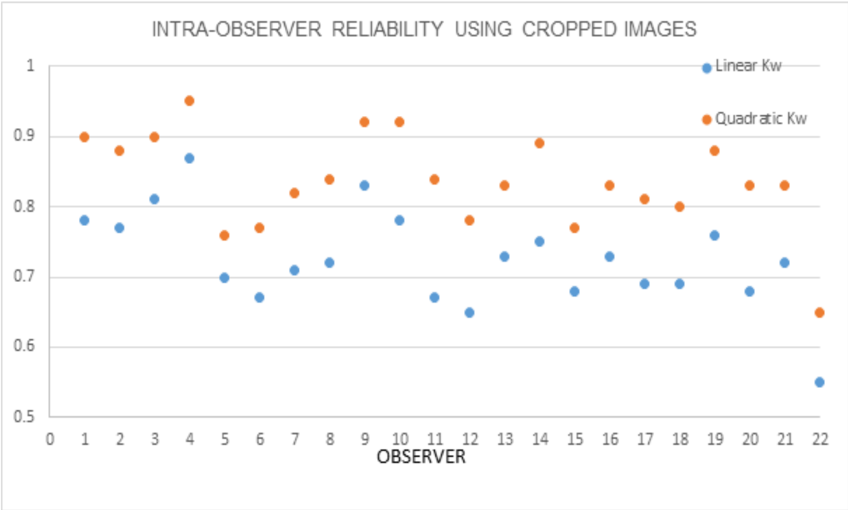
8.5.4 Intra-observer reliability of orthodontic trainees compared with specialists using full images

Orthodontic trainees showed a wide range of agreement from ‘moderate to almost perfect’ intra-observer reliability using linear weight kappa when the modified CVM staging index is applied to full lateral cephalogram images. When quadratic weighted kappa was used the range of agreement was ‘substantial to almost perfect’.

The intra-observer reliability of the orthodontic specialist group showed ‘substantial’ agreement using linear weight kappa when the modified CVM staging index is applied to full lateral cephalogram images. When quadratic weighted kappa was used the range of agreement was ‘substantial to almost perfect’.

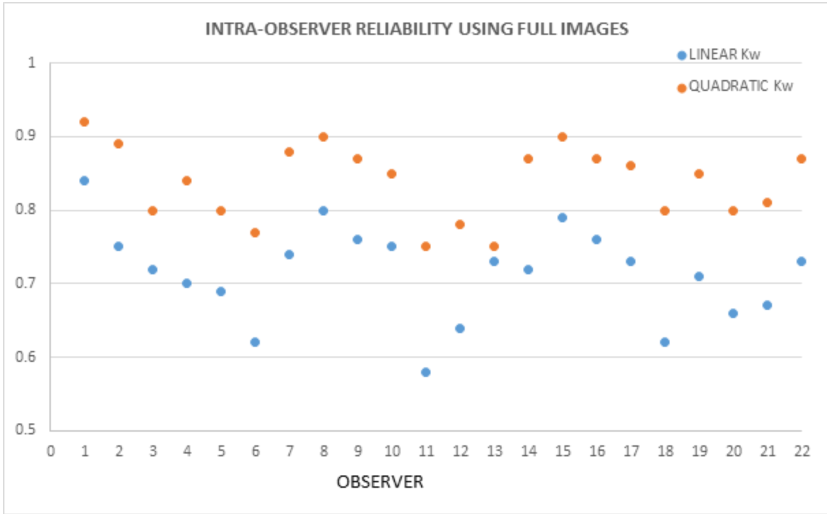
Based on the quadratic weighted kappa findings, the level of clinical experience does not affect the reliability of the CVM staging method when applied to full images (*Table* 8.8 and *Graph* 8.2).

8.5.5 Linear and quadratic weighted kappa intra-observer reliability for trainees and specialists of the modified CVM staging index using **CROPPED** images



Graph 8.1: Linear and Quadratic K^W values for Orthodontic trainees and specialists' intra-observer reliability using cropped images. Observers 1-11 Trainees; 12-22 Specialists

8.5.6 Linear and quadratic weighted kappa intra-observer reliability for trainees and specialists of the modified CVM staging index using **FULL** images



Graph 8.2: Linear and Quadratic K^W values for Orthodontic trainees and specialists' intra-observer reliability using full images. Observers 1-11 Trainees; 12-22 Specialists

8.6 Observer outliers

The data were analysed to examine whether there were any outliers in both groups for cropped and full images using both linear and quadratic weighted kappa (Appendix 9, 10, 11, 12, 13, 14, 15 and 16). Overall, the narrow range of weighted kappa values indicated a limited variability (*Table* 8.9).

	Image	Linear K^W	Strength of Agreement	Range
Phase 1	Cropped	0.67	Substantial	0.62-0.72
	Full	0.67	Substantial	0.58-0.72
Phase 2	Cropped	0.65	Substantial	0.49-0.7
	Full	0.63	Substantial	0.56-0.69

Table 8.9: Overall linear K^w reliability scores of Trainees versus Specialists for cropped and full images in Phase 1 and Phase 2

Overall the average linear weighted kappa score when the Trainees were evaluated against the Specialists for intra-observer reliability was 0.67 for both cropped and full images in Phase 1. The cropped images had a marginally narrower range 0.62-0.72 compared to the full images 0.58-0.72. There was no obvious outlier to account for these findings.

Surprisingly, intra-observer reliability decreased from Phase 1 to Phase 2 with the average linear weighted kappa score for cropped images in Phase 2 reported as 0.65 with a large range of 0.49-0.7. The lower end of this range lies in the 'moderate' strength of agreement with specialist number 11 contributing to this low value with a linear kappa value of 0.55 reported (*Table* 8.7).

Full images also reported a slightly lower linear weighted kappa value of 0.63

but had a narrower range of values 0.56-0.69, which also accounted for the reliability reducing in to the lower category of 'moderate' agreement with trainee number 11 contributing to this low value with a linear kappa value of 0.58 (*Table* 8.8).

	Image	Quadratic K^W	Strength of Agreement	Range
Phase 1	Cropped	0.81	Almost perfect	0.75-0.85
	Full	0.82	Almost perfect	0.76-0.86
Phase 2	Cropped	0.79	Substantial	0.62-0.83
	Full	0.78	Substantial	0.73-0.83

Table 8.10: Overall quadratic K^w reliability scores of Trainees versus Specialists for cropped and full images in Phase 1 and Phase 2

The trend seen in Table 8.9 is repeated in that the quadratic weighted kappa scores of Trainees and Specialists with cropped and full images in Phase 1 achieved 'almost perfect' intra-observer agreement compared to a 'substantial' agreement in Phase 2. Not only did the reliability in CVM staging method decrease in Phase 2 but a wider range of kappa scores is noted. As outlined above, specialist number 11 and trainee number 11 are the obvious outliers partially accounting for these low values (*Table* 8.10).

8.7: Summary of results

- The modified CVM staging method had 'substantial to almost perfect' intra-observer agreement using both cropped and full images.
- Cropping the lateral cephalogram did not affect the intra-observer reliability of the modified CVM staging method as applied by our sample

of 22 orthodontic trainees and orthodontic specialists.

- The modified CVM staging method had ‘fair’ to ‘moderate’ inter-observer reliability of the modified CVM method when applied by our sample of orthodontic trainees and orthodontic specialists.
- The inter-observer reliability for cropped images showed a slightly better strength of agreement of ‘moderate’ compared to a ‘fair’ agreement for full images, however the kappa values were not significantly different, therefore, cropping the lateral cephalogram image did not reduce the inter-observer reliability of the modified CVM staging method .
- The intra-observer reliability for the orthodontic trainee group showed very similar levels of agreement for both cropped and full images over the two phases with linear weighted kappa reporting a ‘moderate’ to ‘almost perfect’ level of agreement and quadratic weighted kappa reporting a ‘substantial’ to ‘almost perfect’ level of agreement.
- The overall inter-observer reliability of the orthodontic trainee group reported a ‘fair’ to ‘moderate’ level of agreement of both cropped and full images.
- The intra-observer reliability for the orthodontic specialist group showed very similar levels of agreement for both cropped and full images over the two phases.
- The overall inter-observer reliability of the orthodontic specialist group suggested a ‘slight’ to ‘moderate’ level of agreement for both cropped and full images.
- When the quadratic weighted kappa values were considered, both the orthodontic trainees and the specialists had ‘substantial to almost perfect’ intra-observer agreement using both cropped and full images.

9. Discussion

9.1 Overview

The typical craniofacial growth pattern follows an orderly and predictable pattern but is characterised by a wide individual variability in both the amount and rate of growth.[62, 89] This variability is innately related to genetic, gender and ethnic differences.[11] Determining periods of accelerated growth remains an important consideration for the optimal timing of orthodontic treatment, however, obtaining an accurate method of assessing an individual's stage of growth has proven difficult.[10] Several physiological indicators have been proposed to aid the evaluation of skeletal maturity.[3, 13, 15, 19, 21, 28, 29, 71, 73, 77, 79, 81-83, 92, 94, 97, 100] Determining skeletal age using CVM staging has gained popularity in recent times as the second to fourth cervical vertebrae are visible on a lateral cephalogram, an image that is routinely taken for treatment planning purposes[135] thus avoiding additional radiographs. A lack of methodological standardisation associated with much of the previous studies on CVM staging reliability have contributed to the uncertainty surrounding CVM staging reliability.[27] Rainey *et al*. addressed these methodological flaws and concluded that CVM staging showed 'substantial' intra- and inter-observer reliability when full lateral cephalograms were assessed. However, visibility of the developing dentition may potentially influence CVM staging and introduce bias. To date, no previous research has investigated whether CVM staging reliability is affected by using cropped images. The results of this research study have shown that cropping lateral cephalograms to remove the potential influence of visibility of the dentition, does not affect CVM staging reliability. This study also confirms previous research that the level of orthodontic experience does not affect reliability of CVM staging.[40]

The results of this study are discussed in further detail below and emphasise quadratic kappa weightings. This weighting was chosen as it is a more accurate reflection of clinical practice. For example, the clinical consequences are greater if the observer reports a CVM staging difference of three categories rather than one category, i.e. CVM stage 1 reported instead of CVM stage 3. Quadratic weightings penalises more for each additional category or CVM staging by which the observer is out, while with linear weightings the penalties are consistent (Tables 5.2 and 5.3).[153] Therefore its use has been recommended for its practical interpretation[153, 156] An extensive search of the literature on CVM staging reliability has revealed numerous methods of assessing CVM staging.[3, 4, 26, 29, 38, 143] The results of this study will initially be compared to studies[28, 33-36, 39, 40, 42, 78, 109, 131, 137, 141] that have used the modified CVM staging index as described by Baccetti *et al*.[28]

9.2 Overall intra-observer reliability results

9.2.1 Overall intra-observer reliability for cropped and full images

This study reports that the overall intra-observer reliability for the application of the modified CVM staging index as applied by 22 orthodontic clinicians to a contemporary sample of images of cropped and full images were 'almost perfect' with a quadratic K^w of 0.84 and an average percentage agreement of 97%. Linear K^w values reflected a 'substantial' reliability with the reported cropped image value of 0.72 and full image value of 0.71, percentage agreement 91%. The P-value for both linear and quadratic weighted kappa using cropped and full images was <0.0001, reflecting a statistical high significance (*Table* 8.1).

When the overall intra-observer reliability for both cropped and full images are compared, it indicates that there is no difference in the intra-observer reliability with both reporting a 'substantial to almost perfect' strength of agreement, depending on which weighted kappa was used. To my knowledge there is no previous study which compares CVM intra-observer reliability using cropped and full lateral cephalogram images. The findings of this study demonstrate that the reliability of the modified CVM staging method for the identification of the peak mandibular growth was not improved by using cropped images.

9.2.2 Comparison with other intra-observer reliability studies using cropped images

The overall intra-observer reliability of the modified CVM staging method using cropped images was 'almost perfect' (quadratic K^w 0.84) with an overall average agreement of 97%. This is an acceptable level of intra-observer reliability to recommend the use of the modified CVM staging index (*Table* 8.1).[28]

These results can be compared to a number of studies which used cropped images to investigate the reliability of the modified CVM staging index[22, 33, 40, 41, 78, 131, 141] and these are summarised in *Table* 9.1. While the majority of studies assessing CVM staging reliability has been conducted using cropped images, direct comparisons cannot be made due to the large variability in methodologies, particularly, differences in statistical approaches, observer number, sample size, and training in CVM staging.

Author	Image number	Observer number	Traced CV	Statistic	Intra-observer reliability	Bias risk
Gabriel[33]	90	10	No	K^w	0.36-0.79	↑
Nestman[41]	90	10	No	Percentage agreement	50-88.2%	↑
Alkhal[35]	25	1	NR	Percentage agreement	96%	↑
Wong[39]	25	1	NR	Percentage agreement	92%	↑
Ball[109]	72	1	Yes	Kappa	0.943-1.0	↑
Rongo[40]	50	30	NR	Linear K^w	0.24-0.81	↑
Perinetti[137]	72	10	Yes	Linear K^w	0.52-0.90	↑
Predko-Engel[141]	50	10	No	K^w	0.44	↑
Engel[131]	29	5	No	Kappa	0.36	↑
Gray[78]	25	1	Yes	Percentage agreement Kappa K^w	84% 0.80 0.89	↑

Table 9.1: Summary of reported intra-observer reliability of the modified CVM method[28] using cropped images.

Key: NR=not reported; CV=cervical vertebrae; K^w=weighted kappa

Gabriel *et al*. cited methodological flaws as a possible cause of inflated levels of reproducibility associated with CVM staging, specifically the use of traced cervical vertebrae which may introduce bias. He also argued that using observers with a 'research level' of understanding[33] may also contribute to exaggerated reliability results. Concerns were expressed regarding the use of small sample sizes, questionable randomisation of image samples and inappropriate statistical methods to determine reliability. Many comparisons can be made between the methodologies used in Gabriel *et al.* and this study such as cropping the lateral cephalograms to facilitate complete visualisation of the cervical vertebrae 1 to 4 only so that additional information which may bias the observer, was eliminated. The cropped images were not traced, again to avoid influencing the observer during the staging process. Similarly, training in the CVM staging method was provided prior to the reliability test using randomised images. In Gabriel *et al.,* the intra-observer reliability was evaluated using weighted kappa, however, the author didn't clarify which weighting was used. The intra-observer reliability for cropped images was reported as a kappa coefficient range between 0.36-0.79 with an average percentage agreement of 62%. The authors reported this as 'moderate agreement' and concluded that the modified CVM staging method was too variable to be used for the assessment of skeletal maturity. The difference in intra-observer reliability reported between Gabriel *et al*. and this study can be possibly explained by the differences in the sample size, sample frame and the number and type of observers used. Gabriel *et al.* used a historical sample of images and no sample size calculation was carried out. In addition, only ten observers were recruited and these were specialist practitioners, while this study recruited twenty-two observers which included orthodontic trainees, hospital based orthodontic specialists and specialist practitioners. Also, the training received in the CVM staging method may also have contributed to the difference in intra-observer reliability. In the Gabriel *et al.* study training consisted of the observers receiving a hard copy handout of a schematic representation of the individual stages without any further explanation provided. This form of training was never recommended by the developing authors as a guideline for the implementation of the CVM staging method in clinical practice.[167] This study used training material from one of the original authors of the modified CVM method.[28] Also, Gabriel *et al*. reported moderate intra-observer reliability, however, no scale for the interpretation of weighted kappa coefficient was referenced. If the Landis and Koch scale was used then these values of 0.36-0.79 would equate to 'poor-substantial' reliability (*Table* 2.9).[147] It would appear that the author based his conclusion primarily on percentage agreement, which is an inappropriate statistical method.[149, 150]

Nestman *et al.*[41] expanded on the study by Gabriel *et al*. [33] in the effort to determine why reliability of the CVM staging method was poor. In this study, the reliability study was replicated using the same number of observers and historical sample of images. Again the type of weighted kappa used to assess intra-observer reliability was not stated and the kappa coefficient was calculated by comparing CVM staging from the study by Gabriel *et al*.[33] and Nestman *et al.*[41] giving an agreement of 50-88.2%. The author cites similar findings from a study by Kucukkeles *et al*.[38] that reported intra-observer reliability of three observers of 90%, 65% and 45%, respectively. However, these results are not comparable as reliability is not a fixed property of an index but rather a product of interactions between the various sources of variability such as the index used, the observers, the sample characteristics and the statistical approach used.[145] While Kucukkeles *et al*.[38] used cropped images, these were traced. A different CVM staging method[29] was used and a cross sectional sample of images from a Turkish population was assessed which differed from the longitudinal sample of images used by Gabriel and Nestman.[33, 41] In the study by Kucukkeles *et al*.[38] only three observers were used. A minimum of ten observers are required to increase the generalisability.[164]

Alkhal *et al*. reported an intra-observer reliability of 96% when the modified CVM staging method was applied to a small sample of twenty five lateral cephalograms randomly selected from a large cross-sectional sample from a Southern Chinese population, three weeks apart.[35] Unfortunately, the author did not report whether training was received or whether cropped or full lateral cephalograms were used. Only one observer was used to determine the intra-observer reliability and this was reported as the principle investigator, therefore, it can be assumed that the observer had a 'research level' of experience in CVM staging, introducing a potential bias in to the results.[35] There was no clarification on the randomisation process used or justification as to why such a small sample of twenty five images were selected for CVM staging. A study by Wong *et al*.[39] based on the same sample as Alkhal *et al*. was subjected to the same limitations, reported an intra-observer reliability of 92%. As both these authors only reported percentage agreement and not the chance corrected kappa statistic, these findings must be interpreted with caution.

A recent study reported an intra-observer reliability of the modified CVM staging method between kappa values 0.943 and 1.0, an 'almost perfect' agreement, using a sample of male subjects from the longitudinal records of the Burlington Growth centre.[109] The author and sole observer was calibrated by the original authors of the modified CVM method twice before staging the sample of seventy two cropped lateral cephalograms, with a two month interval. However, these findings must be interpreted with caution as a historical sample of images was used, only one observer with a 'research level' of CVM understanding was involved in the staging process and the type of kappa used was not stated.

A study investigating the intra-observer reliability of CVM staging reported a linear weighted kappa value of 0.52-0.90, using seventy two images and ten observers with a 4 to 6 week interval between phases.[137] These findings are very similar to the values recorded in this study with a mean linear weighted kappa value of 0.72 reported (range 0.58-0.87). Other similarities with this study and Perinetti *et al*. include the use of observers with a broad range of clinical experience, trained in CVM staging prior to undertaking the reliability test. However, the results of this study are more generalisable as the cervical vertebrae were not traced and low quality images were not excluded therefore, avoiding potential bias. In another study evaluating intra-observer reliability of CVM staging using cropped images, a linear weighted kappa range of 0.24-0.81 was reported, along with an average percentage agreement of 26-78%.[40] The wide range of values reported was explained by the author as being a result of the more junior members of the rating panel having more knowledge of CVM staging as opposed to the specialist orthodontists who did not use CVM staging on a daily basis. If the results from the specialist orthodontists were eliminated when the kappa values would be similar to the linear kappa weighting reported in this study. However, while the study by Rongo *et al*. did not carry out a sample size calculation, the fifty images evaluated were randomly ordered before each reliability phase, with a three week interval stated. Similar to this study, the sample of lateral cephalograms used were of patients attending an orthodontic department for treatment, however, the author did not state whether the images used were traced or not. Due to limitations already mentioned previously the results must be interpreted with caution. A pragmatic reliability study carried out by Predko-Engel *et al*. consisted of ten non-calibrated observers that specifically did not partake in the preparation of the images used for evaluation of the modified CVM staging method.[141] They did not calibrate the observers prior to the staging process as they suggested to do so would deviate from normal daily clinical practice and possibly cause an inflated reliability co-efficient.[141] The images were randomised, however, the author did not describe how this was done. Similar to this study, the images were sourced from their orthodontic department and were cropped and uploaded into a *PowerPoint*™ presentation. The prepared presentation was circulated to the observers, who had to log on to a designated website and rate the images, twice, at least three weeks apart. The mean but unspecified, weighted kappa, was 0.44, which indicated a moderate intra-observer agreement as per the Landis and Koch reference table (*Table* 2.9).[147] However, this result should be interpreted with caution, as it appears this reliability study lacked standardisation across the two phases which is required for the accurate reporting of reliability studies.[145] The web based reliability study lacked control over key conditions such as how the images were viewed, the time allocated to staging each image and the time interval between phases. In addition to not stipulating the type of weighting used, a sample size calculation was not carried out.

A study by Engel *et al.* evaluated CVM reliability using cropped lateral cephalograms of girls from the Nijegen Growth Study.[131] A total of twenty nine randomised images were rated by five calibrated observers on two occasions, at least four weeks apart. The authors did not specify the kappa used but reported a poor intra-observer reliability of 0.36, ranging from the lowest value of 0.18 to the highest value of 0.54. Again, this result must be interpreted with caution due to methodological flaws as previous addressed, such as, lack of sample size calculation and failure to give an account of the randomisation process used. Also of concern, was the lack of standardization in the rating process, as the observers rated the images at a time convenient to them with no time restriction and no control over the environment in which they rated the images. Without strict implementation of consistent conditions during the two occasions the images were rated, intra-observer reliability values can be affected.[145] However this may reflect "real life" situations more accurately.

Gray *et al*. evaluated the reliability of CVM staging using a quantitative method utilising multiple semilandmarks to discriminate the morphological changes associated with maturation changes of the cervical vertebrae C2 to C4.[78] Only one observer, the chief investigator, evaluated twenty five cropped images on two occasions, over a four week period. There was no further information provided by the author whether these images used were randomised, the conditions under which they were assessed or whether any time restrictions were used. Intra-observer reliability was reported using robust statistical analyses including a kappa value (0.80), a weighted kappa value (0.84) and a percentage agreement (84%). While the type of kappa weighting was not provided, these values reflect a 'substantial to almost perfect' level of reliability of the CVM staging method[147] and are similar to the results of this study. However, a direct comparison cannot be made with this study due to the differences in methodologies as outlined.

Other studies have evaluated the intra-observer reliability of cropped images using a variety of different CVM staging indices, while many more failed to clarify the type of image used (*Table* 2.6).[22, 23, 32, 37, 38, 130, 138, 139] One study, using an earlier devised CVM staging index from Baccetti *et al.*[4] and a Canadian sample of ten images reported a high level of intra-observer reliability with an intra-class correlation coefficient of 0.889 (range between 0.723-0.968).[130] While the intra-observer reliability reported is similar to this study, it must be interpreted with caution due to the small sample size, the inappropriate use of a quantitative statistical method for ordinal data,[138] the use of only one observer, with a research level of experience and the evaluation of traced images, all of which may have contributed to this high agreement value. San Roman *et al.*[32] reported positive intra-observer reliability using two different methods of CVM staging indices[26, 29] with a Pearson's correlation coefficient range between 0.96 and 0.99. However, the study shared similar deficiencies in its methodology, as outlined above, including the use of an inappropriate statistical method. A study evaluating the Hassel and Farman staging method[29] using cropped images reported a weighted kappa of 'almost perfect' reliability[22] similar to this study. However, while a large sample size was used the author did not state how many observers were used to establish this intra-observer reliability.

Sohrabi *et al.*[138] assessed intra-observer reliability of 5 observers evaluating 70 cropped images using the earlier CVM staging index by Baccetti et al.[4] A Fleiss kappa statistic was used to assess the intra-observer reliability and reported values of between 0.59 and 0.85, reflecting a 'moderate to almost perfect' agreement, which is in agreement with this findings of this study. It is important to mention that the authors did not provide information pertaining to whether the observers were involved in the study design or whether training in CVM staging was given prior to the reliability study and these factors can therefore affect the generalisability of the results.

9.2.3 Comparison with other intra-observer reliability studies using full images

In this study the overall intra-observer agreement for the application of the modified CVM staging index to full images as applied by the 22 orthodontic clinicians to the contemporary sample of images was reported as 'substantial' for linear weighted kappa (0.71, percentage agreement 91%) and 'almost perfect' for quadratic weighted kappa (0.84, percentage agreement 97% *Table* 8.1 and 8.3). This is an acceptable level of intra-observer reliability to recommend the use of the modified CVM staging index.[28] This study can be compared to a number of studies which have evaluated the intra-observer reliability of the modified CVM staging method[34, 36, 42] (*Table* 2.6) and other CVM staging methods using full lateral cephalogram images.[25, 140]

Lai *et al.*[34] reported a similar intra-observer reliability of the modified CVM staging method[28] of 90% using a Taiwanese cross-sectional sample of 30 randomised images. However, the author gave no details regarding the method of randomisation used or why only 30 full images were used from a total sample of 709 lateral cephalograms. The experience of the observer was not stated and only one observer was used to evaluate the intra-observer reliability, therefore reducing the generalisability of the study.[164] As previously discussed percentage agreement is inappropriate as it does not account for chance agreement[150] and a chance corrected kappa statistic was not reported.[34]

Zhao *et al.*[36] evaluated the reliability of CVM staging using 86 randomised full lateral cephalograms. Eleven observers, with an average of 17 years of clinical experience and independent of research design were recruited, increasing its generalisability. 'Substantial' intra-observer agreement was reported with a weighted kappa range of 0.53-0.86, however if this range is evaluated using the Landis and Koch scale[147] (*Table* 2.9), these values equate to 'moderate to almost perfect' agreement, which is very similar to the results found in this study. An average percentage agreement of 56.9% was reported with a range of 40.7%-79.1% which the authors regarded as a ‘great variation’ in intra-observer agreement as the observer had nearly half the opportunity to assess the image differently during the second occasion. The reason suggested for this ‘great variation’ was that the index was based on the subjective evaluation of the cervical vertebrae morphology and that the morphology of the cervical vertebra showed marked variation between the subjects. The intra-observer reliability findings using full images reported in this study is supported by Rainey *et al.*[42] which shares a similar methodology. Both Rainey *et al*. and this study carried out a sample size calculation, used a contemporary sample of pre-treatment lateral cephalogram images and recruited observers with a wide range of clinical experience. The twenty observers that participated in the Rainey *et al*. study received the same training in CVM staging method as this study, using the same teaching methods and materials. Linear weighted kappa was used to evaluate the intra-observer reliability which was reported as 0.70, a ‘substantial’ agreement and percentage agreement of 89%.[42]

Two studies[25, 140] evaluated the intra-observer reliability of full images using an earlier devised CVM staging methods.[4, 29, 143] Thirteen observers, all of which were trainee orthodontists and independent of the study design assessed fifteen full lateral cephalograms from a University based clinical record archive.[25] These images were identified based on their image quality and screened to ensure that the cervical vertebral features of all five CVM stages were present.[25] This small sample size affects the generalisability of the findings while the screening of images possibly contributes to selection bias. The intra-observer reliability reported was weighted kappa 0.86, reflecting an ‘almost perfect’ reliability, which is in agreement with the quadratic intra-observer reliability of this study. Jaqueira *et al*. assessed three different CVM staging methods.[140] One observer, trained in all three methods assessed the intra-observer reliability, with a 15 day interval between phases. The limitations of using one observer has already been discussed. The study evaluated 23 randomised lateral cephalograms which were selected based on being of 'good quality' with 'no overlapping cervical vertebrae'. This may have introduced a selection bias while the randomisation process was not described, therefore its adequacy cannot be determined. The author concluded that all three methods of CVM staging demonstrated clinical applicability but Baccetti *et al.* method[4] achieved the best intra-observer reliability with a weighted kappa reported of between 0.73-0.76, which is similar to this and other studies.[36, 42]

9.3 Overall inter-observer reliability results

9.3.1 Overall inter-observer reliability for cropped and full images

This study reports that the overall inter-observer reliability of cropped images for the application of the modified CVM staging index as applied by 22 orthodontic clinicians to a contemporary sample of images over the two phases was found to be 'moderate' using the Fleiss-Nee-Landis kappa statistic (Phase 1 = 0.46; Phase 2 = 0.43), percentage agreement 85% and 89%, respectively. The overall inter-observer reliability of full images over the two phases using the Fleiss-Nee-Landis kappa was reported as 'fair to moderate' with (Phase 1 = 0.44; Phase 2 = 0.40), percentage agreement 85% and 88%, respectively (*Table* 8.2). On analysis of the results there appears to be a discrepancy between the kappa values for inter-observer reliability for both cropped and full images and the overall percentage agreement values, with 'fair to moderate' kappa values and high percentage agreements being reported. This was also reported in similar studies.[25, 33] This discrepancy may be accounted for in that the percentage agreement does not take into consideration chance agreement.[149] Including agreements accounted for by chance can over-estimate the level of agreement reported and produce misleading levels of agreement; therefore percentage agreement values must be interpreted with caution.[149]

Both cropped and full images demonstrated a reduction of inter-observer reliability from phase 1 to phase 2. This is an interesting finding as reliability usually should increase with repeated training in the CVM staging method and experience.[40, 42, 137]

The poor inter-observer results reported in this study cannot recommend the use of the modified CVM staging method. While the Landis and Koch reference scale for the interpretation of kappa scores[147] categorises 0.41-0.60 as moderate agreement, recent literature has expressed that these values may be too lenient for health related studies as it implies that such a low score as 0.41 might be acceptable.[166] The literature assessing inter-observer reliability of the modified CVM staging method reports high degrees of variability when compared with the findings of this study.

9.3.2 Comparison with other inter-observer reliability studies using cropped images

Gabriel *et al.* reported the inter-observer reliability of ten observers evaluating ninety radiographs using the Kendall coefficient of concordance, as 0.74 on the first occasion and 0.72 on the second occasion, three weeks later, with overall agreement less than 50%. These values were reported as a ‘moderate’ level of agreement by the authors, again without referencing the categorisation scale. The inter-observer reliability values reported by Gabriel *et al*. are higher than the values reported in this study, however, a similar trend is seen in that the reliability reported in the second occasion is less than the first occasion. Gabriel *et al*. concluded that the CVM staging method is not reproducible and therefore could not recommend its clinical use. However, it has been argued that this conclusion was based only on the percentage agreement value which showed a discrepancy with the Kendall W values.[25] Of interest is the use of the Kendall W statistic, which is used to assess agreement among observers when the set of data is to be ranked.[168] The use of the Fleiss-Nee-Landis kappa statistic for the assessment of inter-observer reliability would have been a more appropriate statistic, as discussed previously.[159] In fact, of the eight studies in the literature that assessed the inter-observer reliability of the modified CVM staging method using cropped images, four inappropriately used the Kendall W statistic[33, 40, 41, 137] and therefore these results should be interpreted with caution (*Table* 9.2).

A study by Ballrick *et al.* evaluating full lateral cephalograms using a different staging index,[4] shared similar methodologies with several studies,[33, 40, 41, 137] including the use of Kendall W statistic. Ballrick *et al*. reported the inter-observer reliability as 0.79 in phase 1 and 0.86 in phase 2. The author compared the inter-observer reliability results with Gabriel *et al*. and concluded that the results suggested that CVMM tends to be unreliable in diagnosing each distinct stage but reliable in detecting the interface between pre- and post-peak phases of mandibular growth, making it a useful clinical tool.

The study by Nestman *et al*. again followed a similar methodology to Gabriel *et al.,* however the reported inter-observer reliability was much lower with an overall Kendall W value of 0.45 over the two different occasions.[41] As the author failed to provide additional information on the range of values achieved by the observers, it is difficult to account for this much lower value when compared to Gabriel *et al*. Nestman *et al.* found that the inter-observer reliability was high for assessing the lower borders of C2 to C4 as either flat or curved but low inter-observer values were found for distinguishing the morphology of the vertebral bodies C3 and C4, concluding an overall poor reliability of the CVMM.[41]

Alkhal *et al*. evaluated the inter-observer reliability using two observers, one of which was involved in the study design and a small number of images. The study reported inter-observer reliability with a kappa value of 0.846, which can be interpreted as 'almost perfect' agreement[147] and a percentage agreement of 96%. This high level of inter-observer reliability differed from the results of this study, however this may be due to the fact that one of the two observers had an expert level of knowledge regarding CVM staging.[27, 33] As previously noted, the author did not provide any details with regard to the randomisation process or to the small number of images used. Wong *et al*. used the same sample as Alkhal *et al.* and used two observers, one being the named author. Inter-observer reliability was reported as 92%[39] but the study lacks generalisability for reasons previously discussed.

Predko-Engel *et al*. reported inter-observer reliability as weighted kappa 0.28 which would reflect a 'fair' level of agreement and concluded that the reliability of CVM staging is questionable, recommending an additional biological indicator to predict peak mandibular growth.[141] While no further information is provided by the author regarding the type of weighted kappa used, the study design is similar to this study in that the observers used had varying levels of CVM staging experience and the digital images were presented on individual computer monitors. Engel *et al*. reported a similar 'fair' inter-observer agreement with a kappa value of 0.30.[131] Similarly, a small sample size of images were evaluated and only five observers were used, both of which affect the generalisability of the results.

Author	Image number	Observer number	Traced CV	Statistic	Intra-observer reliability	Bias risk
Gabriel[33]	90	10	No	Kendall W	Ph 1: 0.72 Ph 2: 0.72	↑
Nestman[41]	90	10	No	Kendall W	Ph1+Ph2: 0.45	↑
Alkhal[35]	25	2	NR	Kappa	0.846	↑
Wong[39]	25	2	NR	Percentage agreement	92%	↑
Rongo[40]	50	30	NR	Kendall W	Ph1: 0.70 Ph2: 0.81	↑
Perinetti[137]	72	10	Yes	Kendall W	Ph1: 0.90 Ph2: 0.91	↑
Predko-Engel[141]	50	10	No	K^w	0.28	↑
Engel[131]	29	5	No	Kappa	0.30	↑

Table 9.2: Summary of reported inter-observer reliability of the modified CVM method[28] using cropped images.

Key: NR=not reported; CV=cervical vertebrae; K^w=weighted kappa

Sohradi *et al.* assessed inter-observer reliability of an earlier CVM staging index[4] using 5 'experienced orthodontists' evaluating 70 randomised cropped images. The average weighted kappa reported was 0.48 over the two occasions (phase 1=0.45 and phase 2=0.51) with an overall percentage agreement of 58%[138], reflecting a 'moderate' level of agreement similar to the findings of this study. Also similar to this study, Sohrabi *et al*. attempted to reduce observer fatigue by implementing regular breaks.[138]

Two studies examined inter-observer reliability of CVM staging using Turkish populations.[31, 139] Similarities between these two Turkish based studies include the use of two observers to rate inter-observer reliability, both of which were named authors and poor reporting of the methodologies particular with regard to prior CVM training and the type of images used. Uysal *et al.* evaluated 30 images using the Hassel and Farman CVM staging method[29] and assessed inter-observer reliability using the Spearman Brown statistic, reporting a value of 0.987.[31] This high inter-observer value while similar to other studies[35, 137] must be interpreted with caution as only two observers assessed the images, which were small in number and an inappropriate statistical method was used. The Spearman Brown formula is a correlation coefficient which measures association and is not a true measure of agreement.[138] Ozer *et al.*[139] assessed inter-observer reliability using Küçükkeles staging method.[38] Two observers, both of which were named authors and 150 images were evaluated, reported a 98% agreement. The short comings associated with the use of percentage agreement has previously been discussed.[149]

Danaei *et al*. evaluated the inter-observer reliability of 178 images using the Hassel and Farman staging method and two observers.[22] A high weighted kappa value of 0.89 was reported, indicating 'almost perfect' reliability, however, one of the observers was a named author which affects the generalisability of the result.

9.3.3 Comparison with other inter-observer reliability studies using full images

The inter-observer reliability using full images in this study was reported as a 'moderate' level of reliability in phase 1 (Fleiss-Nee-Landis kappa 0.44) and a 'poor' level of reliability in phase 2 (Fleiss-Nee-Landis kappa 0.40). Three studies[34, 36, 42] have evaluated the inter-observer reliability of full images using the modified CVM staging method.

Lai *et al*. assessed the inter-observer reliability using three observers on one occasion grading 30 randomised images. The author was involved in staging the images but did not report on the level of experience of the additional two observers and there was no comment about the provision of CVM training. They reported a 93.3% agreement (for observer A and B) and a 90% agreement (for observer A and C). This is similar to the findings reported in this study, however, the short comings associated with the reporting of percentage agreement has already been previously discussed.[149] Inter-observer reliability was also reported using the Spearman's rank correlation coefficient with a value of 0.96-0.98. Again, this statistical method evaluates correlation and is not a measure of agreement, therefore these values should be interpreted with caution.

Zhao *et al.* evaluated the inter-observer reliability between 11 orthodontists, none of which were involved in the study design, staging 86 cephalograms.[36] The authors assessed inter-observer reliability using Kendall's W statistic and percentage agreement with the phase 1 value reported as 0.83 and 39.3%, respectively while phase 2 had a value of 0.84 and 42%, respectively. Using the Kendall's W values, the authors concluded a strong statistical agreement among observers regarding CVM staging. However, as previously outlined above, Kendall's W is not an appropriate statistic to evaluate agreement. Again, a discrepancy can be detected between the statistic and percentage agreement values, similar to this study. The reasons for this has been previously described.[149] Indeed a percentage agreement of between 39-42% reflects a great variability in agreement.

The methodologies of both Rainey *et al.* and this study are very similar, however the reported inter-observer reliability differs, with Rainey *et al.* achieving a higher linear weighted kappa value of 0.68 for phase 1 and 0.66 for phase 2, reflecting a 'substantial' level of agreement.[42] The inter-observer reliability did reduce from phase 1 to phase 2, albeit was not statistically significant.

Jaqueira *et al.* evaluated inter-observer reliability of a different CVM staging method[4] using a small sample of 23 full images and 4 observers. The study reported an inter-observer reliability of weighted kappa between 0.73-0.75.[140] The methodological deficiencies of this study has previously been described, therefore these results should be interpreted with caution.

9.4 Overall reliability of orthodontic trainees

9.4.1 Trainee intra-observer reliability for cropped versus full images

The findings of this study demonstrated that there was no significant difference in the intra-observer reliability achieved by the 11 orthodontic trainees using cropped and full images. The quadratic weighted kappa values achieved for cropped images were reported as a 'substantial-almost perfect' level of agreement (0.76-0.95) with a 98% agreement and this was the same level of agreement achieved for the full images (Quadratic K^w 0.75-0.92, 97% agreement).

A minor difference was noted when the linear weighted kappa values were evaluated. The intra-observer reliability achieved for cropped images were reported as 'substantial-almost perfect' level of agreement (0.67-0.87 with a 92% agreement) and was comparable to the 'moderate-almost perfect' intra-observer reliability values achieved for the full images (0.58-0.84, percentage agreement 90%, *Table* 8.3).

As no previous study has examined intra-observer reliability of cropped versus full images, a direct comparison cannot be made with the available literature, however, the findings based on this study, concluded that the cropping of lateral cephalograms did not affect the reliability of the modified CVM staging as assessed by this sample of University of Liverpool orthodontic trainees.

9.4.2 Orthodontic trainee inter-observer reliability for cropped versus full images

The trainee inter-observer reliability reported for cropped and full images were identical, with the inter-observer reliability, as assessed by the Fleiss-Nee-Landis kappa indicating a 'moderate' level of agreement for cropped images (Phase 1=0.5 and Phase 2=0.5) and full images (Phase 1=0.46 and Phase 2=0.43).

While the Berry-Mielke values indicated a 'fair' agreement for cropped images (Phase 1=0.37 and Phase 2=0.29) and full images (Phase 1=0.31 and Phase 2=0.35). The overall percentage agreement was very similar with 90% and 89% reported for Phase 1 and Phase 2, respectively (*Table* 8.4).

Therefore, it can be concluded that the cropped image did not affect the inter-observer reliability of the modified CVM staging method. However, as previously discussed, the kappa value for inter-observer agreement was below the threshold value of 0.5 which has been expressed by some as an unacceptable level of reliability for health related studies as it implies that such a low score such 0.43 might be deemed acceptable.[166]

9.5 Overall reliability of orthodontic specialists

9.5.1 Specialist intra-observer reliability for cropped versus full images

The findings of this study demonstrated that there was no significant difference in the intra-observer reliability achieved by the 11 orthodontic specialists using cropped and full images. The quadratic weighted kappa values achieved for cropped images were reported as a 'substantial-almost perfect' level of agreement (0.65-0.89) with a 97% agreement and this was the same level of agreement achieved for the full images (Quadratic K^w 0.75-0.90, 97% agreement).

A minor difference was noted when the linear weighted kappa values were evaluated. The intra-observer reliability achieved for cropped images were reported as 'moderate-substantial' level of agreement (0.55-0.76 with a 90% agreement) and was comparable to the 'substantial' intra-observer reliability values achieved for the full images (0.62-0.79, percentage agreement 91% *Table* 8.5).

9.5.2 Specialist inter-observer reliability for cropped versus full images

The orthodontic specialists inter-observer reliability reported for cropped and full images were very similar. The inter-observer reliability, as assessed by the Fleiss-Nee-Landis kappa statistic indicated a 'moderate' level of agreement for cropped images (Phase 1=0.44) and full images (Phase 1=0.43). The level of inter-observer reliability reduced to 'fair' in phase 2 for both cropped (Phase 2=0.39) and full images (Phase 2=0.38).

The Berry-Mielke statistical values were identical for both cropped and full images. The values indicated a 'fair' agreement for phase 1 (cropped=0.25 and full=0.25) and 'slight' agreement for phase 2 (cropped images=0.29 and full images=0.35). The overall percentage agreement reported was very similar for cropped images (80% and 88% reported for Phase 1 and Phase 2, respectively) and full images (80% and 87% reported for Phase 1 and Phase 2, respectively *Table* 8.6).

It can be concluded that the cropped image did not affect the inter-observer reliability of the modified CVM staging method. However, as previously discussed, the kappa value for inter-observer agreement was below the threshold value of 0.5 which is not an acceptable level of reliability.[166]

9.6 Comparison of agreement between observers with different levels of clinical experience

9.6.1 Intra-observer reliability of orthodontic trainees compared with specialists

When linear weighted kappa was applied, the level of agreement for the orthodontic trainees was found to be 'substantial' to 'almost perfect' for cropped images and 'moderate' to 'almost perfect' for full images. This difference in the level of agreement noted for orthodontic trainees between cropped and full images could be accounted for by trainee number 11 who scored 0.67 for the cropped images and 0.58 for the full images.

Similarly, the linear weighted kappa assessment of the intra-examiner reliability of the orthodontic specialists for cropped images was reported as 'moderate' to 'substantial' and ‘substantial’ agreement for full images. This would indicate that CVM staging of full images by the orthodontic specialists was marginally more reliable than cropped images. This difference in CVM staging reliability between the cropped and full stages may be attributable to specialist number 11, who scored 0.55 for the cropped images and 0.73 for the full images. It is interesting to note that the both outliers were from the same orthodontic unit.

When the quadratic weighted kappa values were considered both the orthodontic trainees and the specialists both had ‘substantial’ to ‘almost perfect’ agreement for both cropped and full images. Therefore, as the quadratic weighted kappa statistic is more reflective of clinical practice, these results would suggest that the level of clinical experience did not affect the reliability of the CVM staging method when applied to cropped images (*Table* 8.7 and 8.8)

9.6.2 Inter-observer reliability of orthodontic trainees compared with specialists

No difference was noted in the inter-observer reliability for cropped and full images when assessed by the orthodontic trainees with ‘moderate’ levels of reliability reported for the Fleiss-Nee-Landis kappa statistic and ‘fair’ levels of reliability reported for the Berry-Mielke statistic (*Table* 8.4).

When inter-observer reliability was assessed using the Fleiss-Nee-Landis kappa statistic the specialists reported ‘moderate’ levels of reliability for cropped and full images in phase 1 but ‘fair’ levels in phase 2. This trend was replicated when assessed using the Berry-Mielke statistic with phase 1 reliability of cropped and full images reporting a ‘fair’ level of reliability compared to a ‘slight’ level of reliability in phase 2 (*Table* 8.6).

This would suggest that the orthodontic trainees showed more consistent inter-observer reliability between phase 1 and phase 2 for both cropped and full images than the specialists. However, no obvious reason can be proposed for this difference, as both groups received the same standardised training in CVM staging, the same randomised images presented in the same manner on individualised computer monitors, the same amount of designated breaks and reference materials. Inter-observer reliability would be expected to increase with training and experience.[25, 40, 137, 138] Perhaps, fatigue may be an insidious factor which may account for this reduction in the specialists inter-observer reliability. Similarly several studies have noted a reduction the inter-observer reliability from phase 1 and phase 2[33, 42] but no reasons were suggested as to why this occurred.

A number of studies have examined whether orthodontic experience affects the reliability of the modified CVM staging method. [40, 42, 141] Rongo *et al*. concluded that the level of orthodontic experience did affect the reliability of CVM staging method.[40] By grouping the observers based on the level of orthodontic experience, three grades were established; dental graduates with less than one year of clinical experience, postgraduates with between 2-4 years of clinical experience and specialists with more than 7 years of clinical practice. The dental graduates reported higher intra- and inter-observer (weighted kappa 0.78 and Kendall's W statistic 0.87, respectively) than the specialist group (weighted kappa 0.64 and Kendall's W statistic 0.61, respectively). Rongo *et al*. suggested that the better intra- and inter-observer reliability reported by the 'junior group' was due to the tuition received by this group in their final undergraduate year, while the specialist orthodontists did not routinely use CVM staging on a daily basis. However, this study did not use the most appropriate statistic to assess the inter-observer reliability and the results should therefore be interpreted with caution.

Predko-Engel *et al*. also evaluated the effect of clinical experience on the reliability of CVM staging,[141] however, they concluded that the clinicians with more experience in CVM staging reported more consistent inter-observer reliability than those clinicians with little or no experience in the use of the CVM method (kappa 0.39 and 0.25 for experienced and inexperienced assessors, respectively). Interestingly, 'experienced and inexperienced' assessors in CVM staging reported very similar intra-observer reliability (kappa 0.45 and 0.44 for experienced and inexperienced assessors, respectively).

Rainey *et al*. compared the intra- and inter-observer reliability between the consultant orthodontists and trainee orthodontists and concluded that there was no statistically significant difference identified between these two groups.[42]

9.7 Variability in CVM training provided

Several studies have addressed the issue of training observers in the CVM staging method prior to assessing images.[36, 40, 137] The study conducted by Rainey *et al*.[42] and this study followed a very similar teaching approach of the modified CVM staging method, using material provided by one of the developing authors. Following training, a calibration exercise was carried out before each phase, ensuring that all observers had the same standard of knowledge prior to the commencement of the study.

Like Rainey et al., Perinetti *et al*. underwent repeated training of all assessors in CVM staging prior to the evaluation of images[137] and concluded that regular training in CVM staging was necessary to obtain high accuracy and intra-observer reliability in the visual assessment of CVM stages. Rainey and Perinetti *et al*., [42, 137] and this study all had similar intra-observer reliability values.

Formal training in the CVM staging method was found lacking in other studies and relied on observers receiving a copy of a schematic representation of the individual stages without any further explanation provided.[33, 41, 131, 141] All these studies reported low agreement and concluded that CVM staging was unreliable. Rongo *et al.* argued that the level of practice and knowledge of CVM staging was critical for its reliability with the simple use of diagrams not being sufficient to obtain a good level of knowledge and familiarity with CVM.[40]

9.8 Limitations of the study

9.8.1 Identification of image sample

The sample of lateral cephalogram images used in this study was obtained from the patient databases of the first year registrars who started their training programme in

October 2014. Therefore, this included only patients who had been assessed and deemed eligible for orthodontic treatment under the NHS and within a hospital setting. While this facilitated the systematic collection of suitable images which met the inclusion and exclusion criteria of this study in a convenient time frame the images collected was a sample of convenience and were not selected randomly, as recommended.[27] Therefore, the possibility of a selection bias must be considered. Images were selected regardless of their quality to replicate routine clinical practice, reduce selection bias and increase the generalisability of the results. The image sample generated was compiled from the first 72 consecutive lateral cephalograms satisfying the inclusion and exclusion criteria with no further image collection performed thereafter. No attempt was made to ensure a balance across the six CVM stages. Therefore, this may have affected reliability as studies have suggested that CVM stages 1 and 2 are easier to stage than the other stages.[25, 41, 137]

9.8.2 Quality of images

The use of digital radiographs facilitated the manipulation of the image into a cropped format, whereby only the cervical vertebrae 1 to 4 could be identified. During the cropping process, the image was magnified to replicate the size and dimension of the full image. As a result, this may have distorted and reduced the quality of the cropped image to be assessed. However, the intra-observer reliability results obtained in this study would suggest that the magnification process it did not adversely affect the image quality.

9.8.3 Quality of CVM training

In a reliability study, the aim is to replicate the study conditions as strictly as possible.[145] In this study, each recruited observer, independent of his or her orthodontic experience underwent two training sessions in the CVM staging method. The format and structure of the CVM training was based on material provided by one of the original developers of the modified CVM staging method.[28] These steps facilitated a consistent and standardised approach of each teaching session. Time was allocated after the training session for the clarification of any points. While the teaching material was standardised, variability may have been inadvertently introduced by the presenter or during the allocated clarification time. A pre-recorded video presentation of CVM training would have ensured standardisation across both phases, so that the information provided before each staging phase was as consistent as possible.

9.8.3 Presentation of images

All images were exported from the Trust PACS system, anonymised and uploaded on to a timed *PowerPoint*™ presentation. The 144 images were randomised prior to each staging phase as recommended from the literature to reduce bias[27] and prevent image memory. Incorporating a number of rest periods and breaks throughout the staging process was aimed to reduce assessment variability due to operator fatigue, however, the assessment of 144 images was very demanding on the observers. Each image to be assessed was displayed for only 20 seconds, which is less time a clinician would take if grading a particularly difficult image clinically. Therefore, this time limitation was not reflective of normal clinical practice, however, a specific time period was chosen due to the amount of images that required evaluation.

9.8.4 Presentation environment

All participants were seated in a designated computer laboratory and the images were displayed on individual computer monitors to best replicate the clinical environment. This was an improvement of a recent study which projected the images on a large projector screen.[42] However, one potential criticism of this approach is that the observers were not given the opportunity to enhance images, by way of magnification, brightness or contrast which can be done routinely in practice and this may have facilitated a more accurate assessment of the CVM stage.

9.9 Implications of this study

9.9.1 Implications for clinical practice

The main implication for clinical practice is that this research confirms previous findings that CVM staging is reliable when applied by orthodontists who have been trained in the modified CVM staging method. The findings of this study would suggest that the dentition, which is visible on the unaltered lateral cephalogram, did not influence CVM staging. As the inter-observer reliability results suggested a 'fair-moderate' level of reliability, the CVM staging method should be used in conjunction with other growth indicators.

9.9.2 Implications for Future Research

The clinical application of the modified CVM staging method to determine the stage of mandibular growth for optimal treatment timing is both non-invasive and time efficient. However, several studies have criticised the use of CVM staging to predict the peak in mandibular growth, citing that this method is both unreliable and subjective.[33, 41 141] Higher CVM reliability has been reported for the assessment of the inferior borders of the cervical vertebrae than the assessment of the morphology of the vertebral bodies, which accounts for the overall deterioration of the CVM staging method in the post-pubertal stages (CS5 and CV6).[41] It has been suggested that instead of a 6 stage CVM staging method, a two phase pre- and post-pubertal method should be adapted for clinical practice.[25] Recently, computer software has been developed for the geometric assessment of the cervical vertebra in an attempt to remove subjectivity, reduce examiner influence[169, 170] and categorise the more difficult borderline cases. It has been suggested that CVM staging offers no advantage over chronological age in either the assessment of skeletal age or predicting peak growth.[71, 77-79] Therefore, future research should focus on the geometric morphometric evaluation of the cervical vertebrae and CVM validity.

Chapter 10: Conclusions

- The findings of this study demonstrated that the reliability of the modified CVM staging method for the identification of the peak mandibular growth was not improved by using cropped images.
- The findings suggest that the modified CVM staging method had ‘almost perfect’ agreement for intra-observer reliability using both full and cropped lateral cephalogram images.
- The findings suggest that the modified CVM staging method had ‘fair to moderate’ agreement for inter-observer reliability.
- The level of orthodontic experience did not affect the reliability of the modified CVM staging method.

References

1. Cozza P, Baccetti T, Franchi L, De Toffol L, McNamara Jr JA. Mandibular changes produced by functional appliances in Class II malocclusion: A systematic review. *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;129(5):599.e1-12.
2. Teuscher U. Original article: A growth-related concept for skeletal Class II treatment. *American Journal of Orthodontics* 1978;74:258-75.
3. Franchi L, Baccetti T, McNamara JA, Jr. Mandibular growth as related to cervical vertebral maturation and body height. *American Journal of Orthodontics and Dentofacial Orthopedics* 2000;118(3):335-40.
4. Baccetti T, Franchi L, McNamara Jr JA. An Improved Version of the Cervical Vertebral Maturation (CVM) Method for the Assessment of Mandibular Growth. *Angle Orthodontist* 2002;72(4):316-23.
5. Baccetti T, Franchi L, Toth LR, McNamara JJA. Original Articles: Treatment timing for Twin-block therapy. *American Journal of Orthodontics and Dentofacial Orthopedics* 2000;118:159-70.
6. Mojdehi M, Buschang PH, English JD, Wolford LM. Postsurgical growth changes in the mandible of adolescents with vertical maxillary excess growth pattern. *American Journal of Orthodontics and Dentofacial Orthopedics* 2001;119(2):106-16.
7. Bousquet P, Ansermino H, Canal P, Artz C, Renaud M. Growth and dental implants: assessment and prevention of the long-term aesthetic risk. *L' Orthodontie francaise* 2016;87(3):321-8.
8. Littlewood SJ, Kandasamy S, Huang G. Retention and relapse in clinical practice. *Australian Dental Journal* 2017;62:51-7.
9. Pirinen S. Endocrine regulation of craniofacial growth. *Acta Odontologica Scandinavica* 1995;53(3):179-85.
10. Sadowsky PL. Craniofacial growth and the timing of treatment. *American Journal Of Orthodontics And Dentofacial Orthopedics* 1998;113(1):19-23.
11. Nanda RS. The contributions of craniofacial growth to clinical orthodontics. *American Journal of Orthodontics and Dentofacial Orthopedics* 2000;117(5):553-5.
12. Flores-Mir C, Nebbe B, Major PW. Use of skeletal maturation based on hand-wrist radiographic analysis as a predictor of facial growth: a systematic review. *Angle Orthodontist* 2004;74(1):118-24.
13. Hägg U, Taranger J. Maturation indicators and the pubertal growth spurt. *American Journal of Orthodontics* 1982;82(4):299-309.
14. Tofani MI. Mandibular growth at puberty. *American Journal of Orthodontics* 1972;62(2):176-95.
15. Masoud M, Masoud I, Kent RL, Jr., Gowharji N, Cohen LE. Assessing skeletal maturity by using blood spot insulin-like growth factor I (IGF-I) testing. *American Journal of Orthodontics and Dentofacial Orthopedics* 2008;134(2):209-16.
16. Masoud MI, Marghalani HYA, Alamoudi NM, El Derw D, Masoud IM, Gowharji NF. Longitudinal relationship between insulin-like growth factor-1 levels and vertical facial growth. *Journal of Orofacial Orthopedics* 2015;76(5):440-50.
17. Masoud MI, Marghalani HYA, Bamashmous M, Alamoudi NM, El Derwi D, Masoud IM, et al. Predicting changes in mandibular length and total

anterior facial height using IGF-1, cervical stage, skeletal classification, and gender. *Progress in Orthodontics* 2015;16:7-13.

18. Perinetti G, Paolantonio M, Femminella B, Serra E, Spoto G. Gingival crevicular fluid alkaline phosphatase activity reflects periodontal healing/recurrent inflammation phases in chronic periodontitis patients. *Journal of Periodontology* 2008;79(7):1200-7.
19. Perinetti G, Baccetti T, Contardo L, Di Lenarda R. Gingival crevicular fluid alkaline phosphatase activity as a non-invasive biomarker of skeletal maturation. *Orthodontics and Craniofacial Research* 2011;14(1):44-50.
20. Perinetti G, Baccetti T, Di Leonardo B, Di Lenarda R, Contardo L. Dentition phase and chronological age in relation to gingival crevicular fluid alkaline phosphatase activity in growing subjects. *Progress in Orthodontics* 2011;12(2):100-6.
21. Fishman LS. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wirst films. *Angle Orthodontist* 1982;52(2):88-112.
22. Danaei SM, Karamifar A, Sardarian A, Shahidi S, Karamifar H, Alipour A, et al. Measuring agreement between cervical vertebrae and hand-wrist maturation in determining skeletal age: reassessing the theory in patients with short stature. *American Journal of Orthodontics and Dentofacial Orthopedics* 2014;146(3):294-8.
23. Soegiharto BM, Moles DR, Cunningham SJ. Discriminatory ability of the skeletal maturation index and the cervical vertebrae maturation index in detecting peak pubertal growth in Indonesian and white subjects with receiver operating characteristics analysis. *American Journal of Orthodontics and Dentofacial Orthopedics* 2008;134(2):227-37.
24. Isaacson KG TA, Atack NE, Horner K, Whaites E. *Guidelines for the use of radiographs in clinical orthodontics*. 4th Ed ed. London: British Orthodontic Society, 2015.
25. Ballrick JW, Fields HW, Beck FM, Sun Z, Germak J. The cervical vertebrae staging method's reliability in detecting pre and post mandibular growth. *Orthodontic Waves* 2013;72(3):105-11.
26. Lamparski D. *Skeletal age assessment utilizing cervical vertebrae. 1972. 164 f*. Dissertation (Master of Dental Science)-Faculty of the School of Dental Medicine, University of Pittsburgh, Pittsburgh, 1972.[Links], 1972.
27. Santiago RC, de Miranda Costa LF, Vitral RW, Fraga MR, Bolognese AM, Maia LC. Cervical vertebral maturation as a biologic indicator of skeletal maturity. *Angle Orthodontist* 2012;82(6):1123-31.
28. Baccetti T, Franchi L, McNamara JA. The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Seminars in Orthodontics* 2005;11(3):119-29.
29. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *American Journal of Orthodontics and Dentofacial Orthopedics* 1995;107(1):58-66.
30. Luci Mara Fachardo J, Monica Costa A, Luciano José P, Carlos Eduardo Pinto de A, Leandro Silva M. Determining skeletal maturation stage using cervical vertebrae: evaluation of three diagnostic methods. *Brazilian Oral Research* 2010(4):433-7.

31. Uysal T, Ramoglu SI, Basciftci FA, Sari Z. Chronologic age and skeletal maturation of the cervical vertebrae and hand-wrist: is there a relationship? *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;130(5):622-8.
32. San Román P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. *European Journal of Orthodontics* 2002;24(3):303-11.
33. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: poor reproducibility. *American Journal of Orthodontics and Dentofacial Orthopedics* 2009;136(4):478.e1-7.
34. Lai EH, Liu J, Chang JZ, Tsai S, Yao C, Chen M. Radiographic Assessment of Skeletal Maturation Stages for Orthodontic Patients: Hand-wrist Bones or Cervical Vertebrae? *Journal of the Formosan Medical Association* 2008;107(4):316-25.
35. Alkhal HA, Wong RWK, Rabie ABM. Correlation between chronological age, cervical vertebral maturation and fishman's skeletal maturity indicators in southern Chinese. *Angle Orthodontist* 2008;78(4):591-6.
36. Zhao XG, Lin J, Jiang JH, Wang Q, Ng SH. Validity and reliability of a method for assessment of cervical vertebral maturation. *Angle Orthodontist* 2012;82(2):229-34.
37. Chang HP, Liao CH, Yang YH, Chang HF, Chen KC. Correlation of cervical vertebra maturation with hand-wrist maturation in children. *Kaohsiung Journal of Medical Sciences* 2001;17(1):29-35.
38. Kucukkeles N, Acar A, Biren S, Arun T. Comparisons between cervical vertebrae and hand-wrist maturation for the assessment of skeletal maturity. *Journal of Clinical Pediatric Dentistry* 1998;24(1):47-52.
39. Wong RW, Alkhal HA, Rabie AB. Use of cervical vertebral maturation to determine skeletal age. *American Journal of Orthodontics and Dentofacial Orthopedics* 2009;136(4):484 e1-6.
40. Rongo R, Valleta R, Bucci R, Bonetti GA, Michelotti A, D'Anto V. Does clinical experience affect the reproducibility of cervical vertebrae maturation method? *Angle Orthodontist* 2015;85(5):841-7.
41. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Original article: Cervical vertebrae maturation method morphologic criteria: Poor reproducibility. *American Journal of Orthodontics and Dentofacial Orthopedics* 2011;140:182-8.
42. Rainey BJ, Burnside G, Harrison JE. Reliability of cervical vertebral maturation staging. *American Journal of Orthodontics and Dentofacial Orthopedics* 2016;150(1):98-104.
43. Taranger J. Evaluation of biological maturation by means of maturity criteria. *Acta Paediatrica* 1976;65(S258):77-82.
44. Burstone CJ. Process of maturation and growth prediction. *American Journal of Orthodontics* 1963;49(12):907-19.
45. Carlson DS. Biological rationale for early treatment of dentofacial deformities. *American Journal of Orthodontics and Dentofacial Orthopedics* 2002;121(6):554-8.

46. Aelbers CF, Dermaut L. Orthopedics in orthodontics: Part I, fiction or reality a—review of the literature. *American Journal of Orthodontics and Dentofacial Orthopedics* 1996;110(5):513-9.
47. Franchi L, Baccetti T, De Toffol L, Polimeni A, Cozza P. Phases of the dentition for the assessment of skeletal maturity: a diagnostic performance study. *American Journal of Orthodontics and Dentofacial Orthopedics* 2008;133(3):395-400; quiz 76 e1-2.
48. Baccetti T, McGill JS, Franchi L, McNamara Jr JA, Tollaro I. Skeletal effects of early treatment of Class III malocclusion with maxillary expansion and face-mask therapy. *American journal of orthodontics and dentofacial orthopedics* 1998;113(3):333-43.
49. Stahl F, Baccetti T, Franchi L, McNamara Jr JA. Longitudinal growth changes in untreated subjects with Class II Division 1 malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics* 2008;134(1):125-37.
50. Suda N, Ishii-Suzuki M, Hirose K, Hiyama S, Suzuki S, Kuroda T. Effective treatment plan for maxillary protraction: is the bone age useful to determine the treatment plan? *American Journal of Orthodontics and Dentofacial Orthopedics* 2000;118(1):55-62.
51. Mandall N, DiBiase A, Littlewood S, Nute S, Stivaros N, McDowall R. Is early Class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 15-month follow-up. *Journal of Orthodontics* 2010;37(3):149-61.
52. Hägg U, Pancherz H. Dentofacial orthopaedics in relation to chronological age, growth period and skeletal development. An analysis of 72 male patients with class II division 1 malocclusion treated with the herbst appliance. *European Journal of Orthodontics* 1988;10(1):169-76.
53. Meikle M. *Craniofacial development, growth and evolution*. Bressingham, England: Bateson, 2002.
54. Moussa Pacha M, Fleming PS, Johal A. A comparison of the efficacy of fixed versus removable functional appliances in children with Class II malocclusion: A systematic review. *European Journal of Orthodontics* 2016;38(6):621-30.
55. Carlson DS. Theories of Craniofacial Growth in the Postgenomic Era. *Seminars in Orthodontics* 2005;11(4):172-83.
56. Moore RN. Principles of dentofacial orthopedics. *Seminars in Orthodontics* 1997;3(4):212-21.
57. Scott JH. Original article: The growth and function of the muscles of mastication in relation to the development of the facial skeleton and of the dentition. *American Journal of Orthodontics* 1954;40:429-49.
58. Scott JH. The Growth of the Human Face. *Journal of the Royal Society of Medicine* 1954;47(2):91-100.
59. Moss ML, Young RW. A functional approach to craniology. *American Journal of Physical Anthropology* 1960;18:281-92.
60. Petrovic A. Control of postnatal growth of secondary cartilages of the mandible by mechanisms regulating occlusion. Cybernetic model. *Transactions of the European Orthodontic Society* 1974:64-75.
61. Mitani H, Sato K. Comparison of mandibular growth with other variables during puberty. *Angle Orthodontist* 1992;62(3):217-22.

62. Nanda RS. The rates of growth of several facial components measured from serial cephalometric roentgenograms. *American Journal of Orthodontics* 1955;41(9):658-73.
63. Bambha JK. Longitudinal cephalometric roentgenographic study of face and cranium in relation to body height. *Journal Of The American Dental Association (1939)* 1961;63:776-99.
64. Scammon RE. The first seriatim study of human growth. *American Journal of Physical Anthropology* 1927;10(3):329-36.
65. Harris JA JC, Paterson DG, Scammon RE. *The measurement of the Body in Childhood*. Minneapolis: University of Minnesota Press, 1930:193.
66. Nanda RS, Ghosh J. Longitudinal growth changes in the sagittal relationship of maxilla and mandible. *American Journal of Orthodontics and Dentofacial Orthopedics* 1995;107(1):79-90.
67. O'Reilly MT, Yanniello GJ. Mandibular growth changes and maturation of cervical vertebrae. A longitudinal cephalometric study. *Angle Orthodontist* 1988;58(2):179-84.
68. Perinetti G, Contardo L, Castaldo A, McNamara JA, Jr., Franchi L. Diagnostic reliability of the cervical vertebral maturation method and standing height in the identification of the mandibular growth spurt. *Angle Orthodontist* 2016;86(4):599-609.
69. Houston W. Relationships between skeletal maturity estimated from hand-wrist radiographs and the timing of the adolescent growth spurt. *European Journal of Orthodontics* 1980;2(2):81-93.
70. Baccetti T, Franchi L, De Toffol L, Ghiozzi B, Cozza P. The diagnostic performance of chronologic age in the assessment of skeletal maturity. *Progress in Orthodontics* 2006;7(2):176-88.
71. Beit P, Peltomaki T, Schatzle M, Signorelli L, Patcas R. Evaluating the agreement of skeletal age assessment based on hand-wrist and cervical vertebrae radiography. *American Journal of Orthodontics and Dentofacial Orthopedics* 2013;144(6):838-47.
72. Gupta S, Jain S, Gupta P, Deoskar A. Determining skeletal maturation using insulin-like growth factor I (IGF-I) test. *Progress in Orthodontics* 2012;13(3):288-95.
73. Sierra AM. Assessment of dental and skeletal maturity: a new approach. *Angle Orthodontist* 1987;57(3):194-208.
74. Fishman LS. Chronological versus skeletal age, an evaluation of craniofacial growth. *Angle Orthodontist* 1979;49(3):181-9.
75. Hunter CJ. The correlation of facial growth with body height and skeletal maturation at adolescence. *Angle Orthodontist* 1966;36(1):44-54.
76. Bergersen EO. The male adolescent facial growth spurt: its prediction and relation to skeletal maturation. *Angle Orthodontist* 1972;42(4):319-38.
77. Mellion ZJ, Behrents RG, Johnston LE, Jr. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *American Journal of Orthodontics and Dentofacial Orthopedics* 2013;143(6):845-54.
78. Gray S, Bennani H, Kieser JA, Farella M. Morphometric analysis of cervical vertebrae in relation to mandibular growth. *American Journal of Orthodontics and Dentofacial Orthopedics* 2016;149(1):92-8.

79. Chatzigianni A, Halazonetis DJ. Geometric morphometric evaluation of cervical vertebrae shape and its relationship to skeletal maturation. *American Journal of Orthodontics and Dentofacial Orthopedics* 2009;136(4):481.e1-.e9.
80. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocrine Reviews* 2003;24(5):668-93.
81. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls of the Harpenden Growth Study. *Annals of Human Biology* 1976;3(2):109-126.
82. Sullivan PG. Prediction of the pubertal growth spurt by measurement of standing height. *European Journal of Orthodontics* 1983;5(3):189-97.
83. Taranger J, Hägg U. The timing and duration of adolescent growth. *Acta Odontologica Scandinavica* 1980;38(1):57-67.
84. Tanner JM, Davies PSW. Clinical longitudinal standards for height and height velocity for North American children. *The Journal of Pediatrics* 1985;107(3):317-29.
85. Tanner JM, Whitehouse RH, Takaishi M. Standards from Birth to Maturity for Height, Weight, Height Velocity, and Weight Velocity: British Children, 1965 Part II. *Archives of Disease in Childhood* 1966;41(220):613-35.
86. Ochoa BK, Nanda RS. Comparison of maxillary and mandibular growth. *American Journal of Orthodontics and Dentofacial Orthopedics* 2004;125(2):148-59.
87. Grave K, Brown T. Skeletal ossification and the adolescent growth spurt. *American Journal of Orthodontics* 1976;69(6):611-9.
88. Pike JB. A serial investigation of facial and statural growth in seven to twelve year old children. *Angle Orthodontist* 1968;38(1):63-73.
89. Baughan B, Demirjian A, Levesque GY, Lapalme-Chaput L. The pattern of facial growth before and during puberty, as shown by French-Canadian girls. *Annals of Human Biology* 1979;6(1):59-76.
90. Moore RN, Moyer BA, DuBois LM. Skeletal maturation and craniofacial growth. *American Journal of Orthodontics and Dentofacial Orthopedics* 1990;98(1):33-40.
91. Van der Beek M, Hoeksma J, Prahl-Andersen B. Vertical facial growth and statural growth in girls: a longitudinal comparison. *European Journal of Orthodontics* 1996;18(1):549-55.
92. Lewis AB, Garn SM. The relationship between tooth formation and other maturational factors. *Angle Orthodontist* 1960;30(2):70-7.
93. Lewis AB. Comparisons between dental and skeletal ages. *Angle Orthodontist* 1991;61(2):87-92.
94. Björk A, Helm S. Prediction of the age of maximum puberal growth in body height. *Angle Orthodontist* 1967;37(2):134-43.
95. Uysal T, Sari Z, Ramoglu SI, Basciftci FA. Relationships between dental and skeletal maturity in Turkish subjects. *Angle Orthodontist* 2004;74(5):657-64.
96. Nadler GL. Earlier dental maturation: Fact or fiction? *Angle Orthodontist* 1998;68(6):535-8.

97. Coutinho S, Buschang PH, Miranda F. Relationships between mandibular canine calcification stages and skeletal maturity. *American Journal of Orthodontics and Dentofacial Orthopedics* 1993;104(3):262-8.
98. Kumar S, Singla A, Sharma R, Virdi MS, Anupam A, Mittal B. Skeletal maturation evaluation using mandibular second molar calcification stages. *Angle Orthodontist* 2012;82(3):501-6.
99. Lopes LJ, de Oliveira Gamba T, Visconti MA, Ambrosano GM, Haiter-Neto F, Freitas DQ. Utility of panoramic radiography for identification of the pubertal growth period. *American Journal of Orthodontics and Dentofacial Orthopedics* 2016;149(4):509-15.
100. Baccetti T, Franchi L, De Lisa S, Giuntini V. Eruption of the maxillary canines in relation to skeletal maturity. *American Journal of Orthodontics and Dentofacial Orthopedics* 2008;133(5):748-51.
101. Perinetti G, Contardo L, Gabrieli P, Baccetti T, Di Lenarda R. Diagnostic performance of dental maturity for identification of skeletal maturation phase. *European Journal of Orthodontics* 2012;34(4):487-92.
102. Uslenghi S, Liversidge HM, Wong FS. A radiographic study of tooth development in hypodontia. *Archives of Oral Biology* 2006;51(2):129-33.
103. Demisch A, Wartmann P. Calcification of the mandibular third molar and its relation to skeletal and chronological age in children. *Child Development* 1956:459-73.
104. Morris JM, Park JH. Correlation of dental maturity with skeletal maturity from radiographic assessment: a review. *Journal of Clinical Pediatric Dentistry* 2012;37(3):309-14.
105. Hägg U, Taranger J. Menarche and voice change as indicators of the pubertal growth spurt. *Acta Odontologica Scandinavica* 1980;38(3):179-86.
106. Perinetti G, Castaldo A, Contardo L, Franchi L. Gingival crevicular fluid protein content and alkaline phosphatase activity in relation to pubertal growth phase. *Angle Orthodontist* 2012;82(6):1047-52.
107. Cox L. The biology of bone maturation and ageing. *Acta Pædiatrica* 1997;86(S423):107-8.
108. Houston W. The current status of facial growth prediction: a review. *British Journal of Orthodontics* 1979;6(1):11-17.
109. Ball G, Woodside D, Tompson B, Hunter WS, Posluns J. Relationship between cervical vertebral maturation and mandibular growth. *American Journal of Orthodontics and Dentofacial Orthopedics* 2011;139(5):e455-61.
110. Greulich WW PS. *Radiographic atlas of skeletal development of the hand and wrist*. 2nd ed. Stanford: Standford University Press Calif, 1959.
111. Tanner JM WR, Camerom N, Marshall WA, Healy MJR, Goldstein H. *Assessment of skeletal maturity and prediction of adult height (TW2 Method)*. 2nd ed. London: Academic Press, 1983.
112. Bull RK, Edwards PD, Kemp PM, Fry S, Hughes IA. Bone age assessment: A large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Archives of Disease in Childhood* 1999;81(2):172-3.
113. Milner G, Levick R, Kay R. Assessment of bone age: a comparison of the Greulich and Pyle, and the Tanner and Whitehouse methods. *Clinical Radiology* 1986;37(2):119-21.

114. Perinetti G, Contardo L. Reliability of growth indicators and efficiency of functional treatment for skeletal Class II malocclusion: Current evidence and controversies. *BioMed Research International* 2017, 1-19
115. Fishman LS. Maturational Patterns and Prediction During Adolescence. *Angle Orthodontist* 1987;57(3):178-93.
116. Chapman SM. Ossification of the adductor sesamoid and the adolescent growth spurt. *Angle Orthodontist* 1972;42(3):236-44.
117. Hägg U, Taranger J. Skeletal stages of the hand and wrist as indicators of the pubertal growth spurt. *Acta Odontologica Scandinavica* 1980;38(3):187-200.
118. Pileski RC, Woodside DG, James GA. Relationship of the ulnar sesamoid bone and maximum mandibular growth velocity. *Angle Orthodontist* 1973;43(2):162-70.
119. Patcas R, Signorelli L, Schätzle M, Peltomäki T. Is the use of the cervical vertebrae maturation method justified to determine skeletal age? A comparison of radiation dose of two strategies for skeletal age estimation. *European Journal of Orthodontics* 2013;35(5):604-9.
120. Comprehensive Advanced Life Support. *Labelled Lateral Cephalogram*. 2012. https://calsprogram.org/manual/volume2/Section10_XraySkills/03-XraySk2CerSpineXrayInterp13.html. (accessed 23 November 2016).
121. PA LBaR. *McMinn's Color Atlas of Head and Neck Anatomy*. 4th ed. Philadelphia, PA Mosby Elsevier, 2010.
122. Bench RW. Growth of the cervical vertebrae as related to tongue, face, and denture behavior. *American Journal of Orthodontics* 1963;49(3):183-214.
123. Deen K. *Morphology of the cervical vertebrae*. 2014. http://www.google.co.uk/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwihxabiwpTQAhVFVxoKHfLbCNUQjRwIBw&url=http://medicalfreakz.blogspot.com/2014/07/osteology-of-vertebrae.html&psig=AFQjCNHfPKTFJyyVwWYkZfOGuUFdiceG2A&ust=1478535095423751 (accessed 23 November 2016).
124. Vastardis H, Evans CA. Evaluation of cervical spine abnormalities on cephalometric radiographs. *American Journal of Orthodontics and Dentofacial Orthopedics* 1996;109(6):581-8.
125. Bick EM, Copel JW. Longitudinal growth of the human vertebra. *Journal of Bone and Joint Surgery (American Volume)* 1950;32(4):803-14.
126. Hellsing E. Cervical vertebral dimensions in 8-, 11-, and 15-year-old children. *Acta Odontologica Scandinavica* 1991;49(4):207-13.
127. DeLuca S, Rhea J. Radiographic anatomy of the cervical vertebrae. *Medical Radiography and Photography* 1979;56(2):18-24.
128. García-Fernandez P, Torre H, Flores L, Rea J. The cervical vertebrae as maturational indicators. *Journal of clinical orthodontics* 1998;32(4):221-5.
129. Gandini P, Mancini M, Andreani F. A comparison of hand-wrist bone and cervical vertebral analyses in measuring skeletal maturation. *Angle Orthodontist* 2006;76(6):984-9.
130. Flores-Mir C, Burgess CA, Champney M, Jensen RJ, Pitcher MR, Major PW. Correlation of skeletal maturation stages determined by cervical vertebrae and hand-wrist evaluations. *Angle Orthodontist* 2006;76(1):1-5.

131. Engel TP, Renkema AM, Katsaros C, Pazera P, Pandis N, Fudalej PS. The cervical vertebrae maturation (CVM) method cannot predict craniofacial growth in girls with Class II malocclusion. *European Journal of Orthodontists* 2016, 38(1):1-7.
132. Pancherz H, Szyska M. Analysis of cervical vetebrae instead of hand bones for determination of skeletal and somatic maturity. *Informationen Aus Orthodontie Und Kieferorthopadie* 2000;32(02):151-61.
133. Houston W, Maher R, McElroy D, Sherriff M. Sources of error in measurements from cephalometric radiographs. *European Journal of Orthodontics* 1986;8(3):149-51.
134. Baumrind S, Frantz RC. The reliability of head film measurements: 1. Landmark identification. *American Journal of Orthodontics* 1971;60(2):111-27.
135. Grave K, Townsend G. Cervical vertebral maturation as a predictor of the adolescent growth spurt. *Australian Orthodontic Journal* 2003;19(1):25-32.
136. Cericato GO, Bittencourt MAV, Paranhos LR. Validity of the assessment method of skeletal maturation by cervical vertebrae: A systematic review and meta-analysis. *Dentomaxillofacial Radiology* 2015;44(4)20140270.
137. Perinetti G, Caprioglio A, Contardo L. Visual assessment of the cervical vertebral maturation stages: A study of diagnostic accuracy and repeatability. *Angle Orthodontist* 2014;84(6):951-6.
138. Sohrabi A, Babay Ahari S, Moslemzadeh H, Rafighi A, Aghazadeh Z. The reliability of clinical decisions based on the cervical vertebrae maturation staging method. *European Journal of Orthodontics* 2015;38(1):8-12.
139. Ozer T, Kama JD, Ozer SY. A practical method for determining pubertal growth spurt. *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;130(2):131 e1-6.
140. Jaqueira LMF, Armond MC, Pereira LJ, Alcântara CEPd, Marques LS. Determining skeletal maturation stage using cervical vertebrae: evaluation of three diagnostic methods. *Brazilian oral research* 2010;24(4):433-7.
141. Predko-Engel A, Kaminek M, Langova K, Kowalski P, Fudalej PS. Reliability of the cervical vertebrae maturation (CVM) method. *Bratislava Medical Journal* 2015;116(04):222-6.
142. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Editor's Summary and Q&A. *American Journal of Orthodontics and Dentofacial Orthopedics* 2009;136(4):478-80.
143. Seedat A, Forsberg C. An evaluation of the third cervical vertebra (C3) as a growth indicator in Black subjects. *Journal of the South African Dental Association* 2005;60(4):156, 8-60.
144. Pandis N, Polychronopoulou A, Eliades T. *Sample size estimation: An overview with applications to orthodontic clinical trial designs. American Journal of Orthodontics and Dentofacial Orthopedics* 2011;140(4):e141-6
145. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies* 2011;48(6):661-71.

146. Dunn G. Design and analysis of reliability studies. *Statistical Methods in Medical Research* 1992;1(2):123-57.
147. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977:159-74.
148. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Family Medicine* 2005;37(5):360-3.
149. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 1960;20(1):37-46.
150. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70(4):213-20.
151. Warrens MJ. Equivalences of weighted kappas for multiple raters. *Statistical Methodology* 2012;9(3):407-22.
152. Warrens MJ. Cohen's linearly weighted kappa is a weighted average. *Advances in Data Analysis and Classification* 2012;6:67-79.
153. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 1973, 33(3):613-619.
154. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971;76(5):378-82.
155. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996:199-202.
156. Schuster C. A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. *Educational and Psychological Measurement* 2004;64(2):243-53.
157. Graham P, Jackson R. The analysis of ordinal agreement data: beyond weighted kappa. *Journal of Clinical Epidemiology* 1993;46(9):1055-62.
158. Vanbelle S. A New Interpretation of the Weighted Kappa Coefficients. *Psychometrika* 2016;81(2):399-410.
159. Fleiss JL, Nee JC, Landis JR. Large sample variance of kappa in the case of different sets of raters. *Psychological Bulletin* 1979;86(5):974-7.
160. Berry KJ, Mielke Jr PW. Agreement measure comparisons between two independent sets of raters. *Educational and Psychological Measurement* 1997;57(2):360-4.
161. Mielke JPW, Berry KJ. A note on Cohen's weighted kappa coefficient of agreement with linear weights. *Statistical Methodology* 2009;6:439-46.
162. Vanbelle S, Albert A. A note on the linearly weighted kappa coefficient for ordinal scales. *Statistical Methodology* 2009;6:157-63.
163. Fernando Cesar T, Marcos Shinao Y, Renata Pilli J, Luiz Renato P, Sigmar de Mello R, Danilo Furquim S. Evaluation of the Cervical Vertebrae Maturation Index in Lateral Cephalograms Taken in Different Head Positions. *Brazilian Dental Journal* 2013;24(5):462-6.
164. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? *American Journal of Roentgenology* 2004;182(4):867-9.
165. Cicchetti DV, Fleiss JL. Comparison of the Null Distributions of Weighted Kappa and the C Ordinal Statistic. *Applied Psychological Measurement* 1977;1(2):195-201.
166. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica* 2012;22(3):276-82.

167. Baccetti T, Franchi L, McNamara JA. Reproducibility of the CVM method: a reply. *American Journal of Orthodontics and Dentofacial Orthopedics* 2010;137(4):446-7.
168. Legendre P. Species Associations: The Kendall Coefficient of Concordance Revisited. *Journal of Agricultural, Biological, and Environmental Statistics* 2005;10(2):226-45.
169. Santiago RC, Cunha AR, Júnior G, Fernandes N, Campos M, Costa L, et al. New software for cervical vertebral geometry assessment and its relationship to skeletal maturation-a pilot study. *Dentomaxillofacial Radiology* 2014;43(2):20130238.
170. Padalino S, Sfondrini MF, Chenuil L, Scudeller L, Gandini P. Reliability of skeletal maturity analysis using the cervical vertebrae maturation method on dedicated software. *International Orthodontics* 2014;12(4):483-93.

Appendices

Appendix 1: Clinician Information Leaflet

Department of Orthodontics
Liverpool University Dental Hospital and School of Dentistry,
Pembroke Place,
Liverpool,
L3 5PS
26th April 2016

CLINICIAN INFORMATION SHEET

Reliability of Cervical Vertebrae Maturation (CVM) Staging Method using Full versus Cropped Lateral Cephalograms.

What is the purpose of the study?

As Orthodontists we regularly treat patients who are growing and maturing as they develop from a child, to an adolescent and then to an adult. Predicting how much patients will grow may affect the type of orthodontic appliances and/or treatment we offer them and the result they get from their treatment.

Cervical vertebral maturation (CVM) staging is a measure taken from the cervical vertebrae visible on lateral cephalograms that we use routinely. Previous research suggests that CVM staging may provide valuable information about growth, which may then allow us to predict how much growth we can expect in individual patients. This would then allow us to target treatment better and potentially, reduce the length of orthodontic treatment. Despite previous studies, controversy remains as to whether the reliability of the CVM staging tool is affected when viewing the cervical vertebrae from full or cropped lateral cephalograms.

Therefore, the primary purpose of this study is to assess how reliable the CVM staging method is, using both full and cropped lateral cephalometric radiographs. The secondary purpose is to discover if cropped images affect the reliability of the CVM staging method.

Has the study been approved?

Yes. This study has been reviewed and approved by the London Queen Square Research Ethics Committee.

Who is paying for the study?

The School of Dentistry of the University of Liverpool is paying for the study. The Royal Liverpool and Broadgreen University Hospital NHS Trust is the sponsor.

Who will be conducting the study?

The study is being led by Dr. Jayne Harrison (Consultant in Orthodontics) and carried out by Miss Julia Mangan.

Why have I been asked to take part?

We are asking all consultants and trainee orthodontists in the Mersey and North Wales region to take part. It is our aim to recruit the highest number of observers compared with similar reliability research studies. This will improve the generalisability of our findings.

What will I have to do?

You will have to attend two consecutive educational sessions of the Mersey and North Wales Deanery Audit Meeting and participate in two rounds of this study. The first phase will begin with a thirty-minute training presentation, detailing how to use the CVM staging index. You will then be shown a random sample of 144 lateral cephalograms and asked to stage each image appropriately. The second phase will take place approximately three months later, at the next Mersey and North Wales Audit meeting and you will be asked to stage the same 144 lateral cephalograms, only in a different random order.

Your individual scores will be anonymised and analysed for intra-examiner reliability, and compared with your colleagues, for inter-examiner reliability.

What happens if I don't want to take part?

Participation in this study is voluntary so please feel free to volunteer or decline participation. If you don't want to take part in the study, please feel free to leave the room at any stage. You may also withdraw at any time without explanation.

What if I have a question or there is a problem during the study?

If you have a concern about any aspect of this study, you should ask to speak to a member of the research team on 0151 706 5252 or 5030. They will do their best to answer your questions.

How will you collect and look after my data (information)?

Any data referring to you, the 'observer', will be anonymised and no one will be able to identify any of the data we collect about you. As soon as we have collected the necessary data, we will remove all information that identifies you and replace it by a coded number. Only members of the research team will have access to this information and will be solely responsible for the processing and analysis of your data. The 'observer' reliability data will be anonymised. The person responsible for security and access to your data is Dr. Jayne Harrison, the Chief Investigator of the Study. The data will be stored safely for ten years.

What do I do if I want to take part?

If you would like to take part in our study, please sign all the appropriate parts of the consent form that we will give you.

THANK YOU FOR TAKING THE TIME TO READ THIS LEAFLET.

Appendix 2: Participant Consent form

UNIVERSITY OF LIVERPOOL

Centre Number:

Study Number:

Participant ID:

CONSENT FORM

Title of Project: **Reliability of Cervical Vertebrae Maturation (CVM) Staging Method using Full versus Cropped Lateral Cephalograms.**

Name of Researcher: Julia Mangan

Please read

1. I confirm that I have read and understand the information sheet dated 26th April 2016 (version 2.0) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.

3. I understand that the data collected during the study will be analysed by the study investigators. I give permission for these individuals to analyse my results for Phase I and II of the study.

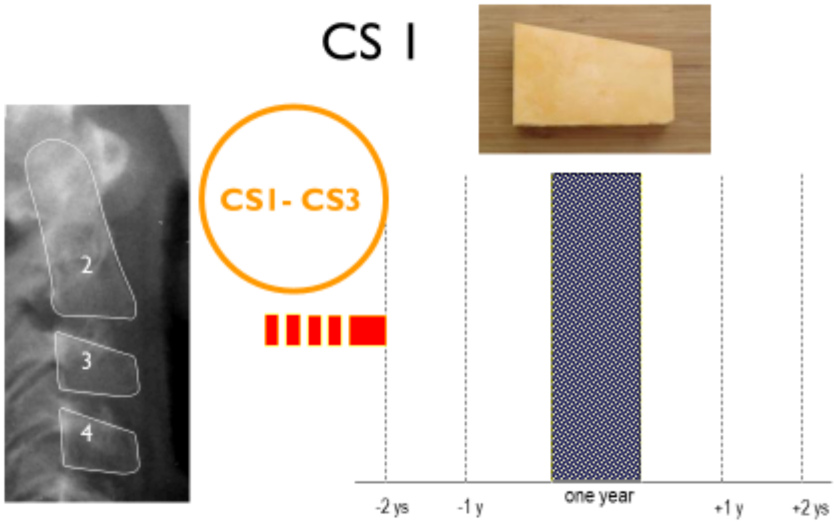
4. I agree to take part in the above study.

Name of Participant	Date	Signature

Name of Person taking consent	Date
Signature	

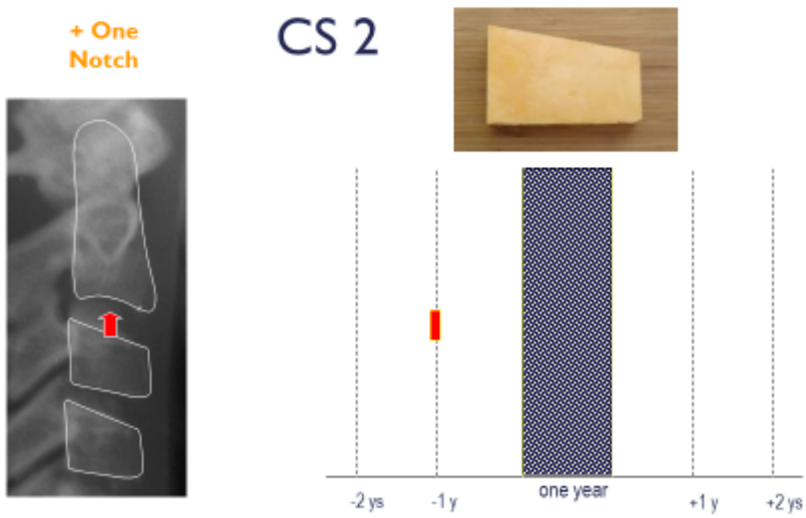
Appendix 3: CVM Training Material

UNIVERSITY OF LIVERPOOL



- All lower borders are flat (7% may show a concavity)
- C3 and C4 are trapezoid in shape

The peak interval will start not earlier than 2 years after this stage

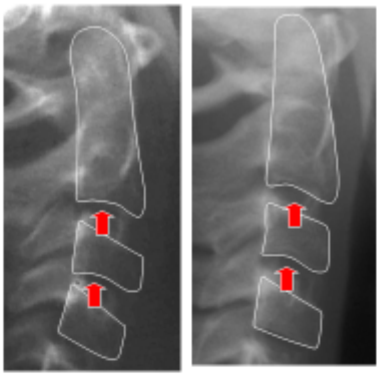


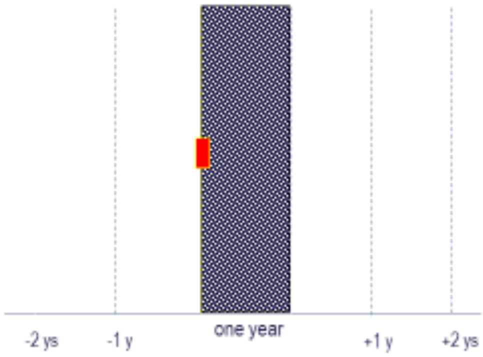
- The lower border of C2 shows a concavity (80% of the subjects)
- C3 and C4 are trapezoid in shape

The peak interval will **start 1 year after** this stage

Two
Notches

CS 3



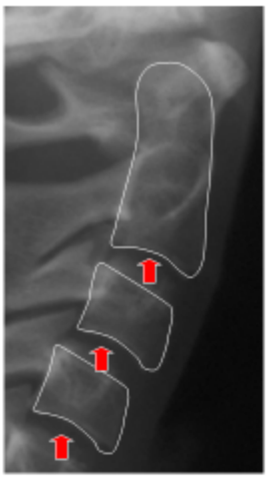


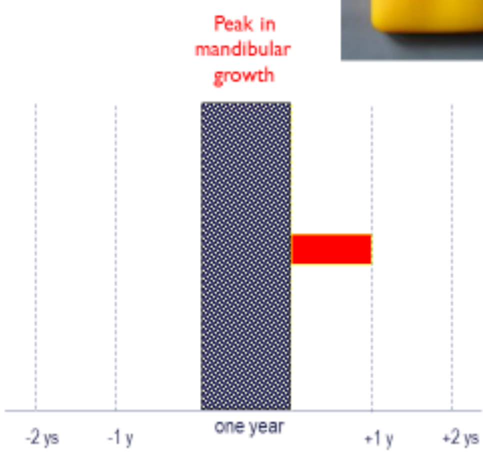
- The lower borders of C2 and C3 may show a concavity
- C3 or C4 may be trapezoid or rectangular horizontal in shape

The peak interval **starts** at this stage

Three
Notches

CS 4



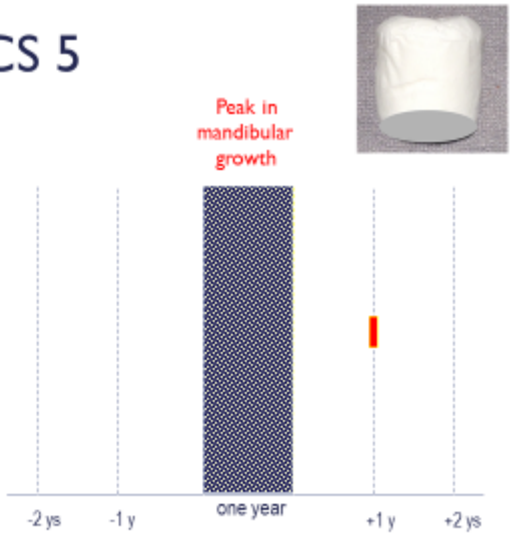


- All lower borders show concavities
- C3 and C4 are rectangular horizontal in shape

The **peak interval ends at this stage**
(or it has ended during the year before this stage)

Square "Marshmallow" Stage

CS 5

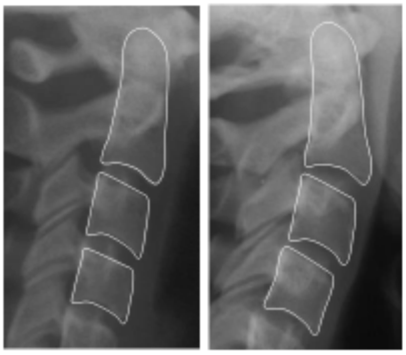


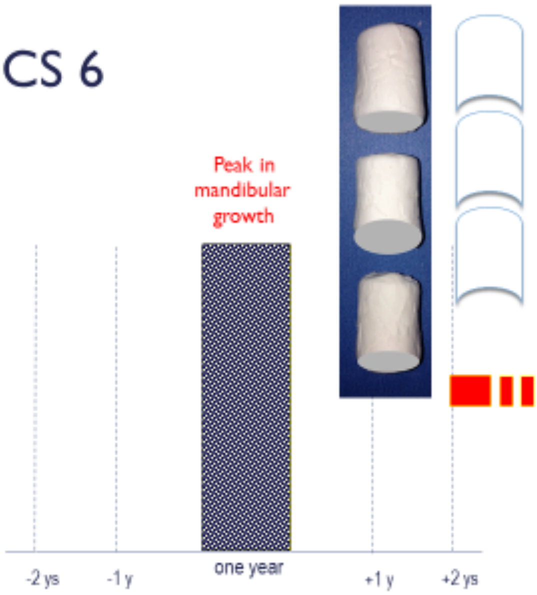
•All lower borders show concavities
•At least one of the bodies of C3 and C4 is squared in shape

The peak interval **ended one year before** this stage

Rectangular Vertical

CS 6





•All lower borders show concavities
•At least one of the bodies of C3 and C4 is rectangular vertical in shape

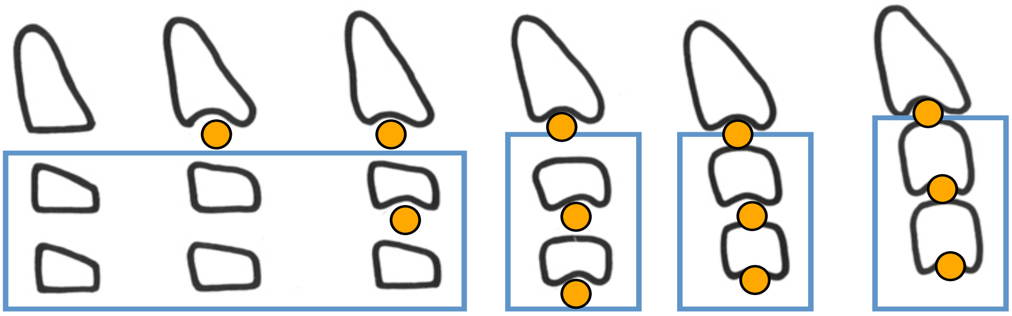
The peak interval **ended at least two years before** this stage

Appendix 4: Reference Material

UNIVERSITY OF LIVERPOOL

Guide for CVM stage determination

1. Look and count notches



Trapezoid + Notches   Rect Horiz   Square   Rect Vert

Stage I   Stage 2   Stage 3   Stage 4   Stage 5   Stage 6

2. Decide on the <u>shape</u> of C3 and C4

Trapezoid   **Horizontal Rectangular**   **Square**   **Vertical**

Rectangular

(Cheese)   **(Soap)**   **(Marshmallow)**

CS 1, 2, 3   **CS4**   **CS5**   **CS6**

Appendix 5: CVM Staging Score Sheet

UNIVERSITY OF LIVERPOOL

Reliability of Cervical Vertebrae Maturation (CVM) Staging Method using Full versus Cropped Lateral Cephalograms.

Score Sheet

Researcher: Julia Mangan

Date:

Phase: 1/2 (please circle)

Participant ID:

Image Number	CVM Stage	Notes
1		
2		
3		
4		
5		
6		

Image Number	CVM Stage	Notes
7		
8		
9		
10		
11		
12		

Image Number	CVM Stage	Notes
13		
14		
15		
16		
17		
18		

Image Number	CVM Stage	Notes
19		
20		
21		
22		
23		
24		

Image Number	CVM Stage	Notes
25		
26		
27		
28		
29		
30		

Image Number	CVM Stage	Notes
31		
32		
33		
34		
35		
36		

Image Number	CVM Stage	Notes
37		
38		
39		
40		
41		
42		

Image Number	CVM Stage	Notes
43		
44		
45		
46		
47		
48		

Image Number	CVM Stage	Notes
49		
50		
51		
52		
53		
54		

Image Number	CVM Stage	Notes
55		
56		
57		
58		
59		
60		

Image Number	CVM Stage	Notes
61		
62		
63		
64		
65		
66		

Image Number	CVM Stage	Notes
67		
68		
69		
70		
71		
72		

Image Number	CVM Stage	Notes
73		
74		
75		
76		
77		
78		

Image Number	CVM Stage	Notes
79		
80		
81		
82		
83		
84		

Image Number	CVM Stage	Notes
85		
86		
87		
88		
89		
90		

Image Number	CVM Stage	Notes
91		
92		
93		
94		
95		
96		

Image Number	CVM Stage	Notes
97		
98		
99		
100		
101		
102		

Image Number	CVM Stage	Notes
103		
104		
105		
106		
107		
108		

Image Number	CVM Stage	Notes
109		
110		
111		
112		
113		
114		

Image Number	CVM Stage	Notes
115		
116		
117		
118		
119		
120		

Image Number	CVM Stage	Notes
121		
122		
123		
124		
125		
126		

Image Number	CVM Stage	Notes
127		
128		
129		
130		
131		
132		

Image Number	CVM Stage	Notes
133		
134		
135		
136		
137		
138		

Image Number	CVM Stage	Notes
139		
140		
141		
142		
143		
144		

----------End of Reliability Test----------

THANK YOU FOR YOUR PARTICIPATION

Appendix 6: Ethical approval

NHS
Health Research Authority

National Research Ethics Service

London - Queen Square Research Ethics Committee
HRA NRES Centre Manchester
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

09 November 2015

Dr Jayne Harrison
Orthodontic Department,
Liverpool University Dental Hospital,
Pembroke Place, Liverpool
L3 5PS

Dear Dr Harrison

Study title:	**To determine the reliability and reproducibility of CVM stage determination in cropped and full lateral cephalometric images amongst orthodontists in training and specialist orthodontists.**
REC reference:	**15/LO/1660**
Protocol number:	**RD&I 5061**
IRAS project ID:	**174153**

Thank you for your letter of 6 November 2015. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our letter dated 15 September 2015

Documents received

The documents received were as follows:

Document	*Version*	*Date*
Participant consent form	2	17 October 2015
Participant information sheet (PIS)	2	17 October 2015

Approved documents

The final list of approved documentation for the study is therefore as follows:

Document	*Version*	*Date*
Other [Response to validation query]		10 September 2015
Participant consent form	2	17 October 2015
Participant information sheet (PIS)	2	17 October 2015
REC Application Form [REC_Form_09092015]		09 September 2015
Research protocol or project proposal [Study Protocol]	Version 2	14 August 2015
Summary CV for Chief Investigator (CI) [CI CV]	Version 1	13 August 2015
Summary CV for student	1	13 August 2015

You should ensure that the sponsor has a copy of the final documentation for the study. It is the sponsor's responsibility to ensure that the documentation is made available to R&D offices at all participating sites.

15/LO/1660	**Please quote this number on all correspondence**

Yours sincerely

Rachel Heron REC Manager

E-mail: nrescommittee.london-queensquare@nhs.net

Copy to: *Ms Heather Rogers, Royal Liverpool and Broadgreen University Hospital Trust*

A Research Ethics Committee established by the Health Research Authority

Appendix 7: Sponsorship

The Royal Liverpool and Broadgreen University Hospitals NHS Trust

Royal Liverpool University Hospital
Prescot Street
Liverpool
L7 8XP

Tel: 0151 706 2000
Fax: 0151 706 5806

TRUST APPROVAL LETTER FOR NON-CTIMP STUDIES

Dr Jayne E Harrison
Royal Liverpool and Broadgreen University Hospitals NHS Trust
Liverpool University Dental Hospital
Pembroke Place
Liverpool
L3 5PS

REC: 15/LO/1660
Date: 26/04/2016

Dear Dr Harrison

RD&I No: 5061
Does cropping lateral cephalograms affect CVM staging reliability?

The above study is a Data only study, sponsored by RLBUHT and with no external funding. The Trust is now happy for you to commence work on this study, using the following ethically approved documents. Please note that the table below lists only the key documents rather than all ethically approved documents. For details of all ethically approved documents please refer to the REC favourable opinion letter and the subsequent amendment letters.

Document	Version	Dated
Participant consent form	2	17 October 2015
Participant information sheet (PIS)	2	17 October 2015
Research protocol or project proposal [Study Protocol]	Version 2	14 August 2015

May I to take this opportunity to remind you of your responsibilities as PI for this study to:-

- Report SAE's as per protocol and Trust policy and record total number on OSIRIS
- Ensure that all screening and recruitment activity is updated on OSIRIS every Friday (training can be obtained if required by phoning Ext 3782)
  - Department of Health target for this study is first patient recruited by **05 July 2016**
  - Please provide a timely response to requests for information regarding achievement of this target

- For Trust sponsored studies, provide RD&I with copies of regulatory annual progress and safety reports to Ethics
- Complete and return the RD&I annual report form in a timely manner
- Comply with the Research Governance Framework 2nd Ed 2005 including but not limited to the Medicines for Human use (Clinical Trials) 2004 act plus it's appendices and the Data Protection Act 1998
- Read, disseminate to research team and acknowledge to RD&I, Trust research SOP announcements (details of relevant SOP's can be found at http://staffintranet/departments_and_services/corporate_services/research_and_development/documents/documents.aspx)
- Inform RD&I of any amendments to, or changes of status in, the study.
- Ensure any conditions to approval stipulated by the MHRA/ REC have been addressed prior to implementation of approved changes
- Maintain the study site file (if not provided by the sponsor a template is available on the Trust intranet)
- Provide copies of publications

Investigators who do not comply with the above will be dealt with in accordance with the Trust Disciplinary policy and/or will have their research stopped.

I wish you every success with your research. Please contact the RD&I Department if you require any advice on the above points.

Yours sincerely

Julia West
Operational Director RD&I

cc Head of Directorate
RLBUHT

....

I agree to the terms and conditions of the Trust research approval for RD&I **5061, Reliability of Cervical Vertebrae Maturation (CVM) Staging Method in Full versus Cropped Lateral Cephalograms** and am aware of my responsibilities under the Research Governance framework and Trust Research SOP's.

Signed: .... Dated: ....

Please return a copy of this letter to the RD&I Department, 4th Floor Linda McCartney Centre, Royal Liverpool Hospital, Prescot Street, Liverpool, L7 8XP

Appendix 8: Trainee and specialist CVM scores Phase 1 cropped images

IMAGE	ID	CROP	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	St11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Cn11
1	AC4902	C	5	6	6	6	5	5	5	6	6	6	5	6	5	5	5	5	5	6	6	6	6	6
2	AC6408	C	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4
3	AD3155	C	4	4	4	4	4	5	5	4	4	4	4	5	3	5	4	4	4	3	4	5	4	4
4	AF1739	C	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	1
5	AG0197	C	5	6	5	6	4	5	4	5	5	5	5	5	4	5	4	4	4	4	5	6	4	4
6	AH3750	C	4	5	5	5	5	6	4	5	5	5	5	4	4	5	4	4	5	4	4	4	4	4
7	AH9908	C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	AJ2214	C	6	6	5	6	5	6	6	6	5	6	5	6	5	5	6	5	6	6	6	5	6	5
9	AJ5642	C	1	3	5	3	3	3	3	1	3	1	3	1	2	2	1	3	1	3	3	3	4	3
10	AQ5118	C	3	3	3	4	4	6	3	4	4	5	4	3	3	5	4	2	4	4	4	4	3	3
11	AR4249	C	4	3	4	4	4	3	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4
12	CD0213	C	2	2	2	3	3	2	2	2	2	2	3	2	2	2	2	2	2	1	2	2	2	2
13	CD5417	C	5	6	5	6	6	5	6	5	5	5	5	6	5	6	5	5	6	5	5	5	6	5
14	CR7290	C	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	3	3	2
15	DH3042	C	2	2	2	2	2	2	2	2	2	2	3	2	2	3	2	2	3	2	2	3	2	2
16	DS2486	C	5	4	4	4	4	4	4	4	4	4	4	4	4	5	5	4	4	2	4	5	4	4
17	EA9539	C	1	2	1	2	2	2	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	3
18	EC0069	C	5	6	3	6	6	3	5	6	6	6	5	3	5	6	5	3	5	4	4	3	5	6
19	EI4661	C	5	4	4	4	4	6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
20	ES3432	C	1	2	2	1	2	2	1	2	2	2	2	1	2	2	2	1	2	1	2	2	2	1
21	ET2562	C	4	4	4	5	4	5	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	5

22	EV7611	C	4	4	5	5	4	5	4	4	4	4	4	5	4	5	4	4	4	5	4	4	4	4
23	FC9398	C	5	5	5	5	6	5	4	5	5	5	5	5	4	3	5	5	5	4	5	5	4	5
24	FM9738	C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1
25	GB1726	C	5	6	5	6	6	6	6	6	6	6	6	6	5	6	5	5	5	5	5	5	6	5
26	GC3175	C	4	4	4	4	5	4	4	4	5	4	4	3	4	5	4	4	4	4	4	4	4	4
27	GC6633	C	4	4	5	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4
28	GW5276	C	4	4	4	3	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
29	HB9067. 2	C	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5
30	HC8727	C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	HJ3507	C	4	3	4	4	6	3	4	5	3	4	3	4	4	4	4	3	3	3	4	5	4	4
32	IH4765	C	2	2	2	1	2	1	1	2	2	1	1	1	1	1	2	1	1	1	2	2	2	2
33	IP9461	C	5	5	5	5	5	5	4	6	6	5	5	4	5	5	5	4	5	5	5	5	5	5
34	JB1885	C	1	1	2	1	6	1	1	1	3	2	3	1	1	2	4	1	1	1	2	1	1	3
35	JC5483	C	2	3	3	2	3	3	2	3	3	3	3	3	3	3	3	2	2	2	2	3	3	2
36	JC6039	C	5	5	5	6	6	5	5	6	6	5	5	6	5	5	5	5	5	5	5	5	5	6
37	JF3821	C	4	4	5	4	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	4	5
38	JH0519	C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
39	JK0063	C	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
40	JP7514	C	5	6	6	6	6	6	6	6	6	6	6	5	5	6	5	5	5	5	6	6	6	6
41	JS5897	C	3	3	3	3	3	3	3	3	3	2	3	1	2	3	3	2	3	1	4	4	4	4
42	JT0057	C	5	4	5	5	5	5	4	5	5	5	4	4	4	5	4	4	4	5	4	5	4	5
43	JT2381	C	4	4	4	3	2	3	4	4	4	4	2	3	3	4	4	3	4	4	3	5	4	4
44	KR0817	C	4	4	4	5	4	6	4	4	4	4	4	3	3	4	4	3	4	4	4	4	4	4
45	KR5274	C	5	5	5	5	5	5	5	5	5	5	5	1	4	5	4	5	2	5	5	5	4	5
46	LC5760	C	5	6	5	6	6	6	6	6	6	6	6	6	6	6	5	5	6	6	6	6	5	6
47	LD5766	C	1	3	3	1	3	3	1	3	1	1	1	1	1	1	1	1	1	1	1	3	1	3

48	LH3739	C	5	4	5	5	4	5	4	4	5	4	5	3	4	5	4	3	4	5	4	5	4	5
49	LM9995	C	1	2	3	2	2	3	3	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2
50	LR6511	C	2	2	2	2	3	2	2	3	3	2	4	1	3	4	2	2	1	1	3	2	3	3
51	LT0567	C	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	1	3	3	3	3
52	LT2656	C	6	5	5	6	5	5	5	6	5	6	5	5	4	5	5	5	5	5	5	5	5	5
53	LW9018	C	4	5	5	5	4	4	4	5	5	5	4	5	4	5	4	4	4	5	4	5	4	4
54	MC1740	C	4	5	5	3	4	4	5	5	3	6	5	4	3	5	4	3	4	5	5	5	4	4
55	MC6551	C	3	2	2	5	3	2	2	2	5	2	3	2	2	3	2	2	3	3	3	5	4	3
56	MJ5271	C	4	4	4	5	5	6	5	5	5	5	5	4	4	5	4	5	4	4	5	5	4	5
57	MM5621	C	4	3	3	5	5	1	5	5	5	6	5	1	1	5	5	2	4	5	3	5	4	5
58	MM5931	C	2	2	3	3	3	3	2	2	2	3	3	2	2	3	3	1	3	2	3	3	3	3
59	MR2611	C	5	6	6	6	5	5	5	6	5	6	5	6	5	5	5	5	5	5	5	5	5	6
60	NW9845	C	5	5	5	5	5	5	5	5	6	5	5	4	5	6	5	5	5	5	5	5	5	6
61	OR1149	C	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5
62	RA7261	C	5	5	5	5	6	6	6	6	6	6	5	6	4	6	6	5	6	5	6	6	4	5
63	RB4817	C	4	5	5	5	5	4	5	5	4	5	4	5	4	5	4	4	5	4	4	5	4	5
64	RD3119	C	3	3	4	4	2	4	4	4	3	4	3	3	3	4	4	4	3	3	3	4	4	4
65	RD5450	C	4	4	4	4	4	6	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4
66	RP3260	C	4	5	5	4	5	4	4	5	5	6	4	5	4	5	4	4	5	4	5	6	5	4
67	SC4690	C	5	4	5	5	5	5	5	5	5	5	4	5	4	5	4	4	4	4	4	5	4	5
68	SM0223	C	3	3	3	4	3	3	4	4	3	4	4	3	3	2	3	3	3	4	3	3	3	4
69	SR0071	C	5	4	5	4	4	5	4	5	5	5	4	4	4	4	4	4	3	5	4	5	4	4
70	SR2649	C	5	5	4	4	5	5	5	5	5	5	4	5	4	5	5	3	5	5	5	5	4	5
71	TC5697	C	2	3	4	3	3	2	2	1	3	2	3	3	3	3	1	1	2	1	3	3	3	3
72	TW7660	C	5	3	6	3	5	3	3	3	3	3	4	3	3	4	3	3	3	2	3	6	3	6

Trainee and specialist CVM scores Phase 1 full images

Image	ID	Full	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	St11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Cn11
1	AC4902	F	5	6	6	6	6	5	6	6	6	6	5	6	6	6	5	6	6	6	5	6	6	6
2	AC6408	F	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4
3	AD3155	F	4	4	4	5	4	5	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4
4	AF1739	F	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	2	1	1	1	2	1	1
5	AG0197	F	5	5	6	5	6	5	6	6	6	5	5	6	5	6	5	4	6	5	5	5	4	5
6	AH3750	F	5	5	5	5	5	6	4	5	6	4	5	5	4	5	4	4	4	4	4	5	4	5
7	AH9908	F	1	2	1	3	1	2	2	1	1	1	1	2	1	2	2	2	1	1	1	2	1	2
8	AJ2214	F	5	6	6	6	6	5	5	6	6	6	5	6	6	6	5	6	6	6	6	5	5	6
9	AJ5642	F	3	1	1	3	1	3	4	3	3	3	4	1	3	2	1	2	1	1	3	3	4	3
10	AQ5118	F	3	3	3	4	5	5	5	5	5	5	5	3	3	5	5	3	4	5	2	4	3	3
11	AR4249	F	4	4	4	4	4	4	3	4	4	4	4	3	4	4	4	3	4	4	4	4	3	4
12	CD0213	F	2	2	2	2	2	2	2	2	2	2	3	2	3	2	2	1	2	1	3	2	2	2
13	CD5417	F	5	6	6	5	6	5	5	5	6	5	6	6	6	6	6	5	6	6	5	6	5	6
14	CR7290	F	2	2	2	2	3	3	2	2	3	2	2	2	2	3	2	2	2	2	2	3	2	2
15	DH3042	F	2	3	3	2	2	2	3	2	2	2	3	2	3	3	2	2	2	2	2	3	3	3
16	DS2486	F	5	4	5	4	5	4	5	5	4	4	5	6	5	5	5	5	5	5	5	5	4	4
17	EA9539	F	1	3	1	1	1	3	3	1	1	1	3	1	1	3	1	3	3	3	1	1	1	3
18	EC0069	F	5	6	6	6	3	3	6	6	6	6	6	3	6	6	6	4	6	6	4	5	3	6
19	EI4661	F	4	4	4	4	4	5	3	3	4	4	4	3	4	4	4	2	4	3	4	4	4	**4**
20	ES3432	F	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	2	2	1	2
21	ET2562	F	4	4	4	4	5	5	4	5	5	4	4	3	4	5	4	5	4	5	5	4	4	4
22	EV7611	F	4	5	4	4	4	4	4	4	6	4	5	5	4	5	5	4	5	5	4	4	5	4
23	FC9398	F	5	5	6	4	6	6	4	5	6	5	5	6	6	6	5	5	6	6	6	6	5	4

24	FM9738	F	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	2
25	GB1726	F	5	6	6	5	5	5	6	5	6	6	5	6	5	5	5	5	5	5	5	6	5	5
26	GC3175	F	4	4	4	4	4	4	5	4	4	4	5	4	4	5	4	4	4	5	4	4	4	4
27	GC6633	F	4	4	4	4	4	6	4	4	5	4	4	4	4	5	4	4	4	4	3	4	4	4
28	GW5276	F	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4
29	HB9067	F	5	5	6	4	5	4	5	6	6	5	5	5	5	5	5	5	5	4	5	5	5	5
30	HC8727	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	HJ3507	F	4	3	4	5	3	3	5	3	6	5	4	4	4	5	4	3	3	5	4	5	4	5
32	IH4765	F	2	2	2	2	2	2	2	2	2	2	3	2	2	2	1	1	2	2	2	2	2	2
33	IP9461	F	5	5	5	4	5	6	5	5	6	5	5	5	4	5	5	4	5	5	5	6	5	5
34	JB1885	F	1	1	4	1	1	1	1	1	1	1	1	1	1	4	1	1	1	1	2	1	1	2
35	JC5483	F	2	3	3	3	3	3	3	3	3	3	3	2	3	2	2	2	3	3	2	2	2	3
36	JC6039	F	5	5	6	6	6	5	6	5	6	5	5	6	5	5	5	6	5	5	5	6	5	6
37	JF3821	F	4	4	5	4	4	5	5	5	5	4	4	4	4	5	4	4	4	5	5	4	4	4
38	JH0519	F	2	2	2	3	2	2	2	2	2	2	2	3	2	3	2	2	2	2	2	2	2	2
39	JK0063	F	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4
40	JP7514	F	5	5	5	6	6	6	6	6	6	5	5	6	5	5	5	5	6	5	5	4	5	5
41	JS5897	F	3	4	4	3	4	3	3	4	3	3	3	1	3	4	2	2	4	2	4	4	4	3
42	JT0057	F	5	4	5	4	5	5	5	5	5	5	5	4	4	5	4	4	5	5	5	4	5	4
43	JT2381	F	4	4	4	4	4	6	4	4	4	5	4	4	4	5	5	4	3	4	4	5	4	4
44	KR0817	F	4	4	4	5	5	5	4	5	5	5	5	4	4	5	4	4	5	5	4	5	4	4
45	KR5274	F	5	5	4	5	5	5	5	6	6	5	5	6	5	6	5	5	5	5	5	4	5	5
46	LC5760	F	5	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	6	6	5	6	6	6
47	LD5766	F	1	1	1	3	3	1	3	1	3	1	3	1	3	1	2	3	1	1	3	1	1	3
48	LH3739	F	4	5	4	5	5	4	4	5	5	5	5	3	4	5	4	4	5	5	5	4	4	4
49	LM9995	F	2	2	2	3	2	3	2	2	3	2	3	2	2	3	2	2	2	2	3	3	2	2

50	LR6511	F	2	3	3	3	1	2	3	2	2	3	3	2	3	4	3	1	4	3	3	3	3	3
51	LT0567	F	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3
52	LT2656	F	5	5	5	5	6	5	5	5	6	5	5	6	5	5	5	5	5	5	5	5	5	6
53	LW9018	F	4	5	4	5	4	4	4	5	6	5	5	4	4	4	4	4	5	5	4	4	4	4
54	MC1740	F	4	5	5	5	5	4	5	5	5	5	5	4	4	5	4	4	5	5	4	5	4	5
55	MC6551	F	3	4	4	5	2	4	4	3	4	4	4	3	3	4	3	3	3	4	4	5	4	3
56	MJ5271	F	4	5	4	5	4	4	5	5	4	4	4	4	4	5	4	4	4	4	4	4	4	4
57	MM5621	F	4	3	3	5	1	1	4	5	1	5	5	4	4	5	4	4	5	4	2	2	1	4
58	MM5931	F	2	3	3	2	3	3	2	3	2	3	3	3	3	3	2	2	2	2	3	2	3	2
59	MR2611	F	5	6	6	6	5	5	6	5	5	6	6	6	5	5	6	6	6	6	5	6	6	5
60	NW9845	F	5	5	5	5	4	6	4	5	5	5	5	5	5	5	5	5	4	5	4	5	5	5
61	OR1149	F	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	6	5	5
62	RA7261	F	5	6	6	6	6	5	6	6	6	6	6	6	6	6	5	5	5	5	6	6	5	6
63	RB4817	F	4	5	4	4	5	6	5	5	6	4	4	5	4	4	5	4	4	5	5	5	4	5
64	RD3119	F	3	3	4	3	4	4	4	4	4	4	4	4	4	3	3	3	3	4	3	3	4	4
65	RD5450	F	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4
66	RP3260	F	5	6	6	5	6	6	5	5	6	6	5	6	4	5	5	5	5	4	4	5	5	5
67	SC4690	F	4	4	4	5	5	5	5	5	5	5	4	4	4	5	5	4	5	5	4	4	5	5
68	SM0223	F	3	2	3	4	2	4	3	3	2	2	3	4	2	4	2	2	2	4	2	3	3	4
69	SR0071	F	5	4	4	4	5	6	5	5	6	5	5	5	4	5	5	4	5	4	4	5	5	5
70	SR2649	F	5	5	5	5	5	4	5	5	6	6	5	5	5	5	5	5	5	5	5	5	5	5
71	TC5697	F	2	2	4	3	3	3	3	1	3	3	3	2	3	3	2	1	3	2	3	4	3	2
72	TW7660	F	3	3	3	3	3	3	3	6	3	3	3	3	3	5	4	3	3	3	3	3	3	3

Trainee and specialist CVM scores Phase 2 cropped images

IMAGE	**ID**	**Crop**	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	St11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Cn10	Cn11
1	AC4902	C	5	6	6	6	6	5	6	5	6	6	6	6	6	6	6	5	6	5	6	6	6	6
2	AC6408	C	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	2	3	3	4	4	4	4
3	AD3155	C	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4
4	AF1739	C	1	1	1	1	2	4	2	1	1	2	2	2	2	1	1	2	1	2	2	1	2	1
5	AG0197	C	5	6	5	5	6	5	6	5	5	5	6	4	5	6	5	5	5	5	5	5	6	5
6	AH3750	C	5	5	5	5	5	6	4	5	5	4	5	3	4	4	4	4	4	4	4	5	5	5
7	AH9908	C	1	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	2	1
8	AJ2214	C	5	6	5	6	6	5	6	6	5	6	6	6	5	5	5	5	3	5	5	6	5	6
9	AJ5642	C	1	1	3	3	4	3	1	4	3	1	1	4	3	1	1	3	1	2	3	3	3	3
10	AQ5118	C	1	2	3	3	4	4	3	4	3	4	5	3	3	5	1	2	3	1	3	4	4	5
11	AR4249	C	4	4	4	4	4	4	4	4	4	4	3	4	3	4	3	3	3	4	4	4	4	4
12	CD0213	C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2
13	CD5417	C	6	6	5	6	6	6	6	6	6	6	6	5	5	6	5	6	6	5	6	6	6	6
14	CR7290	C	2	2	2	2	2	2	2	2	3	2	2	2	3	2	2	2	2	2	2	3	3	2
15	DH3042	C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
16	DS2486	C	5	4	4	4	6	4	5	5	4	4	4	5	4	4	4	4	5	4	4	5	4	4
17	EA9539	C	1	1	1	3	1	1	3	1	2	2	2	2	3	1	1	1	1	1	2	1	1	3
18	EC0069	C	5	6	5	6	3	6	5	6	6	6	6	6	5	6	5	5	3	5	5	5	5	6
19	EI4661	C	4	4	4	4	4	6	4	4	4	4	4	4	4	3	4	4	4	4	4	5	4	4
20	ES3432	C	1	1	2	1	2	2	2	2	2	1	2	2	2	1	1	1	2	1	2	2	2	2
21	ET2562	C	5	5	4	5	5	6	5	5	5	5	5	4	4	5	4	4	4	4	5	5	4	4
22	EV7611	C	5	5	5	5	5	5	4	4	5	4	5	5	4	5	4	4	4	5	4	4	4	5
23	FC9398	C	5	5	6	4	6	6	5	5	6	6	6	5	5	5	4	4	5	5	5	6	6	5

24	FM9738	C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
25	GB1726	C	5	6	5	6	6	6	6	5	6	6	6	6	5	6	5	5	5	5	5	6	6	5
26	GC3175	C	5	4	4	5	4	4	4	4	4	4	4	4	4	5	3	3	4	4	3	5	4	5
27	GC6633	C	5	4	4	4	4	5	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	5
28	GW527 6	C	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4
29	HB9067	C	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	4	5	5
30	HC8727	C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	HJ3507	C	1	1	4	4	3	6	4	4	3	4	5	3	4	5	4	3	3	1	3	4	4	5
32	IH4765	C	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	1	1	2	2	2	2	2
33	IP9461	C	5	5	5	5	5	6	5	5	5	5	5	5	4	5	4	4	5	5	5	5	4	5
34	JB1885	C	1	1	1	1	1	1	4	1	3	1	1	1	1	1	1	1	1	1	1	1	1	3
35	JC5483	C	2	1	2	2	3	3	2	2	2	2	2	2	3	2	2	2	2	2	3	3	2	2
36	JC6039	C	6	6	5	6	5	5	5	5	6	5	5	5	5	6	5	5	5	5	5	6	5	6
37	JF3821	C	4	4	4	4	4	6	4	5	5	4	4	5	4	5	4	5	4	5	4	6	4	5
38	JH0519	C	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	2
39	JK0063	C	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
40	JP7514	C	6	5	6	6	6	6	6	5	6	5	6	5	6	6	5	5	6	6	6	6	6	6
41	JS5897	C	2	2	3	2	3	3	3	3	3	3	2	1	2	2	2	1	1	1	3	3	4	1
42	JT0057	C	5	5	5	5	5	6	4	4	5	5	4	6	4	5	4	4	4	5	5	5	4	4
43	JT2381	C	3	4	4	4	1	6	4	3	3	4	3	3	2	3	3	3	1	3	2	4	4	5
44	KR0817	C	4	3	4	4	4	6	4	4	4	5	4	4	4	3	4	3	4	5	4	5	4	5
45	KR5274	C	5	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	5	6
46	LC5760	C	5	6	6	6	6	6	6	6	6	6	5	6	6	6	5	5	6	5	6	6	5	3
47	LD5766	C	1	1	3	1	3	3	3	1	3	1	3	1	3	1	1	1	1	1	3	3	4	5
48	LH3739	C	4	5	5	5	4	5	4	4	5	5	5	4	3	5	4	3	4	4	4	5	4	2
49	LM9995	C	1	1	2	3	2	1	1	2	2	2	2	1	2	2	1	1	1	1	2	2	1	3

50	LR6511	C	2	1	3	3	3	3	2	3	3	2	3	2	3	1	2	3	1	1	3	3	2	3
51	LT0567	C	3	2	3	3	3	3	3	3	3	3	3	3	3	2	1	2	1	2	3	3	3	6
52	LT2656	C	5	5	5	5	5	5	5	5	6	5	5	6	5	5	5	5	5	5	5	5	5	5
53	LW9018	C	5	5	4	5	5	5	5	5	5	5	5	5	4	5	4	4	3	4	5	5	4	3
54	MC1740	C	5	3	4	4	5	6	4	5	4	5	5	4	5	4	4	2	4	4	4	4	4	3
55	MC6551	C	2	3	4	4	2	6	4	3	4	3	4	4	3	3	4	2	2	4	4	4	3	5
56	MJ5271	C	5	4	4	5	4	6	4	4	6	5	5	4	4	5	4	3	4	4	5	5	4	5
57	MM5621	C	5	1	3	5	5	1	5	5	5	5	5	5	4	4	1	2	5	1	1	2	5	3
58	MM5931	C	3	2	3	3	2	3	3	2	3	2	3	3	3	3	3	3	2	2	2	3	3	6
59	MR2611	C	5	6	6	6	6	5	5	5	5	5	6	6	5	5	5	3	5	5	5	6	6	6
60	NW9845	C	5	5	5	5	5	6	5	5	6	5	5	5	5	6	5	5	5	5	6	6	5	5
61	OR1149	C	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
62	RA7261	C	5	5	5	5	6	5	5	5	6	5	5	5	6	5	4	5	5	5	5	6	6	5
63	RB4817	C	4	5	5	5	5	5	5	4	5	5	5	4	4	5	4	4	4	4	5	6	5	4
64	RD3119	C	3	3	4	4	4	4	4	3	3	3	4	4	3	4	3	3	3	2	3	4	4	4
65	RD5450	C	5	4	4	4	4	6	4	4	5	4	4	5	4	3	4	1	4	1	4	4	4	4
66	RP3260	C	4	4	5	4	5	4	4	5	4	5	5	5	4	4	4	5	4	4	5	5	6	4
67	SC4690	C	5	4	4	5	5	6	4	4	5	5	4	5	4	5	4	4	4	4	4	5	4	3
68	SM0223	C	3	3	3	4	3	3	4	2	3	4	4	4	3	3	3	2	2	3	4	3	3	5
69	SR0071	C	5	4	4	4	4	6	4	4	3	4	4	4	4	4	4	4	5	5	4	4	5	5
70	SR2649	C	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	4	3
71	TC5697	C	2	2	3	3	3	4	3	3	3	3	2	3	3	2	2	1	2	1	3	4	3	3
72	TW7660	C	5	3	5	3	5	3	3	3	3	4	4	3	3	3	3	1	2	5	3	3	3	3

Trainee and specialist CVM scores Phase 2 full images

Image	ID	Full	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10	St11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Cn10	Cn11
1	AC4902	F	5	6	6	5	6	5	6	6	6	6	6	6	5	6	5	6	6	5	5	6	6	6
2	AC6408	F	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	2	3	3	4	4	4	4
3	AD3155	F	4	4	4	5	4	4	4	5	4	4	4	4	4	5	4	4	4	5	4	5	4	4
4	AF1739	F	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	2	1	1	1
5	AG0197	F	5	6	4	5	6	5	6	6	6	5	6	6	5	6	5	5	5	5	6	5	5	6
6	AH3750	F	4	5	5	4	5	6	5	5	5	4	5	5	4	5	5	4	4	5	4	6	4	5
7	AH9908	F	1	1	2	2	1	1	2	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1
8	AJ2214	F	5	6	6	5	6	5	6	6	6	6	6	6	5	6	5	6	6	6	5	5	5	6
9	AJ5642	F	3	1	4	3	1	3	4	3	3	2	1	5	3	3	3	2	3	3	3	3	1	3
10	AQ5118	F	3	3	3	4	2	4	4	4	5	5	4	4	3	3	5	2	3	2	3	3	4	4
11	AR4249	F	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	3	4	4	4	4
12	CD0213	F	2	2	2	2	2	2	2	2	2	2	3	2	3	4	2	2	2	1	3	2	2	2
13	CD5417	F	5	5	5	6	6	5	6	6	6	6	6	6	5	6	5	5	6	5	5	6	6	6
14	CR7290	F	2	2	2	2	2	3	2	2	3	2	3	2	2	2	2	2	2	2	2	2	3	2
15	DH3042	F	2	2	3	2	2	2	2	2	2	2	2	2	3	5	2	2	2	2	2	3	3	3
16	DS2486	F	4	4	5	4	5	4	5	4	4	5	4	4	4	5	4	4	4	5	4	5	4	4
17	EA9539	F	1	1	3	3	1	3	3	1	1	5	5	3	3	3	1	2	3	3	1	1	3	1
18	EC0069	F	6	5	6	6	5	5	6	6	6	6	5	6	6	6	6	5	6	5	3	3	5	6
19	EI4661	F	4	4	4	4	4	4	4	4	4	3	3	5	4	4	3	3	4	4	4	4	3	4
20	ES3432	F	1	1	1	1	1	2	2	1	1	1	3	1	2	1	1	1	1	1	1	2	1	1
21	ET2562	F	4	4	4	5	4	5	4	4	6	4	5	4	4	5	4	5	4	5	4	5	4	4
22	EV7611	F	4	4	4	5	5	5	4	5	5	4	4	4	4	5	4	4	5	4	4	5	4	4
23	FC9398	F	4	5	5	4	5	6	5	5	6	6	5	6	6	5	5	5	6	5	6	6	5	5

24	FM9738	F	1	1	2	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1
25	GB1726	F	6	6	6	5	6	6	5	5	6	5	5	4	5	6	5	5	6	5	5	5	5	5
26	GC3175	F	4	4	4	4	4	4	5	4	5	4	5	4	4	4	4	4	4	4	4	5	4	4
27	GC6633	F	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4
28	GW5276	F	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
29	HB9067	F	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	4	5
30	HC8727	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	HJ3507	F	4	5	4	4	4	4	4	5	3	5	4	3	4	5	4	4	3	5	3	5	3	4
32	IH4765	F	1	1	2	1	1	2	2	1	1	2	2	2	2	1	1	1	1	1	2	2	1	2
33	IP9461	F	5	5	5	4	5	6	5	5	5	5	6	4	4	6	4	5	5	5	5	6	4	5
34	JB1885	F	1	1	3	4	4	4	3	4	4	1	4	1	6	5	1	4	1	1	4	1	2	4
35	JC5483	F	2	2	3	2	3	3	2	1	3	2	3	2	3	2	2	2	2	2	3	3	2	2
36	JC6039	F	5	5	6	6	6	5	5	5	6	5	5	6	6	6	5	6	6	5	5	5	5	6
37	JF3821	F	4	5	5	4	4	5	4	5	5	4	5	5	4	6	4	4	5	4	4	5	5	5
38	JH0519.9	F	2	1	2	2	2	2	2	2	2	2	4	2	4	3	2	2	3	2	2	2	3	2
39	JK0063	F	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	3	4	4	5	4	4
40	JP7514	F	5	5	6	5	5	5	5	5	6	5	6	5	5	5	5	5	4	6	5	5	4	5
41	JS5897	F	2	2	4	3	3	4	3	4	4	3	4	2	4	3	2	2	2	1	3	3	4	4
42	JT0057	F	5	5	5	5	5	5	5	6	6	5	4	4	5	5	4	5	5	5	4	6	4	5
43	JT2381	F	4	4	4	4	4	4	4	4	4	4	3	6	4	3	4	4	2	3	4	4	4	4
44	KR0817	F	4	4	5	4	4	5	4	5	5	4	6	4	4	5	4	4	5	5	4	4	4	5
45	KR5274	F	5	6	5	5	5	5	5	5	6	5	5	6	5	6	5	5	5	5	5	4	5	5
46	LC5760	F	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	6	5	6	6	6	6
47	LD5766	F	1	1	4	1	1	2	3	1	1	3	1	3	3	1	4	1	2	1	1	2	1	3
48	LH3739	F	4	4	4	5	4	4	4	4	4	5	4	4	4	5	4	4	4	4	4	5	4	4
49	LM9995	F	3	2	2	3	3	3	2	2	3	2	3	2	2	4	2	2	2	1	3	2	3	2

50	LR6511	F	2	2	3	2	1	3	3	3	3	2	3	3	3	5	4	3	3	1	3	3	3	3
51	LT0567	F	3	3	2	3	3	3	3	3	3	3	3	3	3	4	3	2	2	2	3	3	3	3
52	LT2656	F	5	6	5	5	5	5	6	5	6	6	5	5	5	5	5	5	5	5	5	6	5	6
53	LW9018	F	4	5	4	5	5	6	5	5	6	5	5	4	4	5	4	4	5	4	4	4	4	5
54	MC1740	F	5	4	4	5	5	5	5	5	6	6	5	4	4	5	4	4	5	4	4	5	5	4
55	MC6551	F	3	3	4	4	5	6	4	3	5	5	5	4	4	3	3	4	6	4	4	6	3	4
56	MJ5271	F	4	5	4	4	5	5	5	5	6	5	5	4	4	5	5	4	5	4	4	4	4	4
57	MM5621	F	5	1	5	5	5	3	5	5	3	5	5	3	4	5	4	5	5	1	3	6	3	6
58	MM5931	F	2	2	2	3	2	3	3	3	3	3	3	2	3	2	2	2	2	2	2	3	3	3
59	MR2611	F	6	6	5	5	6	5	5	5	6	6	5	6	6	5	5	5	6	5	5	6	5	6
60	NW9845	F	5	5	6	5	5	5	5	5	6	5	5	5	5	5	5	6	5	5	5	5	4	6
61	OR1149	F	4	5	4	5	5	5	5	5	5	5	5	5	4	5	4	5	5	5	4	5	4	5
62	RA7261	F	6	5	6	5	6	6	6	6	6	6	6	6	6	6	5	6	6	5	6	6	6	6
63	RB4817	F	4	4	4	5	4	4	5	5	4	5	5	4	4	4	4	4	4	4	5	4	3	4
64	RD3119	F	3	2	4	3	4	4	4	2	4	3	4	4	3	4	3	3	2	2	4	4	4	4
65	RD5450	F	4	4	4	4	4	6	4	4	5	4	4	4	4	3	4	3	4	1	4	4	4	4
66	RP3260	F	5	5	6	4	6	5	5	6	6	5	6	4	5	5	6	6	5	4	5	5	4	4
67	SC4690	F	4	4	4	5	4	4	4	4	5	5	5	5	4	5	4	4	4	4	4	4	5	4
68	SM0223	F	3	3	4	4	3	5	4	3	3	4	4	4	4	4	3	3	4	3	4	4	3	3
69	SR0071	F	4	4	4	4	4	4	4	5	5	5	4	4	4	5	5	4	4	5	4	6	5	4
70	SR2649	F	5	5	5	5	5	4	5	5	6	5	5	6	4	5	5	5	6	5	5	5	5	5
71	TC5697	F	2	2	4	4	3	4	3	1	3	2	3	3	3	3	1	1	3	1	3	3	3	1
72	TW7660	F	6	3	6	3	4	3	5	6	3	5	3	3	6	6	5	3	3	5	3	6	3	3

Appendix 9: Linear weighted kappa trainees and specialists Phase 1 cropped images

	SR1	SR2	SR3	SR4	SR5	SR6	SR7	SR8	SR9	SR10	SR11	Cn1	Cn2	Cn3	Cn4	Cn5	Cn6	Cn7	Cn8	Cn9	Cn10	Cn11
SR1		0.7	0.68	0.68	0.6	0.63	0.74	0.72	0.66	0.72	0.67	0.65	0.68	0.65	0.76	0.66	0.75	0.72	0.7	0.64	0.68	0.7
SR2	0.7		0.74	0.74	0.69	0.67	0.75	0.8	0.7	0.76	0.7	0.71	0.7	0.68	0.67	0.64	0.74	0.67	0.76	0.65	0.73	0.69
SR3	0.68	0.74		0.63	0.6	0.65	0.65	0.71	0.64	0.71	0.64	0.61	0.56	0.65	0.61	0.59	0.6	0.61	0.69	0.71	0.64	0.7
SR4	0.68	0.74	0.63		0.65	0.65	0.73	0.74	0.73	0.75	0.71	0.66	0.6	0.69	0.6	0.61	0.66	0.66	0.67	0.64	0.66	0.73
SR5	0.6	0.69	0.6	0.65		0.57	0.65	0.71	0.7	0.66	0.72	0.54	0.56	0.7	0.63	0.52	0.65	0.52	0.7	0.63	0.61	0.59
SR6	0.63	0.67	0.65	0.65	0.57		0.68	0.66	0.63	0.67	0.63	0.61	0.61	0.62	0.6	0.63	0.63	0.6	0.68	0.63	0.54	0.72
SR7	0.74	0.75	0.65	0.73	0.65	0.68		0.77	0.67	0.78	0.73	0.7	0.68	0.74	0.75	0.72	0.77	0.74	0.76	0.65	0.72	0.73
SR8	0.72	0.8	0.71	0.74	0.71	0.66	0.77		0.77	0.88	0.69	0.67	0.6	0.73	0.73	0.59	0.72	0.69	0.72	0.7	0.65	0.72
StR9	0.66	0.7	0.64	0.73	0.7	0.63	0.67	0.77		0.73	0.73	0.58	0.59	0.74	0.65	0.57	0.66	0.65	0.71	0.71	0.66	0.69
SR10	0.72	0.76	0.71	0.75	0.66	0.67	0.78	0.88	0.73		0.72	0.68	0.62	0.77	0.74	0.6	0.74	0.7	0.73	0.68	0.65	0.7
SR11	0.67	0.7	0.64	0.71	0.72	0.63	0.73	0.69	0.73	0.72		0.55	0.66	0.76	0.69	0.63	0.72	0.67	0.76	0.62	0.66	0.55
Con1	0.65	0.71	0.61	0.66	0.54	0.61	0.7	0.67	0.58	0.68	0.55		0.68	0.61	0.62	0.65	0.65	0.63	0.64	0.56	0.61	0.55
Con2	0.68	0.7	0.56	0.6	0.56	0.61	0.68	0.6	0.59	0.62	0.66	0.68		0.58	0.77	0.73	0.68	0.64	0.68	0.49	0.7	0.58
Con3	0.65	0.68	0.65	0.69	0.7	0.62	0.74	0.73	0.74	0.77	0.76	0.61	0.58		0.66	0.54	0.7	0.64	0.68	0.68	0.63	0.67
Con4	0.76	0.67	0.61	0.6	0.63	0.6	0.75	0.73	0.65	0.74	0.69	0.62	0.77	0.66		0.66	0.79	0.67	0.72	0.59	0.72	0.64
Con5	0.66	0.64	0.59	0.61	0.52	0.63	0.72	0.59	0.57	0.6	0.63	0.65	0.73	0.54	0.66		0.65	0.67	0.67	0.52	0.6	0.6
Con6	0.75	0.74	0.6	0.66	0.65	0.63	0.77	0.72	0.66	0.74	0.72	0.65	0.68	0.7	0.79	0.65		0.71	0.75	0.61	0.71	0.62
Con7	0.72	0.67	0.61	0.66	0.52	0.6	0.74	0.69	0.65	0.7	0.67	0.63	0.64	0.64	0.67	0.67	0.71		0.68	0.6	0.61	0.63
Con8	0.7	0.76	0.69	0.67	0.7	0.68	0.76	0.72	0.71	0.73	0.76	0.64	0.68	0.68	0.72	0.67	0.75	0.68		0.69	0.75	0.71
Con9	0.64	0.65	0.71	0.64	0.63	0.63	0.65	0.7	0.71	0.68	0.62	0.56	0.49	0.68	0.59	0.52	0.61	0.6	0.69		0.62	0.7
Con10	0.68	0.73	0.64	0.66	0.61	0.54	0.72	0.65	0.66	0.65	0.66	0.61	0.7	0.63	0.72	0.6	0.71	0.61	0.75	0.62		0.63
Con11	0.7	0.69	0.7	0.73	0.73	0.59	0.72	0.73	0.72	0.69	0.7	0.55	0.58	0.67	0.64	0.6	0.62	0.63	0.71	0.7	0.63	

Appendix 10: Linear weighted kappa trainees and specialists Phase 1 full images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Cons1	Cons2	Cons3	Cons4	Cons5	Cons6	Cons7	Cons8	Con9	C0n10	C0n11
StR 1		0.71	0.71	0.66	0.67	0.64	0.68	0.73	0.73	0.74	0.68	0.72	0.77	0.58	0.8	0.73	0.72	0.67	0.71	0.69	0.79	0.73
StR2	0.71		0.79	0.7	0.7	0.57	0.7	0.72	0.64	0.77	0.71	0.72	0.73	0.65	0.76	0.67	0.79	0.74	0.65	0.69	0.72	0.75
StR3	0.71	0.79		0.62	0.71	0.53	0.69	0.69	0.64	0.73	0.66	0.72	0.74	0.66	0.66	0.58	0.72	0.69	0.68	0.72	0.71	0.7
StR4	0.66	0.7	0.62		0.63	0.55	0.72	0.69	0.64	0.76	0.67	0.62	0.68	0.63	0.66	0.61	0.7	0.68	0.58	0.66	0.62	0.74
StR5	0.67	0.7	0.71	0.63		0.68	0.67	0.75	0.76	0.71	0.63	0.73	0.69	0.61	0.68	0.63	0.76	0.67	0.69	0.65	0.69	0.67
StR6	0.64	0.57	0.53	0.55	0.68		0.57	0.61	0.67	0.61	0.53	0.58	0.51	0.57	0.58	0.5	0.57	0.6	0.57	0.58	0.64	0.55
StR7	0.68	0.7	0.69	0.72	0.67	0.57		0.72	0.65	0.74	0.72	0.66	0.68	0.63	0.71	0.64	0.71	0.73	0.78	0.66	0.7	0.77
StR8	0.73	0.72	0.69	0.69	0.75	0.61	0.72		0.69	0.77	0.67	0.68	0.67	0.7	0.69	0.62	0.76	0.73	0.62	0.6	0.66	0.67
StR9	0.73	0.64	0.64	0.64	0.76	0.67	0.65	0.69		0.73	0.62	0.62	0.61	0.6	0.6	0.52	0.65	0.65	0.6	0.66	0.6	0.63
StR10	0.74	0.77	0.73	0.76	0.71	0.61	0.74	0.77	0.73		0.77	0.66	0.74	0.66	0.75	0.62	0.78	0.75	0.66	0.7	0.75	0.71
StR11	0.68	0.71	0.66	0.67	0.63	0.53	0.72	0.67	0.62	0.77		0.57	0.73	0.63	0.68	0.56	0.72	0.7	0.7	0.64	0.68	0.7
Con1	0.72	0.72	0.72	0.62	0.73	0.58	0.66	0.68	0.62	0.66	0.57		0.68	0.61	0.62	0.65	0.72	0.63	0.64	0.56	0.61	0.55
Con2	0.77	0.73	0.74	0.68	0.69	0.51	0.68	0.67	0.61	0.74	0.73	0.68		0.58	0.67	0.73	0.68	0.64	0.68	0.49	0.7	0.58
Con3	0.58	0.65	0.66	0.63	0.61	0.57	0.63	0.7	0.6	0.66	0.63	0.61	0.58		0.66	0.54	0.7	0.64	0.68	0.68	0.63	0.67
Con4	0.8	0.76	0.66	0.66	0.68	0.58	0.71	0.69	0.6	0.75	0.68	0.62	0.67	0.66		0.66	0.79	0.67	0.72	0.59	0.72	0.64
Con5	0.73	0.67	0.58	0.61	0.63	0.5	0.64	0.62	0.52	0.62	0.56	0.65	0.73	0.54	0.66		0.65	0.67	0.67	0.52	0.6	0.6
Con6	0.72	0.79	0.72	0.7	0.76	0.57	0.71	0.76	0.65	0.78	0.72	0.72	0.68	0.7	0.79	0.65		0.71	0.75	0.61	0.71	0.62
Con7	0.67	0.74	0.69	0.68	0.67	0.6	0.73	0.73	0.65	0.75	0.7	0.63	0.64	0.64	0.67	0.67	0.71		0.68	0.6	0.61	0.63
Con8	0.71	0.65	0.68	0.58	0.69	0.57	0.78	0.62	0.6	0.64	0.63	0.64	0.68	0.68	0.72	0.67	0.75	0.68		0.69	0.75	0.71
Con9	0.69	0.69	0.72	0.66	0.65	0.58	0.66	0.6	0.66	0.7	0.64	0.56	0.49	0.68	0.59	0.52	0.61	0.6	0.69		0.62	0.7
Cns10	0.79	0.72	0.71	0.62	0.69	0.64	0.7	0.66	0.6	0.75	0.68	0.61	0.7	0.63	0.72	0.6	0.71	0.61	0.75	0.62		0.63
Cns11	0.73	0.75	0.7	0.74	0.67	0.55	0.77	0.67	0.63	0.71	0.7	0.55	0.58	0.67	0.64	0.6	0.62	0.63	0.71	0.7	0.63	

Appendix 11: Quadratic weight kappa trainees and specialists Phase 1 cropped images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Con11
StR 1		0.87	0.81	0.86	0.75	0.78	0.88	0.89	0.85	0.87	0.84	0.8	0.83	0.85	0.88	0.82	0.87	0.84	0.86	0.83	0.84	0.86
StR2	0.87		0.85	0.86	0.76	0.81	0.89	0.9	0.84	0.88	0.84	0.82	0.85	0.85	0.81	0.8	0.86	0.83	0.89	0.79	0.87	0.84
StR3	0.81	0.85		0.77	0.74	0.78	0.81	0.79	0.77	0.78	0.8	0.74	0.75	0.79	0.72	0.76	0.74	0.76	0.83	0.85	0.8	0.84
StR4	0.86	0.86	0.77		0.77	0.78	0.87	0.87	0.89	0.87	0.86	0.76	0.77	0.86	0.78	0.76	0.84	0.83	0.84	0.8	0.86	0.85
StR5	0.75	0.76	0.74	0.77		0.66	0.76	0.8	0.82	0.79	0.83	0.65	0.67	0.8	0.79	0.64	0.73	0.67	0.79	0.72	0.7	0.85
StR6	0.78	0.81	0.78	0.78	0.66		0.79	0.79	0.76	0.77	0.76	0.76	0.76	0.77	0.73	0.76	0.77	0.77	0.83	0.74	0.71	0.71
StR7	0.88	0.89	0.81	0.87	0.76	0.79		0.9	0.84	0.9	0.87	0.81	0.81	0.89	0.87	0.85	0.89	0.88	0.9	0.8	0.87	0.85
StR8	0.89	0.9	0.79	0.87	0.8	0.79	0.9		0.86	0.95	0.84	0.78	0.78	0.88	0.87	0.77	0.85	0.85	0.86	0.82	0.81	0.85
StR9	0.85	0.84	0.77	0.89	0.82	0.76	0.84	0.86		0.86	0.87	0.71	0.76	0.88	0.83	0.74	0.82	0.82	0.85	0.82	0.84	0.85
StR10	0.87	0.88	0.78	0.87	0.79	0.77	0.9	0.95	0.86		0.87	0.78	0.75	0.9	0.88	0.74	0.87	0.87	0.86	0.81	0.82	0.82
StR11	0.84	0.84	0.8	0.86	0.83	0.76	0.87	0.84	0.87	0.87		0.71	0.79	0.88	0.84	0.77	0.82	0.81	0.89	0.77	0.84	0.85
Con1	0.8	0.82	0.74	0.76	0.65	0.76	0.81	0.78	0.71	0.78	0.71		0.83	0.73	0.76	0.81	0.81	0.77	0.79	0.7	0.74	0.67
Con2	0.83	0.85	0.75	0.77	0.67	0.76	0.81	0.78	0.76	0.75	0.79	0.83		0.77	0.77	0.86	0.83	0.78	0.83	0.65	0.82	0.74
Con3	0.85	0.85	0.79	0.86	0.8	0.77	0.89	0.88	0.88	0.9	0.88	0.73	0.77		0.84	0.72	0.84	0.8	0.85	0.83	0.83	0.83
Con4	0.88	0.81	0.72	0.78	0.79	0.73	0.87	0.87	0.83	0.88	0.84	0.76	0.77	0.84		0.77	0.89	0.82	0.85	0.73	0.8	0.8
Con5	0.82	0.8	0.76	0.76	0.64	0.76	0.85	0.77	0.74	0.74	0.77	0.81	0.86	0.72	0.77		0.8	0.82	0.82	0.68	0.78	0.73
Con6	0.87	0.86	0.74	0.84	0.73	0.77	0.89	0.85	0.82	0.87	0.82	0.81	0.83	0.84	0.89	0.8		0.85	0.86	0.77	0.84	0.78
Con7	0.84	0.83	0.76	0.83	0.67	0.77	0.88	0.85	0.82	0.87	0.81	0.77	0.78	0.8	0.82	0.82	0.85		0.83	0.74	0.8	0.76
Con8	0.86	0.89	0.83	0.84	0.79	0.83	0.9	0.86	0.85	0.86	0.89	0.79	0.83	0.85	0.85	0.82	0.86	0.83		0.82	0.88	0.83
Con9	0.83	0.79	0.85	0.8	0.72	0.74	0.8	0.82	0.82	0.81	0.77	0.7	0.65	0.83	0.73	0.68	0.77	0.74	0.82		0.79	0.82
Con10	0.84	0.87	0.8	0.86	0.7	0.71	0.87	0.81	0.84	0.82	0.84	0.74	0.82	0.83	0.8	0.78	0.84	0.8	0.88	0.79		0.8
Con11	0.86	0.84	0.84	0.85	0.85	0.71	0.85	0.85	0.85	0.82	0.85	0.67	0.74	0.83	0.8	0.73	0.78	0.76	0.83	0.82	0.8	

Appendix 12: Quadratic weight kappa trainees and specialists Phase 1 full images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Con11
StR 1		0.71	0.85	0.85	0.83	0.8	0.85	0.88	0.88	0.89	0.85	0.88	0.9	0.8	0.91	0.87	0.88	0.86	0.87	0.86	0.89	0.87
StR2	0.71		0.89	0.84	0.84	0.78	0.84	0.86	0.82	0.88	0.84	0.84	0.87	0.82	0.89	0.84	0.92	0.88	0.82	0.86	0.84	0.87
StR3	0.85	0.89		0.79	0.83	0.73	0.81	0.81	0.8	0.86	0.79	0.84	0.86	0.84	0.84	0.76	0.87	0.83	0.84	0.87	0.82	0.84
StR4	0.85	0.84	0.79		0.75	0.71	0.88	0.83	0.79	0.89	0.85	0.8	0.84	0.81	0.85	0.79	0.83	0.85	0.76	0.82	0.77	0.88
StR5	0.83	0.84	0.83	0.75		0.84	0.79	0.83	0.86	0.82	0.76	0.85	0.82	0.73	0.84	0.81	0.84	0.8	0.84	0.82	0.84	0.79
StR6	0.8	0.78	0.73	0.71	0.84		0.74	0.75	0.81	0.76	0.71	0.78	0.69	0.73	0.77	0.71	0.72	0.77	0.73	0.81	0.82	0.73
StR7	0.85	0.84	0.81	0.88	0.79	0.74		0.85	0.83	0.89	0.88	0.81	0.85	0.81	0.86	0.81	0.85	0.87	0.89	0.83	0.81	0.9
StR8	0.88	0.86	0.81	0.83	0.83	0.75	0.85		0.82	0.89	0.83	0.82	0.84	0.85	0.87	0.8	0.87	0.86	0.78	0.77	0.78	0.83
StR9	0.88	0.82	0.8	0.79	0.86	0.81	0.83	0.82		0.85	0.8	0.79	0.79	0.76	0.82	0.74	0.8	0.82	0.79	0.85	0.8	0.81
StR10	0.89	0.88	0.86	0.89	0.82	0.76	0.89	0.89	0.85		0.89	0.83	0.88	0.83	0.9	0.8	0.9	0.89	0.77	0.87	0.84	0.86
StR11	0.85	0.84	0.79	0.85	0.76	0.71	0.88	0.83	0.8	0.89		0.76	0.87	0.8	0.84	0.76	0.86	0.85	0.79	0.81	0.77	0.86
Con1	0.88	0.84	0.84	0.8	0.85	0.78	0.81	0.82	0.79	0.83	0.76		0.83	0.73	0.76	0.81	0.87	0.77	0.83	0.7	0.74	0.67
Con2	0.9	0.87	0.86	0.84	0.82	0.69	0.85	0.84	0.79	0.88	0.87	0.83		0.77	0.77	0.86	0.83	0.79	0.83	0.65	0.82	0.74
Con3	0.8	0.82	0.84	0.81	0.73	0.73	0.81	0.85	0.76	0.83	0.8	0.73	0.77		0.84	0.72	0.84	0.83	0.85	0.83	0.83	0.83
Con4	0.91	0.89	0.84	0.85	0.84	0.77	0.86	0.87	0.82	0.9	0.84	0.76	0.77	0.84		0.77	0.89	0.85	0.85	0.73	0.8	0.8
Con5	0.87	0.84	0.76	0.79	0.81	0.71	0.81	0.8	0.74	0.8	0.76	0.81	0.86	0.72	0.77		0.8	0.82	0.82	0.68	0.78	0.73
Con6	0.88	0.92	0.87	0.83	0.84	0.72	0.85	0.87	0.8	0.9	0.86	0.87	0.83	0.84	0.89	0.8		0.85	0.86	0.77	0.84	0.78
Con7	0.86	0.88	0.83	0.85	0.8	0.77	0.87	0.86	0.82	0.89	0.85	0.77	0.78	0.8	0.82	0.82	0.85		0.83	0.74	0.8	0.76
Con8	0.87	0.82	0.84	0.76	0.84	0.73	0.89	0.78	0.79	0.79	0.77	0.79	0.83	0.85	0.85	0.82	0.86	0.83		0.82	0.88	0.83
Con9	0.86	0.86	0.87	0.82	0.82	0.81	0.83	0.77	0.85	0.87	0.81	0.7	0.65	0.83	0.73	0.68	0.77	0.74	0.82		0.79	0.82
Con10	0.89	0.84	0.82	0.77	0.84	0.82	0.81	0.78	0.8	0.84	0.77	0.74	0.82	0.83	0.8	0.78	0.84	0.8	0.88	0.79		0.8
Con11	0.87	0.87	0.84	0.88	0.79	0.73	0.9	0.83	0.81	0.86	0.86	0.67	0.74	0.83	0.8	0.73	0.78	0.76	0.83	0.82	0.8	

Appendix 13: Linear weighted kappa trainees and specialists Phase 2 cropped images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Cons1	Cons2	Cons3	Cons4	Cons5	Cons6	Cons7	Cons8	Cons9	Cons10	Cons11
StR 1		0.75	0.66	0.73	0.68	0.55	0.65	0.68	0.69	0.72	0.68	0.66	0.61	0.72	0.71	0.56	0.71	0.74	0.64	0.6	0.6	0.47
StR2	0.75		0.71	0.76	0.65	0.57	0.7	0.68	0.67	0.75	0.71	0.68	0.59	0.79	0.75	0.6	0.67	0.73	0.7	0.66	0.63	0.5
StR3	0.66	0.71		0.76	0.74	0.62	0.71	0.7	0.74	0.72	0.76	0.69	0.72	0.68	0.69	0.55	0.62	0.66	0.77	0.76	0.76	0.57
StR4	0.73	0.76	0.76		0.7	0.59	0.75	0.7	0.77	0.81	0.78	0.75	0.66	0.78	0.65	0.52	0.61	0.58	0.72	0.74	0.64	0.61
StR5	0.68	0.65	0.74	0.7		0.55	0.68	0.72	0.69	0.74	0.74	0.68	0.66	0.63	0.53	0.48	0.66	0.55	0.72	0.71	0.69	0.41
StR6	0.55	0.57	0.62	0.59	0.55		0.54	0.55	0.65	0.6	0.6	0.55	0.51	0.6	0.51	0.42	0.45	0.53	0.63	0.69	0.54	0.46
StR7	0.65	0.7	0.71	0.75	0.68	0.54		0.69	0.71	0.76	0.77	0.72	0.69	0.69	0.67	0.48	0.66	0.57	0.74	0.66	0.71	0.52
StR8	0.68	0.68	0.7	0.7	0.72	0.55	0.69		0.68	0.77	0.7	0.72	0.68	0.67	0.65	0.58	0.67	0.59	0.71	0.66	0.66	0.47
StR9	0.69	0.67	0.74	0.77	0.69	0.65	0.71	0.68		0.72	0.72	0.71	0.67	0.72	0.59	0.47	0.58	0.56	0.75	0.75	0.65	0.53
StR10	0.72	0.75	0.72	0.81	0.74	0.6	0.76	0.77	0.72		0.81	0.72	0.65	0.76	0.66	0.53	0.65	0.61	0.6	0.7	0.67	0.47
StR11	0.68	0.71	0.76	0.78	0.74	0.6	0.77	0.7	0.72	0.81		0.7	0.68	0.76	0.65	0.51	0.61	0.58	0.72	0.69	0.71	0.57
Cons 1	0.66	0.68	0.69	0.75	0.68	0.55	0.72	0.72	0.71	0.72	0.7		0.65	0.67	0.65	0.55	0.64	0.64	0.69	0.64	0.62	0.49
Cons2	0.61	0.59	0.72	0.66	0.66	0.51	0.69	0.68	0.67	0.65	0.68	0.65		0.61	0.7	0.57	0.66	0.59	0.75	0.65	0.68	0.5
Cons3	0.72	0.79	0.68	0.78	0.63	0.6	0.69	0.67	0.72	0.76	0.76	0.67	0.61		0.68	0.68	0.57	0.67	0.67	0.69	0.62	0.52
Cons4	0.71	0.75	0.69	0.65	0.53	0.51	0.67	0.65	0.59	0.66	0.65	0.65	0.7	0.68		0.69	0.72	0.75	0.7	0.57	0.62	0.46
Cons5	0.56	0.6	0.55	0.52	0.48	0.42	0.48	0.58	0.47	0.53	0.51	0.55	0.57	0.68	0.69		0.62	0.67	0.58	0.47	0.53	0.4
Cons6	0.71	0.67	0.62	0.61	0.66	0.45	0.66	0.67	0.58	0.65	0.61	0.64	0.66	0.57	0.72	0.62		0.69	0.64	0.52	0.6	0.44
Cons7	0.74	0.73	0.66	0.58	0.55	0.53	0.57	0.59	0.56	0.61	0.58	0.64	0.59	0.67	0.75	0.67	0.69		0.66	0.53	0.55	0.45
Cons8	0.64	0.7	0.77	0.72	0.72	0.63	0.74	0.71	0.75	0.6	0.72	0.69	0.75	0.67	0.7	0.58	0.64	0.66		0.74	0.64	0.5
Cons9	0.6	0.66	0.76	0.74	0.71	0.69	0.66	0.66	0.75	0.7	0.69	0.64	0.65	0.69	0.57	0.47	0.52	0.53	0.74		0.67	0.52
Cons10	0.6	0.63	0.76	0.64	0.69	0.54	0.71	0.66	0.65	0.67	0.71	0.62	0.68	0.62	0.62	0.53	0.6	0.55	0.64	0.67		0.5
Cons11	0.47	0.5	0.57	0.61	0.41	0.46	0.52	0.47	0.53	0.47	0.57	0.49	0.5	0.52	0.46	0.4	0.44	0.45	0.5	0.52	0.5	

Appendix 14: Linear weighted kappa trainees and specialists Phase 2 full images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Con11
StR 1		0.77	0.65	0.7	0.75	0.58	0.68	0.73	0.58	0.7	0.51	0.62	0.68	0.53	0.77	0.74	0.71	0.68	0.71	0.64	0.63	0.68
StR2	0.77		0.57	0.6	0.76	0.59	0.66	0.73	0.64	0.68	0.57	0.67	0.55	0.57	0.7	0.72	0.7	0.72	0.66	0.6	0.62	0.72
StR3	0.65	0.57		0.57	0.66	0.58	0.72	0.64	0.55	0.59	0.54	0.6	0.68	0.57	0.61	0.65	0.64	0.57	0.79	0.58	0.55	0.71
StR4	0.7	0.6	0.57		0.69	0.63	0.7	0.69	0.59	0.67	0.59	0.6	0.62	0.6	0.58	0.66	0.65	0.57	0.65	0.56	0.64	0.67
StR5	0.75	0.76	0.66	0.69		0.66	0.74	0.75	0.72	0.69	0.69	0.62	0.6	0.65	0.63	0.75	0.72	0.61	0.71	0.65	0.62	0.74
StR6	0.58	0.59	0.58	0.63	0.66		0.65	0.62	0.68	0.57	0.67	0.56	0.61	0.51	0.52	0.57	0.65	0.5	0.66	0.63	0.57	0.64
StR7	0.68	0.66	0.72	0.7	0.74	0.65		0.75	0.61	0.79	0.68	0.69	0.66	0.61	0.65	0.64	0.68	0.57	0.66	0.67	0.58	0.77
StR8	0.73	0.73	0.64	0.69	0.75	0.62	0.75		0.69	0.7	0.65	0.58	0.6	0.67	0.71	0.7	0.68	0.64	0.64	0.66	0.62	0.79
StR9	0.58	0.64	0.55	0.69	0.72	0.68	0.61	0.69		0.6	0.64	0.62	0.5	0.54	0.54	0.59	0.64	0.49	0.63	0.58	0.59	0.7
StR10	0.7	0.68	0.59	0.67	0.69	0.57	0.79	0.7	0.6		0.63	0.64	0.62	0.55	0.68	0.63	0.7	0.59	0.59	0.67	0.6	0.69
StR11	0.51	0.57	0.54	0.59	0.69	0.67	0.68	0.65	0.64	0.63		0.54	0.53	0.56	0.49	0.6	0.6	0.47	0.62	0.53	0.63	0.64
Cons 1	0.62	0.67	0.6	0.6	0.62	0.56	0.69	0.58	0.62	0.64	0.54		0.6	0.5	0.63	0.6	0.7	0.55	0.67	0.55	0.63	0.71
Cons2	0.68	0.55	0.68	0.62	0.6	0.61	0.66	0.6	0.5	0.62	0.53	0.6		0.51	0.62	0.61	0.61	0.51	0.68	0.59	0.57	0.67
Con3	0.53	0.57	0.57	0.6	0.65	0.51	0.61	0.67	0.54	0.55	0.56	0.5	0.51		0.5	0.57	0.62	0.54	0.51	0.55	0.5	0.58
Con4	0.77	0.7	0.61	0.58	0.63	0.52	0.65	0.71	0.54	0.68	0.49	0.63	0.62	0.5		0.72	0.62	0.65	0.63	0.57	0.61	0.66
Con5	0.74	0.72	0.65	0.66	0.75	0.57	0.64	0.7	0.59	0.63	0.6	0.6	0.61	0.57	0.72		0.69	0.72	0.68	0.55	0.6	0.73
Con6	0.71	0.7	0.64	0.65	0.72	0.65	0.68	0.68	0.64	0.7	0.6	0.7	0.61	0.62	0.62	0.69		0.62	0.61	0.62	0.61	0.68
Con7	0.68	0.72	0.57	0.57	0.61	0.5	0.57	0.64	0.49	0.59	0.47	0.55	0.51	0.54	0.65	0.72	0.62		0.55	0.54	0.51	0.58
Con8	0.71	0.66	0.79	0.65	0.71	0.66	0.66	0.64	0.63	0.59	0.62	0.67	0.68	0.51	0.63	0.68	0.61	0.55		0.61	0.63	0.67
Con9	0.64	0.6	0.58	0.56	0.65	0.63	0.67	0.66	0.58	0.67	0.53	0.55	0.59	0.55	0.57	0.55	0.62	0.54	0.61		0.52	0.64
Con10	0.63	0.62	0.55	0.64	0.62	0.57	0.58	0.62	0.59	0.6	0.63	0.63	0.57	0.5	0.61	0.6	0.61	0.51	0.63	0.52		0.63
Con11	0.68	0.72	0.71	0.67	0.74	0.64	0.77	0.79	0.7	0.69	0.64	0.71	0.67	0.58	0.66	0.73	0.68	0.58	0.67	0.64	0.63	

Appendix 15: Quadratic weight kappa trainees and specialists Phase 2 cropped images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Con11
StR 1		0.88	0.83	0.85	0.83	0.67	0.8	0.81	0.84	0.87	0.81	0.84	0.79	0.83	0.82	0.71	0.85	0.84	0.79	0.77	0.77	0.58
StR2	0.88		0.85	0.85	0.76	0.7	0.81	0.78	0.82	0.85	0.8	0.82	0.76	0.87	0.89	0.79	0.8	0.88	0.86	0.83	0.77	0.59
StR3	0.83	0.85		0.87	0.86	0.76	0.83	0.83	0.87	0.87	0.89	0.85	0.85	0.85	0.83	0.7	0.77	0.8	0.89	0.89	0.88	0.7
StR4	0.85	0.85	0.87		0.82	0.73	0.86	0.85	0.89	0.92	0.89	0.89	0.84	0.9	0.8	0.71	0.77	0.75	0.83	0.86	0.8	0.7
StR5	0.83	0.76	0.86	0.82		0.65	0.79	0.84	0.83	0.84	0.84	0.82	0.85	0.81	0.71	0.65	0.81	0.72	0.83	0.83	0.84	0.58
StR6	0.67	0.7	0.76	0.73	0.65		0.68	0.68	0.74	0.76	0.75	0.7	0.65	0.73	0.69	0.56	0.6	0.65	0.76	0.84	0.69	0.61
StR7	0.8	0.81	0.83	0.86	0.79	0.68		0.78	0.87	0.88	0.88	0.82	0.83	0.85	0.78	0.67	0.77	0.71	0.81	0.81	0.83	0.68
StR8	0.81	0.78	0.83	0.85	0.84	0.68	0.78		0.83	0.87	0.83	0.84	0.82	0.84	0.76	0.72	0.78	0.72	0.8	0.82	0.83	0.56
StR9	0.84	0.82	0.87	0.89	0.83	0.74	0.87	0.83		0.88	0.87	0.86	0.84	0.86	0.76	0.68	0.78	0.72	0.85	0.88	0.83	0.67
StR10	0.87	0.85	0.87	0.92	0.84	0.76	0.88	0.87	0.88		0.92	0.87	0.83	0.91	0.81	0.69	0.8	0.77	0.84	0.85	0.84	0.62
StR11	0.81	0.8	0.89	0.89	0.84	0.75	0.88	0.83	0.87	0.92		0.83	0.84	0.9	0.78	0.66	0.78	0.72	0.83	0.85	0.87	0.72
Cons 1	0.84	0.82	0.85	0.89	0.82	0.7	0.82	0.84	0.86	0.87	0.83		0.84	0.84	0.79	0.72	0.79	0.77	0.82	0.82	0.78	0.63
Cons2	0.79	0.76	0.85	0.84	0.85	0.65	0.83	0.82	0.84	0.83	0.84	0.84		0.81	0.81	0.72	0.82	0.74	0.87	0.82	0.82	0.64
Con3	0.83	0.87	0.85	0.9	0.81	0.73	0.85	0.84	0.86	0.91	0.9	0.84	0.81		0.83	0.83	0.77	0.79	0.82	0.85	0.81	0.65
Con4	0.82	0.89	0.83	0.8	0.71	0.69	0.78	0.76	0.76	0.81	0.78	0.79	0.81	0.83		0.81	0.83	0.87	0.86	0.78	0.73	0.58
Con5	0.71	0.79	0.7	0.71	0.65	0.56	0.67	0.72	0.68	0.69	0.66	0.72	0.72	0.83	0.81		0.78	0.8	0.75	0.69	0.68	0.54
Con6	0.85	0.8	0.77	0.77	0.81	0.6	0.77	0.78	0.78	0.8	0.78	0.79	0.82	0.77	0.83	0.78		0.79	0.77	0.73	0.77	0.51
Con7	0.84	0.88	0.8	0.75	0.72	0.65	0.71	0.72	0.72	0.77	0.72	0.77	0.74	0.79	0.87	0.8	0.79		0.81	0.74	0.68	0.56
Con8	0.79	0.86	0.89	0.83	0.83	0.76	0.81	0.8	0.85	0.84	0.83	0.82	0.87	0.82	0.86	0.75	0.77	0.81		0.88	0.8	0.63
Con9	0.77	0.83	0.89	0.86	0.83	0.84	0.81	0.82	0.88	0.85	0.85	0.82	0.82	0.85	0.78	0.69	0.73	0.74	0.88		0.84	0.68
Con10	0.77	0.77	0.88	0.8	0.84	0.69	0.83	0.83	0.83	0.84	0.87	0.78	0.82	0.81	0.73	0.68	0.77	0.68	0.8	0.84		0.67
Con11	0.58	0.59	0.7	0.7	0.58	0.61	0.68	0.56	0.67	0.62	0.72	0.63	0.64	0.65	0.58	0.54	0.51	0.56	0.63	0.68	0.67	

Appendix 16: Quadratic weight kappa trainees and specialists Phase 2 full images

	StR 1	StR2	StR3	StR4	StR5	StR6	StR7	StR8	StR9	StR10	StR11	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Con11
StR 1		0.86	0.82	0.82	0.87	0.72	0.85	0.88	0.77	0.83	0.7	0.79	0.76	0.75	0.87	0.86	0.84	0.82	0.8	0.82	0.8	0.83
StR2	0.86		0.72	0.78	0.86	0.77	0.79	0.83	0.82	0.8	0.72	0.81	0.68	0.7	0.83	0.84	0.82	0.87	0.82	0.75	0.82	0.8
StR3	0.82	0.72		0.77	0.81	0.76	0.88	0.8	0.74	0.8	0.71	0.77	0.83	0.76	0.79	0.79	0.8	0.71	0.74	0.77	0.71	0.82
StR4	0.82	0.78	0.77		0.86	0.8	0.86	0.83	0.81	0.82	0.78	0.78	0.77	0.76	0.73	0.83	0.82	0.72	0.8	0.73	0.82	0.84
StR5	0.87	0.86	0.81	0.86		0.83	0.86	0.87	0.86	0.81	0.82	0.76	0.79	0.78	0.77	0.89	0.85	0.77	0.86	0.83	0.81	0.87
StR6	0.72	0.77	0.76	0.8	0.83		0.82	0.76	0.86	0.77	0.83	0.74	0.75	0.7	0.69	0.74	0.8	0.66	0.81	0.76	0.75	0.79
StR7	0.85	0.79	0.88	0.86	0.86	0.82		0.88	0.82	0.9	0.81	0.85	0.82	0.78	0.84	0.82	0.84	0.75	0.8	0.82	0.77	0.89
StR8	0.88	0.83	0.8	0.83	0.87	0.76	0.88		0.84	0.7	0.77	0.75	0.8	0.82	0.82	0.86	0.81	0.78	0.78	0.81	0.77	0.89
StR9	0.77	0.82	0.74	0.81	0.86	0.86	0.82	0.84		0.78	0.8	0.78	0.72	0.75	0.75	0.8	0.8	0.7	0.8	0.76	0.78	0.84
StR10	0.83	0.8	0.8	0.82	0.81	0.77	0.9	0.7	0.6		0.81	0.8	0.74	0.71	0.82	0.79	0.86	0.77	0.71	0.82	0.79	0.82
StR11	0.7	0.72	0.71	0.78	0.82	0.83	0.81	0.77	0.8	0.81		0.67	0.73	0.75	0.66	0.79	0.79	0.66	0.75	0.71	0.78	0.78
Con1	0.79	0.81	0.77	0.78	0.76	0.74	0.85	0.75	0.78	0.8	0.67		0.69	0.65	0.81	0.76	0.82	0.75	0.8	0.72	0.78	0.82
Con2	0.76	0.68	0.83	0.77	0.79	0.75	0.82	0.8	0.72	0.74	0.73	0.69		0.74	0.72	0.78	0.72	0.65	0.77	0.69	0.72	0.8
Con3	0.75	0.7	0.76	0.76	0.78	0.7	0.78	0.82	0.75	0.71	0.75	0.65	0.74		0.69	0.77	0.74	0.67	0.7	0.72	0.71	0.77
Con4	0.87	0.83	0.79	0.73	0.77	0.69	0.84	0.82	0.75	0.82	0.66	0.81	0.72	0.69		0.83	0.8	0.77	0.75	0.76	0.77	0.82
Con5	0.86	0.84	0.79	0.83	0.89	0.74	0.82	0.86	0.8	0.79	0.79	0.76	0.78	0.77	0.83		0.85	0.83	0.84	0.77	0.78	0.87
Con6	0.84	0.82	0.8	0.82	0.85	0.8	0.84	0.81	0.8	0.86	0.79	0.82	0.72	0.74	0.8	0.85		0.79	0.77	0.79	0.79	0.83
Con7	0.82	0.87	0.71	0.72	0.77	0.66	0.75	0.78	0.7	0.77	0.66	0.75	0.65	0.67	0.77	0.83	0.79		0.72	0.74	0.71	0.71
Con8	0.8	0.82	0.74	0.8	0.86	0.81	0.8	0.78	0.8	0.71	0.75	0.8	0.77	0.7	0.75	0.84	0.77	0.72		0.75	0.8	0.81
Con9	0.82	0.75	0.77	0.73	0.83	0.76	0.82	0.81	0.76	0.82	0.71	0.72	0.69	0.72	0.76	0.77	0.79	0.74	0.75		0.7	0.8
Con10	0.8	0.82	0.71	0.82	0.81	0.75	0.77	0.77	0.78	0.79	0.78	0.78	0.72	0.71	0.77	0.78	0.79	0.71	0.8	0.7		0.79
Con11	0.83	0.8	0.82	0.84	0.87	0.79	0.89	0.89	0.84	0.82	0.78	0.82	0.8	0.77	0.82	0.87	0.83	0.71	0.81	0.8	0.79	