Questioning Assessment:

Policy and Classroom Practices in Bahrain Primary Schools.

By

Latifa Ali Al-Mannai

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy.

May 1996

Table of Content

List of Tables

Acknowledgements

I am greatly indebted to my supervisor David Hamilton who gave me valuable feedback on the work of this thesis. During three years of work, David was so helpful, patient and understanding. I am grateful for his extremely careful readings of the several drafts of the chapters. Without his constructive suggestions and prudent comments, this work could not be possible. I am also indebted for the valuable help he gave me in improving my writing in English.

I owe a special debt to my boss Nakhleh Wehbeh who has worked effectively for my grant from the Ministry of Education. He was also of great help to me in developing the original plan of this thesis. Also his comments on successive drafts of the chapters have been of great value although I am still responsible for the views represented in this thesis.

I should particularly like to thank the teachers of the two primary schools who allowed me into their classrooms for a full term. I am also grateful to the two headteachers who facilitated access to their schools and provided me with all documents that I requested.

Finally, I must thank my friends Aneesa Kamdan and Jamilla Mjairan for their consistent help in looking for, and also sending, all the data I needed from Bahrain. Their help was of great benefit for my work.

Abstract

Title: Questioning assessment: policy and classroom practices in Bahrain primary schools.

(Ph.D thesis submitted to University of Liverpool, 1996)

By
Latifa Ali Al-Mannai

This thesis is about the assessment practices in Bahrain primary schools. It is built upon a small scale observation study of six primary teachers. Teachers' practices are viewed within the theory of assessment, within the context of classrooms and within national assessment policy.

The first chapter sets out the general grounds of the research. It explores the research approach used in this thesis; it gives the background to the choice of the topic; and it reports the strategy and methods adopted in the research.

Chapters two and three review assessment in theory. Chapter two places assessment within historical perspectives. It shows how assessment has been developed, as a concept and in practice, over this century. Assessment within the context of classrooms is reviewed in chapter three. It explores classroom assessment as interpersonal judgement, as an interaction process, as an information gathering process and as an intentional activity.

Chapter 4 traces the development of Bahrain primary schools since their establishment in the early years of this century. Four phases are discussed, the initiative period (1919 - 1939), the configuration period (1940 - 1970), the transitional period (1971 - 1981) and the progressive period (1982 - 1996).

Chapter five is concerned with classroom life. It reviews the professional status of primary school teachers, their background, their qualifications and their classroom context. The observed four classes are then described, together with further information about their teachers and classroom relationships.

Chapters six, seven and eight consider the context of classroom assessment. Chapter six traces the ideas of teachers' judgement and expectation in the context of four primary classrooms. Chapter seven is concerned with teacher-pupil interaction. Chapter 8 is devoted to pencil and paper testing since it still dominates teachers' assessment practices in Bahrain classrooms.

Finally, chapter 9 is concerned with broader implications of the research reported in the previous chapters. The findings are related to official policy and the future of assessment in Bahrain primary schools.

1

Research Background

Until 1988, no officials held responsibility for research and evaluation in the Ministry of Education in Bahrain. Research studies were mainly carried out by concerned members of the Ministry, as they planned new programmes. Post-graduate students also played a part in educational research. Most research has been in the areas of curriculum and teaching methods (Al-Saleh, 1992). The concern of these studies, however, remained within the interest of the researchers themselves. Using experimental design and field trials, such research was preoccupied with the measurement of inputs and outputs. It focused, for instance, on topics such as: 'The effect of mastery learning strategy in mathematics on the achievement of third grade intermediate students', 'Interest in science of sixth grade children in Bahrain and some related factors' and 'Evaluation of skills in silent reading of sixth grade students in public Bahrain schools' (AL-Saleh, 1992).

Wehbeh *et al* (1991), scrutinized 18 research plans submitted for Masters' degrees in education by postgraduate students in the College of Education. Twelve used experimental methods while the remainder used analytical and descriptive approaches. Wehbeh and co-workers criticized the methodology used in educational research at the College of Education and by postgraduate students of the college who expected to work in the Ministry of Education. They cast doubt, for instance, on the usefulness of such research in decision making.

> These research plans concentrated on trial and error methods, and have not yet proceeded towards action research. Every researcher wanted to trial his/her solution, while the real problem, its context, development and consequences remained unexamined (p. 25).

Alternative models in educational research, however, began to appear in Bahrain after the formation of the Educational Research and Development Centre in 1983 which become fully operational in 1988. Researchers advocating alternative methods faced resistance from a number of traditional researchers in Bahrain. Such research was however, respected for its interpretation of the realities of schooling. It integrated researchers into schooling and education. These researches were able to penetrate into the real practice and thus, narrow the gap between policy and practice. Such research, therefore, has received increased funding from the Ministry; for example, in 1991 and 1992, the funds devoted to research were increased from £ 45,000 to £ 115,500.

New thinking in educational research:

The move toward an alternative model of research in education echoed the

widespread claims for the usefulness of such research methodology. The call to move from earlier research methods, which emphasized measurement and objectivity of data, to a new model of research methods is widely heard in the area of social and educational research. Earlier research methods had grown out of a scientific search for cause and effect relationships. They had been influenced by psychological investigations of the human intellect. Such research, as Hamilton (1980) indicated was intensified in the work of Francis Galton on individual differences and Ronald Fisher on agricultural field testing. The ideal strategy for such research was experimental control of the situation to eliminate non-relevant variables. In turn, relationships of cause and effect could be deduced or inferred. And, in their turn, such findings could be generalised to new contexts.

At the present, this type of research is less influential in education. It is now widely accepted that educational research is a social activity. Thus, all interactions involved in social research cannot be treated as independent from the data collected and the methodology used. Thus, standardization employed in such research does not guarantee valid and free of bias data. This type of research, as Nisbet (1980) suggested:

> starts from *our* assumptions, *our* framework of thought, and it imposes that framework on what we innocently call the 'subjects' in our experiments. Not surprisingly, the framework usually does not fit; and so we adjust our control mechanisms until we have a situation where we can use our preconceived models - and, not surprisingly, this kind of research produces results of limited value and limited application (p. 6).

The most obvious shortcomings of such research were summarized by Parlett and

Hamilton (1977):

(a) The research method entails a strict control of the situation under concern. By such procedures, the real situation is distorted. Thus, the controlled situation renders results irrelevant to the actual situation, which in itself divorces the study from the real world.

> Whichever approach is used, there is a tendency for the investigator to think in terms of 'parameters' and 'factors' rather than 'individuals' and 'institutions'. Again, this divorces the study from the real world (p. 8).

(b) Researchers are constrained by the research designs, which assume that the programme should not be changed under the changing circumstances. They, therefore, can 'rarely serve an effective or "formative" cybernetic function' (p. 9).

(c) In seeking quantitative information by objective means, traditional evaluation research disregards important data. These data are neglected because they are regarded as 'subjective', 'anecdotal' or 'impressionistic', although they may be salient to the innovation.

> [T]he evaluator is likely to be forced to utilise information of this sort if he is satisfactorily to explain his findings, weight their importance and place them in context (p. 9).

(d) In searching for generalization, details of particular importance to the innovation, individuals and institutions concerned are discounted. Thus, 'they are ironed out and lost to discussion' (p. 9).

(e) The concern over objectivity induces results equally relevant to all parties. These studies, therefore, 'rarely acknowledge the diversity of questions posed by different interest-groups' (p. 9)..

An alternative model was suggested by Parlett and Hamilton in which description and interpretation rather than measurement and prediction is emphasized. It was described as a general research strategy that:

> aims to be both adaptable and eclectic. The choice of research tactics follows not from research doctrine, but from decisions in each case as to the best available techniques: the problem defines the methods used, not *vice versa*. Equally, no method (with its own built-in limitations) is used exclusively or in isolation; different techniques are combined to throw light on a common problem (p. 13).

Researchers following this approach work as social anthropologists or as natural historians. The researcher studies the situation as it is without any attempt to control or alter any variables. Concepts in this model emerge from the data rather then being imposed on the basis of *a priori* theory. The study may begin with general issues or questions but as the research goes on, the researcher begins to focus her/his attention on other issues that emerge during the study. Thus, research questions may become different from the original ones.

Parlett and Hamilton's concern for particularities of research settings prompted, among other things, a new interest in case study research - Towards a science of the singular (Simons, 1980). A basis for case study research has been recently re-stated by Robert Stake in The Art of Case Study Research (1995). Stake suggests that case study research has three distinct features.

His first claim is that case study research relies on a distinction between explanation and understanding: earlier methods of inquiry emphasized the

explanation and control of the situation while the alternative presses for understanding the situation as it happens. Explanation in earlier methods searches for cause and effect relationships, which define the most general variables and at the same time eliminates the particularity of the case in seeking for generalization of results. Understanding in qualitative inquiry searches for descriptions of situations without necessarily explaining why things happen as they do. In this sense, the context of the situation itself is of particular importance to the researchers on their way to understand the case under review.

> researchers treat the uniqueness of individual cases and contexts as important to understanding. Particularization is an important aim, coming to know the particularity of the case (p. 39).

Stakes's second claim is that earlier methods come mainly from relying on interpretation of measurement and statistical data while the new model relies on description of events. As researchers in the earlier methods sought out cause and effect relationships between a small number of variables, their personal interpretation was intentionally minimized while other objective measures were emphasized. In the alternative model, the researcher's personal interpretation is important. Variables are not operationally defined or controlled. Rather, they may become apparent as unexpected events occur. Thus, the researcher's interpretation is vital to the direction and re-direction of the inquiry.

> It is essential to have the interpretive powers of the research team in immediate touch with developing events and ongoing revelations, partly to redirect observation and to peruse emerging issues (pp. 41, 42).

Finally, Stake claimed the researcher's pre-eminence as an interpreter. He builds this role upon the argument that humans construct reality from their own experience of the world. That is, reality as we perceive it is not absolute but bounded by our experiences. In this sense, 'researchers nourish the belief that knowledge is constructed rather than discovered' (p. 99). Researchers are, therefore, encouraged to deliver their own understanding which can be held also in common. Researchers, at the same time should provide the readers with full descriptions of the events in order to help them to interpret and establish their own generalization.

> The researcher struggles to liberate the reader from simplistic views and illusion. The researcher is the agent of new interpretation, new knowledge, but also new illusion. Sometimes, the researcher points to what to believe, sometimes facilitating reader understandings that exceed the comprehension of the researcher (p. 99).

This new alternative approach has recently been adopted for research sponsored by the Educational Research and Development Centre in Bahrain.

> The ERDC prefers a qualitative approach to basic research, new data, techniques and conceptions of validity (ridding the objectivity concept of its classical requirements- reliability and normativity) have been adopted in order to conduct research on authentic educational problems (Wehbeh, 1995, p. 49).

Several issues have been studied using this model in which its value was apparent for decision-making and illumination of the reality of teaching and learning. As I recently worked in some of this research, I realized the major differences between the traditional model previously used and the new model. It is the alternative

model that I adopt in dealing with the issues of this thesis.

Issues in question:

The research reported in this thesis has been concerned with assessment policy and classroom practices in Bahrain primary schools. The issue of assessment started to attract my attention when I worked in the Educational Research and Development Centre (in 1989). In research that I participated in, the quality of Bahrain schools was continuously questioned. Results of studies gave indications that students' achievement in public schools was lower than envisaged. These doubts affected my own thinking. Much educational investment had been made since 1970 to improve education and raise the quality of students' learning. In the 1990s, this investment should have been reflected in improved students' achievement. Classroom practices came to the forefront of my concern. It was an area that had not been investigated.

My professional interest in assessment was extended in 1992 when I participated in gathering information about schools' assessment practices. The information was needed for a 'comparative study of current theories and practices in assessing students' achievement at primary and secondary level' sponsored by the International Association for Educational Assessment (IAEA). I realized that a gap exists between schools and the thinking of departments in the Ministry of Education with regard to the policies and practices of assessment and testing. The department's view, first circulated in the Annual Conference of Education in 1989, was that schools should give attention to formative and diagnostic assessment. But the practice of schools seemed to bear little resemblance to the aspiration of the

Ministry. Each school I visited, at that time, had its own context of assessment although all seemed to agree in sending an official summative report at the end of term to the Ministry of Education. The evidence of these discrepancies suggested that this area was in need of investigation. It become the focus of my Ph D studies.

Much of literature which had tackled this issue also encouraged such concern. Assessment has become a well-established area of international research and practice. The importance of studies in assessment have been realized by many professionals and researchers. Assessment is considered an important aspect for understanding the process of teaching and learning in classrooms and uncovering the reality of an education system. As Rowntree (1987) pointed out:

> If we wish to discover the truth about an educational system, we must look into its assessment procedures. What student qualities and achievement are actively valued and rewarded by the system? How are its purposes and intentions realized? To what extent are the hopes and ideals, aims and objectives professed by the system ever truly perceived, valued and striven for those who make their way within it? (p. 1).

Broader issues have also been realized in studying assessment practices. That is, the relationship between assessment, schools and society as a whole. As Broadfoot (1979a) put it:

> Indeed it could be argued that assessment practices are one of the clearest indices of the relationship between school and society since they provide for communication between the two (p. 11).

At the level of national interest, awareness of the importance of assessment and the

need to change the current practices began to emerge in Bahrain in the late 1980s. The Minister's speech on the Conference for Testing in the Arab World in March 1987, included criticism of, and an invitation to change, the present system of assessment and testing. He stated:

> This applies more forcefully to evaluation and testing in the education realm. It has been a process of adaptation, imitation and distortion of foreign systems designed by aliens .... The same thing is happening now for evaluation and testing. We are still holding on to it at its worst, while its creators have moved on to a better system of evaluation and testing (Fakhro, 1987, pp. 6, 7).

In 1989, the Ministry of Education devoted the Fifth Educational Conference for introducing and discussing new approaches in education. The two main papers discussed were: "formative and comprehensive assessment" and "assessment for teaching". All primary teachers participated in this conference. In the same year, the Ministry of Education also approved an Educational plan for the period 1989/90 to 1994/95. The plan gives as much stress to new approaches of assessment as to teaching and curriculum. It emphasises the introduction of diagnostic and formative assessment in primary education (Directorate of Plans and Programming, 1990). For such consideration to take place, present practices in primary schools deserve attention. Research in this area has not been conducted. Little is known about the practices of assessment in primary schools and very little consideration had been given to their educational and social consequences.

My research, therefore, addresses this issue. It has been concerned with two related areas: the policy context of assessment and the practice context of assessment. In

turn, these areas of interest were translated into an overarching research interest:

> What are the present assessment practices at the policy and classroom level in the primary schools in Bahrain and what purposes do they serve?

From these concerns a further set of questions was generated.

1. What is the historical context that produce the present assessment policies? (chapter 4)
2. What are the assessment policies of education at the present? (chapter 4)
3. How do these policies relate to the national educational objectives? (chapter 4 and 9)
4. What are the main consequences of these policies for classroom practice? (chapter 6, 7 and 8)
5. What are the main assessment methods and procedures that teachers use for assessing students' learning? (chapter 7 and 8)
6. Do these methods correspond with the policy's expectation? (chapter 9)
7. What is the impact of classroom assessment on students' learning? (chapter 7 and 8)
8. What types of achievement are valued in classrooms? (chapter 6).
9. What are the teachers' expectation of students? (chapter 6, 7 and 8)

Strategy:

Classroom assessment, as has been defined in this research, is not just a one shot objective test. Rather, it is a judgemental process which takes place in the interactional situations between teacher and students and aims to improve students' learning. This comprehensive definition entails the need to look thoroughly into

classroom contexts. Decision to choose classroom as the main setting for conducting observation came after conducting an orientation phase.

The aim of the orientation phase was to identify the main assessment concepts and strategies used in schools and to specify the main area in which observation could take place. The work extended from 20th December 1993 to 25th January 1994. It consisted of several tasks: visiting seven primary schools, meeting with two headteachers, attending nine lessons, interviewing eight teachers and observing two testing periods.

The advantage of doing the orientation phase was two-fold. First: It illuminated several critical issues in assessment, such as the different practices between the teachers in the first stage and teachers in the second stage and the general strength of feeling among teachers about shortcomings in the national promotion policies. Secondly: the orientation phase facilitated the selection of the observation setting. My choice was the school as a whole. Observing different classrooms, talking with teachers and meeting with headteachers provided me with an overall idea of primary classrooms: the teaching styles, the characteristics of students, physical environments and different procedures for planning a syllabus, teaching and assessment. These data may be important for illuminating the administrative and organizational contexts that surround the assessment context rather than the context of assessment itself. I thought that choosing school as a whole would drive me away from the aim of the research itself. The aim of this research was rather to document the realities of classroom assessment as they happened and describe the

assessment culture prevailing in them. These could mainly be illuminated by a more focused classroom observation.

Observation as a main method of the research was adopted for two reasons. First, an extended classroom observation could, more than other methods, illuminate both the uniqueness of the context of classrooms themselves and the generalities that surround the implementation of the common curriculum, the new teaching methods and the policies of assessment. It could also illuminate how the relationships between students, teachers and the activities are developed through the whole term and how these relationships work to enhance or inhibit quality assessment. Secondly, the extended observation could reduce the effect of the presence of a researcher on the events. The teacher and the students in the early days of observation would be aware of the presence of a stranger in the class. This awareness could have an effect on the process of the session. The session could not be considered as a normal classroom situation. This is applied if the observation period is short. When the period of observation is continuous and extended, then the teacher and the students will get used to the presence of the researcher and the normality of the classroom could then be discerned.

I intended to focus my observation on a small number of teachers who teach year three and year six (the end of the first and the second stage respectively) for a full term. Thus, the practice of six teachers in two primary schools (girls and boys) became the focus of my observation. The distribution of six teachers comes as follow: two class-teachers from the first stage (one in a boys' school and one in a

girls' school) and four integrated-subjects teachers from the second stage of primary education (two from a boys' school and two from a girls' school). In the second stage, I wanted to observe Arabic and Mathematics sessions in each school for two reasons: first to see whether there are differences in assessment approaches between the two subjects and secondly to get some insights into whether the content of curriculum has an effect on teachers' approaches on assessment.

Since no *a priori* grounds existed for selecting particular classrooms, a series of case studies was devised. The aim was to capture the widest possible range of contexts, in order to generate evidence and foster insight into classroom assessment. My concern was that the two schools needed to embrace, as closely as possible, the variability of Bahraini primary schools. That is, schools where the majority of the students come from middle class families; most teachers in these schools are Bahraini and hold at least a higher degree qualification in teachers' training and, finally, the school has a good record of achievement (students promoted are more than 90% of the total population in each year in the last three years). There were many schools which fit these criteria. Originally, there is a total of 195 (120 girls and 75 boys) classes in year-three and 113 (65 girls and 48 boys) classes in year-six existed in 39 girls and 24 boys schools. But my first interest was to choose schools that implement both systems, the class-teacher in the first stage (year 1 - 3) and the integrated-subjects teacher in the second stage (year 4 - 6). These were accounted for by 24 girls' and 14 boys' schools. I excluded some schools from this list: three schools in the capital Manama where several nationalities of students exist; three village schools where the majority of students

come from the working class families; the two pioneered schools of the class-teacher system; Hafsa and Badar and finally two boys' schools with female staff. Finally, there were 20 girls' and only 8 boys' schools from which I could choose. I chose two primary schools which are situated in two towns, both are near my work and my home. In each of the two schools one year-three class and one year-six class were chosen to be observed. The two schools attracted students of different social and economic background but the majority of their students are from middle class families (e.g. civil service employees, engineers and business men).

In September 1994, I addressed letters to the two headteachers of the chosen primary schools asking for their permission to make observations in two classrooms. I gave them a general idea about what I was trying to do and also specified my requirements. Both headteachers generously accepted my request and recommended three teachers from whom I also sought agreement.

Before the school days started, I had a meeting with all the teachers. One year-six female teacher declined to take part because she was teaching the sixth year for the first time. She said that she was not yet familiar with the required syllabus. The headteacher, therefore, recommended another one who welcomed my presence in one of her classes. A general idea was given to teachers about the aim of my research. I also clarified that I would ask for documents and reports such as students' activities, test scripts, students' assessment results and students' final records. I was aware that classroom observation is not familiar in Bahrain, and this is probably the first time someone has ever asked to carry out intensive observation

in classrooms. I, therefore, clarified for teachers that I would use a note-book to record what was happening in the classroom, and that what I would record was only for my own research. I also offered that they could ask to read my note book, if they wished. During the observation period, none of the teachers has taken up my invitation to read the notebook.

My observation settings were as follows:

Mrs A: Class-teacher in year three (girls' school).

Mr B: Class-teacher in year three (boys' school).

Mrs C: Arabic, religion and social studies teacher in year six (girls' school).

Mrs D: Mathematics and sciences teacher in year six (girls' school).

Mr E: Arabic, religion and social studies teacher (boys' school).

Mr F: Mathematics and sciences teacher (boys' school).

The observation period extended for a full term. It started on 25th September 1994 (the first day of the first term in 1994/95) and ended on 25th January 1995 (the end of the first term). Each teacher was observed for two sessions per week (1 hour and fifty minutes). A total of 21 to 23 hours were observed for each teacher (between 23 to 26 sessions). I did not make any contact with teachers during the sessions, although in very few incidents some teachers initiated a conversation. These incidents, however, had lasted for a few minutes. I had an impression that at least four teachers thought that I was probably building upon their experience of inspection, mainly focusing my attention on the wrong things that happened in the class. In one occasion, for example, when an incident in the class gave an

indication of poor performance or bad behaviour, the teacher looked into my direction probably to check if I was recording at that time. The teacher pressed on explaining the incident, a reaction which I felt it would not be the normal teacher's response in the normal classroom's sessions. This was more obvious in the boys' schools where the performance of many students was poor. Outside the classroom, I tried to give teachers the impression that I understood their situation and I gave them a chance to explain the whole situation. This is actually the main reason for not conducting an interview with the two headteachers. I was worried that contact with headteachers might be mistakenly interpreted by teachers.

In three of the classrooms observed I sat in the back of the class, while in the fourth (year six boys' school) I sat in the front as the classroom itself was small and there was no other space. My presence was clarified to the students from the beginning of the term and, at my request, teachers had pointed out that my presence in the classroom has nothing to do with making judgement or official records about their progress or their behaviour. My relationships with students were very limited although in the girls' schools, students tried to make contact with me on several occasions. In these occasion I tried to be polite with the students, very briefly answered any question they asked but refused firmly to make longer conversation.

My observation technique was unstructured, that is, I chose to record the classroom events as they occurred. In the beginning, I concentrated on what the teacher did. At first, the note taking was intuitive. Teacher's talk and the overall context of the

lesson dominate the observation notes. Students' behaviour formed only a small part of my classroom fieldnotes. I gave my observation note to two of my colleagues in the Ministry and asked them to give me their suggestions for note taking improvement. The points mentioned above were the main shortcoming they gave in which I tried later to improve. Within the first four weeks, I altered my observation procedures. The focus of my observation became more intelligible than before. Although the focus was still the teachers' practice, the students' behaviour became an important aspect through which the teacher's practice could also be interpreted. Time became also another aspect of my concern. Thus, I started to observe and record students' actions and begun also to focus some attention on particular students.

In the second month, I asked teachers whether I could use a tape recorder in the class. All female teachers accepted while the male teachers refused. Tape recording helped me to focus my observation on other incidents which seemed important while not worrying about writing what teachers said. The transcripts of these recordings (nearly 300 A4 pages) complemented my observation notes. In the boys' schools I continued with observation and writing notes but tried to look more closely at what was happening instead of worrying about writing what the teachers had said. Reporting the exact words of teachers came to be less important than the meaning of what the teacher was saying. After I left the classroom at the end of the day, however, I read the observation notes and added what I might have missed.

From the first sight, it was obvious that the observation notes reveal more differences between teachers than similarities. It was important, therefore, to find a unifying framework within which the observation notes could be analyzed. Some unifying themes have been selected to structure the observation data. The themes were those which could illuminate general issues identified in the main research questions. These themes are: teachers' and students' background, teachers' general pattern of teaching (chapter 5), the nature of achievement emphasised in classrooms (chapter 6), teachers' informal assessment or, the information gathering process in the classroom (chapter 7)and finally teachers' formal assessment or, in other words, end of term tests (chapter 8).

The study of observational data is a combination of qualitative and quantitative analysis. That is, choosing single events which convey important meanings together with aggregating particular incidents for quantitative analysis (Stake, 1995). Some incidents, for example, appear to be of important meaning although they happened only once, such as the incident of Mr F with Mansoor (see chapter 6). Other interpretations of observation data came after an aggregation of several incidents which hold the same meaning even though the frequencies of these incidents may not be too many. This was the case, for example, with the assessment of reading in Arabic (see chapter 7). On the other hand, quantitative analysis was carried out for other conclusions such as looking for the pattern of teachers' interaction with high and low achieving students (see chapter 7). For this purpose, categories of interaction were determined, for instance, teachers' questions and teachers' responses followed by an aggregation of the incidents which correspond to the

meaning of categories.

Finally, a consolidation phase of gathering information was carried out. The aim was to clarify, assert, and complement the other data collected in the observation period. It was conducted regularly as the observation took place and after it was completed. It included interviews with the teachers observed, collecting students' activities and tests scripts, examining the Ministry's official documents and finally interviews with some parents. Interviews with teachers were conducted either during the normal classroom sessions or in the staff room. Questions were concentrated on what goes on in the classroom sessions and teachers' general constraints. Teachers also provided me regularly with students' script papers in formal tests, monthly and at the end of term test. I also collected all students' original scripts in the end of term test. These papers were the main source of information used in chapter 8. Meeting with parents was also another important source of information. It was conducted when the two schools arranged the 'open day'. The aim of this day was originally to inform parents of their children's results in the mid term test and also discuss any other problem concerning students' academic performance and behaviour. I conducted the meeting immediately after each mother or father met the teacher. This was helpful because I could pick up parents' reactions to teachers' comments and to the schools' assessment results. Although I didn't use much of this meeting information, it helped originally to understand the social context within which teachers perform their role. Other data were also used, particularly the Ministry of Education documents. These were mainly useful for understanding the official policy of

assessment in the central departments of education and differentiate this from the practice within these departments.

I am not claiming for a generalization of this small observation research to the whole of primary schools in Bahrain. I am aware that reality still consists of much richer and deeper information which might be different than the result of this study. My concern, however, is to provide an understanding of assessment practices approached in Bahrain schools. That is, to elicit some of the features of their assessment culture referred to by Nisbet (1993), when he noted that

> each country has its own "assessment culture", a body of practice established over time and linked with deeply held values, which (like other culture) provide a necessary continuity and stability but are strongly and often irrationally resistant to change (p. 26).

The research, therefore, tries to illuminate such culture which can be mainly seen in the day-to-day practice of teachers. Also, to interpret the relevance of this culture to assessment polices officially dictated to schools. I also hope that this research will generate more interest in classroom research in Bahrain schools in the future.

2

Historical Perspective

Assessment practices, as Broadfoot (1979a) suggested, like any other social phenomenon, need to be understood within their wider historical and international context. The aim of this chapter is to explore how assessment, as a concept and practice, has developed in the course of twentieth century.

Within the history of formal education, ideas about assessment have undergone substantial change. Assessment has progressed from testing for selection purposes to a more broader concept which involves multiple purposes. Behind this progress there are changes in perspective towards individuals' development and learning. These historical differences are most clearly indicated in the gradual differentiation in the role of assessment for education and for society. The differentiation of 'formative' from 'summative' assessment (Scriven, 1967) is a clear example of

changing views of assessment. Behind these two perspectives, formative and summative assessment, there is a history of conflicting and controversial theories and practices. Three phases can be identified. During the first half of the 20th century, assessment was an end-product measurement process. By the mid 20th century assessment had become an educational instrument for curriculum evaluation. In the late 20th century, assessment is viewed as part of the pedagogic process, focusing upon individuals as active social learners. These three phases will be discussed, in turn, in this chapter.

Although developments, identified in the following discussion, focus on assessment practices in Europe and U.S.A, it is by no mean concerned only with these countries. Assessment is a world-wide notion and many practices are similar among different countries. The focus on the U.S.A and Europe comes mainly for two reasons. First, the aim of the chapter is to explore the development of assessment, mainly developed in U.S.A and Europe, rather than to provide international comparison of the systems of assessment. Secondly, it is more likely that assessment practices in other countries, for example the third world, are similar to that in Western countries, since the majority of them have adopted versions of the Western educational system.

Assessment as a Measurement Process

The notion of "assessment" is only of recent use in education (Eisner, 1993). Previously, assessment was tackled in terms of examinations. Examinations are formal, bureaucratic measurement devices that serve educational purposes, mainly grouping, selection and certification. They are seen as reliable measures of students' performance and, thus, a fair way for selecting students for future careers. It is, therefore, useful to look at this term, 'examinations': where has it come from? why has it been adopted? and for what reasons has it dominated educational practices in all countries around the world?

Têng Ssu-Yü (1943) suggested that examinations can be traced back to 2000 years in ancient China. China, he asserted, was the first nation to use an examination system in schools and for civil services:

> the competitive examination system, oral and written, used both in schools and for civil service, was not invented in the ancient West, but in China, and presumably was of purely Chinese origin (p. 270).

As described by Têng, it was through the trading relationships between the West and the East in the 16th and 17th centuries that the Chinese examination system attracted Western attention. From the writing of the Western missionaries and travellers in the Orient and reflections of the outstanding thinkers in France and England, who praised the Chinese civilization, the West became interested in the Chinese examination system. It was seen that the stability and longevity of the Chinese government was due to the adaptation of an examination system. For

example, Têng quoted Thomas Tyler Meadows' 1847 recommendation that England should adopt civil service examinations like those that had existed for thousands of years in China. Meadows' view was that " the long duration of the Chinese empire is solely and altogether owing to the good government which consists in the advancement of men of talent and merit only" (p. 289). In the nineteenth century, England started to adopt civil services examinations similar to the Chinese competitive examinations. Têng concluded:

> In the light of all this contemporary evidence, there can remain no doubt that the Chinese system of examinations for government positions was responsible for the introduction of similar systems into western Europe (p. 305).

The appearance of school examinations in Europe and U.S.A was also contemporaneous with the movement towards mass education in the second half of the nineteenth century (Broadfoot, 1979a, p. 28). The industrial revolution at that time marked a substantial change in the relationship between the family and schooling. Schools become an institution which often took children from one social milieu (e.g. rural, agriculture) and prepared them for work in another setting (e.g. urban, industrial). Schools become important institutions of social redirection. They were expected to produce competent individuals for new forms of productive work as well as, in a minority of cases, for admission to higher studies. Instruction had to be reformed in order to cope with such changes.

The idea of defining a general standard of performance for an age group and grade was born by the turn of this century (Thorndike and Hagen, 1955, p. 2). Pupils

began to be instructed in groups and were systematically selected and promoted from grade to grade (Hamilton, 1989, chapter 6). Consequently, attention focused on finding a model for judging pupils' performance, promoting them and directing them towards a future career. A system of examinations could be used for these purposes. It had the advantage of simplicity. And it provided an effective way for sorting pupils.

Placing examinations within such a 'high stakes' context had inevitable consequences for such a model. Measures should be reliable and objective. Scientific conceptions, most respected at that time, of rationality, precision, quantification and objectivity were to be applied in the field of education. Mental measurement was no different from the measurement of other attributes (Hamilton, 1980). Providing objective measurement of individual differences became important.

Such an issue was also advanced by experimental psychology - or psycho-physics. As Thorndike and Hagen (1955) point out:

> From experimental psychology came a legacy of respect for careful experimental method and precision of technique, a number of techniques that could be carried over to more general psychological and educational measurement problems. (p. 4).

Interest in individual differences was extended by the invention of intelligence tests during the first two decades of the twentieth century. These were first developed to identify subnormal children in schools (Gipps and Murphy, 1994, p. 60); but they soon penetrated into the school examination system. The 'boom' period in test

development, as Thorndike and Hagen call it (p. 5), commenced in 1915. Test development at this time also fed a growing interest in competitive examinations and school selection.

Influenced by biological assumptions, test constructors went on to propose a distribution of performance which has been called the normal or bell-shape curve (Hamilton, 1980, p. 156). This assumption has remained influential for many years. A prominent example of this practice was the Detroit XYZ Plan developed in U.S.A by Dr. Charles Barry in 1919 (Morgenstern, 1966, p.11). As determined by intelligence and achievement tests, pupils were to be grouped into three categories:

> 20 percent of the pupils who received the highest scores on intelligence tests in each grade were called the X group; those who had scores in the middle 60 percent constituted the Y group, and those with scores in the lowest 20 percent formed the Z group (p. 11).

The wholesale exploitation of intelligence testing began during World War I. It was used by the U.S Army to allocate army positions for military recruits. In British schools intelligence tests began to be used for allocating pupils to different forms of educational provisions and streaming them within the schools (Gipps and Murphy, 1994, p. 75). School selection tests came to public attention with the introduction of secondary school scholarship examinations for pupils (Sharp, 1984, p. 36). A typical example of selective examination was the 11-plus in British education which, after World War Two, was used for allocating children to grammar and secondary modern schools.

For such selection and grouping purposes, educational measurement was dominated by technical concerns. The most searching question in this sense was whether examinations provide a reliable, valid and fair measure. Technical quality in test administration and construction became very important. The aim was to maximize discrimination between individuals and to ensure that examinees would have the same score if their work is marked by two or more examiners. Multiple-choice standardized achievement tests met these technical requirements, particularly the goal of inter-marker reliability.

The pursuit of validity proved more difficult. Not all qualities can be examined using objective tests. Creativity, imagination, co-operativeness and critical thinking, for example, require other measures which may be regarded as less objective and less reliable. The emphasis on reliability and the mass production of standardized achievement tests undermined the assessment of important qualities, and thus, undermined the overall validity of such testing.

The reconciliation of reliability and validity is important to the purposes of examinations. Using imperfect examinations for prediction and selection may have unintended social consequences for society and individuals. From this perspective, psychological measures used in selection and prediction began to be criticised. Criticisms were directed against the assumptions underlying the construction of tests. For example, what is intelligence? Is it a stable and inheritable characteristic which is responsible for determining children's scholastic aptitude? If so, examination results and decisions are a direct consequence of pupils' intellectual

abilities. But if such measurement is imperfect or invalid, intelligence testing also has undesirable social implication, as Broadfoot (1979a) had suggested, of serving as 'a mechanism of social control', that is 'unsurpassed in teaching the doomed majority that their own failure was the result of their inbuilt inadequacy' (p. 44).

The assumption that intelligence and abilities are innate also underestimates the influence of other factors such as the social and economical environment. More importantly, it also underestimates the potential of education as a form of intervention that complements genetic effects. As Ryan (1972) puts the counter position.

> A child's ability to profit from school education will depend not only on his cognitive abilities, but also on the kind of education offered, the teachers and their values and abilities, the child's home background, and many other primarily cultural factors (p. 41).

Examinations were also criticised as selective devices. It was argued that differentiation policies justified by objective measures reinforce socio-economic differences. It has been claimed, therefore, that assessment errors deny working-class children the opportunity to gain access to more prestigious education and jobs. Furthermore, the predictive validity of such procedures has also been questioned. Researchers have claimed, for instance, that only a low correlation exists between results of selective examination and the future academic success (Rowntree, 1987, p. 17). Examinations, the critics believed, work as mechanisms for perpetuating social stratification and preserving the position of elites. Broadfoot (1981) expressed this view when she noted:

> examinations, and indeed a whole range of other assessment procedures such as the intelligence tests first developed around the turn of this century, come to play a quite crucial role in the regulation of access to privilege, and more importantly, in legitimating the basis for selection (p. 199).

As a result, selection policies for young children in Western Europe have been largely abandoned or postponed (Broadfoot, 1979a). Selection, however, continued to be the most important role of assessment until the present time. This role of assessment is reinforced when there is pressure of scarce educational resources, high population growth, an increasing demand on education and inadequate school performance. As Heyneman and Ransom (1992) put it in regard to developing countries.

> Examinations are a uniform mechanism for identifying talent and measuring achievement. Especially in environments where education resources are limited or unequally distributed among schools, examinations can help to ensure that society is investing in those who will in turn make the most useful contributions to society (p. 108).

Such a view, however, should take into account the fairness and appropriateness of the test content and the test use to the different groups in society. That is, procedures need to be taken to ensure that examinations do not discriminate against certain ethnic group and advantage other groups. Selection may be justified in higher education. But when it is applied, in the form of streaming and grouping, in the lower grades of education it should be associated with educational purposes, providing the appropriate instruction for each individual rather than labelling children prematurely as bright or slow learners.

Assessment As Educational Measurement

New thinking about assessment took place in the context of the educational reform movement in the 1950s in the U.S.A. The recognition that traditional forms of assessment do not provide information of *what* has been achieved provided justification for the changes. Assessment should be more descriptive and, thus, geared to educational objectives. The wide circulation of Ralph Tyler's Basic Principles of Curriculum and Instruction (1949) fostered the recognition that assessment like instruction should be geared to school curricula. This direct association of assessment and instruction brought assessment into a new educational context.

The key change, reflected in Tyler's work, was that assessment should be used for evaluating performance rather than the measurement of innate abilities. Attention shifted to questions about the aim of education itself (Tyler, 1986, p. 12). Society had to seek ways to improve educational provision in order to help each student reach his/her potential.

Curriculum development, therefore, became a focus of attention. It was deemed to be a potential means of improving the quality of education. One of the influential figures within this movement was Ralph Tyler who, while professor at Chicago University, designed a new frame of reference for the conduct of the school curriculum, instruction and evaluation (1949). This was a rationale for "viewing,

analyzing and interpreting the curriculum and instructional program of an educational institution" (p. 1). Educational goals were to be couched in terms of curriculum objectives. And these curricular objectives were to be classified in behavioral terms. Long-term educational goals became short-term educational objectives. Moreover, these short-term objectives steer classroom life. They: 'become the criteria by which materials are selected, content is outlined, instructional procedures are developed and tests and examinations are prepared' (Tyler, 1949, p. 3).

In turn, assessment (or, in Tyler's term, 'evaluation') became an important element within curriculum and instruction. It was a process 'for finding out how far the learning experiences as developed and organized are actually producing the desired results" (p. 105). Tyler's contribution to assessment practice was the recognition that a programme is needed for evaluating what the students had learned and also the curriculum itself. These factors stimulated educators to look for new approaches to assessment. In Wood's (1987) term, they sought:

> an approach to measurement which would be reflective of and responsive to what is peculiar to education, in particular the cycle of planning, instruction, learning, achievement and measurement. (p. 235).

The fundamental change in assessment, highlighted by Tyler (1949), was that assessment is not just an end-product measurement. It involves continuous and frequent assessment,

> so that a continuous record of progress can be obtained and evidence accumulated to indicate whether desirable objectives are being realized and to

> indicate places where these changes are not actually taking place (p. 107).

Another novel feature of the Tyler's rationale is that tests are not the only acceptable form of assessment. Other qualities, assessed with other devices, are equally important. Assessment methods, in this sense, should be sensitive to the desired behaviours identified in the educational objectives. In other words, the objectives of education define the instrument of assessment and not vice versa. A further important change is the recognition that a single score or mark is not an adequate summary of what has been achieved. Tyler (1949) suggested that the appraised characteristics and units of measurement should indicate the strength and weaknesses of an individual's learning.

> This kind of analytic summary which indicates particular strengths and weaknesses is, of course, invaluable in using the results to improve the curriculum (p. 117).

Changes in assessment had undergone further development with the introduction of two taxonomies of educational objectives (cognitive and affective) developed by Bloom and his colleagues, co-workers of Tyler. Bloom's Taxonomy of Educational Objectives (1956) grouped the educational objectives into six major classes from the simplest to the complex: knowledge, comprehension, application, analysis, synthesis and evaluation (p. 22). These classes are arranged in a hierarchy, so that each level demands skills and abilities of the lower ones. Objectives within each classification were to be made explicit, observable and expressed in terms of action. Later Bloom and his colleagues (1981) stressed that, as they saw it, assessment was

> the systematic collection of evidence to determine whether in fact certain changes are taking place in the learners as well as to determine the amount or degree of change in individual students (p. 5)

Glaser (1963) made a further contribution to assessment when he called for criterion-referenced measures which depend upon an *absolute* standard of quality as opposed to the long-standing norm-referenced forms of measurement which depend upon a *relative* standard of quality (e.g., as defined by a bell-shaped curve). In this respect, Glaser (1977) later suggested an instructional model based on defining foreseeable outcomes in terms of certain measurable products of students' performance. He also emphasized the importance of initial and continuous assessment which is 'referenced to and evaluated in terms of competence criteria and the values to be optimised' (p. 40). Criterion-referenced testing, and performance-based curricula look at individual performance in relation to her/himself rather than in relation to others. The shift in this movement, as Carver (1974) pointed out, was from measuring between-individual differences (norm-referenced) to measuring within-individual differences (criterion-referenced). It, therefore, raised the issue of performance standards. Assessment, in this sense, had moved from mere testing and examinations, and a preoccupation with aptitude and selection, to an evaluation process concerned with achievement and improvement in the quality of education. As Eisner (1993) pointed out:

> Educational evaluation had a mission broader than testing. It was concerned not simply with the measurement of student achievement, but with the quality of curriculum content, with the character of the activities in which students were engaged, with the ease with which teachers could gain access to

> curriculum materials, with the attractiveness of the curriculum's format, and with multiple outcomes, not only single ones (p. 221).

This approach to system-wide curriculum development and evaluation, called by Carver (1974) 'edumetric' as opposed to 'psychometric', gained wide support and has been intensively used in the U.S.A and elsewhere. As Lewy and Bathory (1994) concluded:

> The Taxonomy has been translated into several European languages, and in most countries surveyed, concise versions of the Taxonomy have been included in teachers' guides, curriculum guides, and textbooks on didactics, measurement, and evaluation (p. 160).

The Tyler-Bloom-Glaser movement has had a lasting impact on assessment because it linked assessment to educational processes. It has also revised thinking about assessment as a technical process. It, for example, has provided a model for differentiating higher from lower level-cognitive behaviours (Airasian, 1994, p. 99). And it has drawn attention to measuring students' performance without reference to others' performance (Glaser, 1977, p.42).

Wood (1987) listed six main themes in which educational measurement is distinct from psychometrics. Educational measurement, he stated:

- deals with the individual's achievement relative to himself rather than to others;
- seeks to test for competence rather than for intelligence;
- takes place in relatively uncontrolled conditions and so does not produce 'well behaved' data;
- looks for 'best' rather than 'typical' performance;

- is most effective when rules and regulations characteristic of standardized testing are related;
- embodies a constructive outlook on assessment where the aim is to help rather than sentence the individual (p. 240).

Nevertheless, the Tyler-Bloom-Glaser assessment rationale has not been universally accepted. Criticism of this approach has focused on curriculum and instruction as well as assessment.

One important criticism of the Tyler-Bloom-Glaser rationale was that it retained too much from psychometrics. The first key aspect in the educational measurement movement had remained: that is, to provide a rationale for educators in relation to curriculum, instruction and assessment. It, thus, did not go beyond psychometric measurement. Wood (1987), for example, stated that criterion-referenced testing 'remains the embodiment of educational measurement; notions like mastery testing (although not mastery learning) and minimum competency testing are only developed versions of the original conception, given a particular twist' (p. 235). He pointed out that criterion-referenced testing was still used for selection, screening and monitoring purposes as had been the case for norm-referenced testing. Gipps (1994) also pointed to the overlap between norm-referenced and criterion-referenced testing and stated that ' a norm-referenced interpretation can be put on a criterion-referenced measure, for example, most children of 7 will be at Level Two' (p. 80).

A further criticism of Tyler-Bloom-Glaser movement was related to its rationale of

educational objectives. Eisner (1985) identified several limitations. His first objection was that planning objectives in advance may overlook the unexpected outcomes of instruction. He confirmed that 'the dynamic and complex process of instruction yields outcomes far too numerous to be specified in behavioral and content terms in advance' (p. 32). The second point that Eisner made was that in some subject areas, such as mathematics and language, it is possible to specify with great precision objectives in advance. But in other areas, such as art, where the desired objectives are rather more abstract, behaviour cannot be easily identified. In addition, Eisner argued that many outcomes of curriculum and instruction are not amenable to measurement. Specifying educational objectives assumes that achievement can be measured against a standard of performance. This assumption fails to distinguish adequately between the application of standards for achievement and the need to make qualitative judgements in some fields of activities such as valuing a poem, a novel or a play.

Some have argued that the objectives approach has also limited classroom instruction and teacher assessment (e.g. Anderson, 1994). Shepard (1991) highlighted the undesired outcomes of behavioural approaches. Criterion-referenced measurement, grounded on behavioral learning theory, she argued, is based on two principles. First, that complex qualities can be broken down into discrete skills, taught and tested in isolation.

> Learning is seen to be linear and sequential. Complex understandings can occur only by accretion of elemental, prerequisite leanings (p. 6).

Learning in this behaviourist theory becomes meaningless because knowledge is presented to students as fragmented bits and pieces of information. The context within which knowledge and skills are constructed is easily neglected. Children, in this sense, are seen as passive absorbers, waiting to be filled with fixed knowledge.

The second principle identified by Shepard is concerned with testing and instruction.

> To facilitate learning, assessment should be closely allied with instruction. Tests should exactly specify desired behavioral outcomes of instruction and should be used at each learning juncture; that is, one should "test-teach-test" (p. 7).

This principle as Shepard pointed out means that the test is simplified to measure a mastery of small learning steps without any inference to a broader set of test questions. It also supports a narrow conception of learning goals and clearly does not consider complex criteria of performance. Shephard also stated that the behaviourist perspective treated tests and learning objectives as equivalent and, therefore, that 'teaching to tested objectives is synonymous with good instruction' (p. 7).

In fact, it has also been argued that a taxonomy could have a restrictive effect on instruction and assessment if teachers depend heavily on it. As stated by Rowntree (1987):

> Hence, the danger of entering an assessment situation with too specific and exclusive a set of learning objectives or 'constructs' in mind (p. 175).

Rowntree believed that the taxonomy should, for classroom assessment, be regarded cautiously as 'suggestive, illuminative and stimulating rather than as comprehensive, prescriptive and indubitable' (p. 105).

Another major criticism of educational measurement and its emphasis on testing as a means for selection, has been that it shapes classroom life and, as a result, hinders the achievement of equal opportunity. Selection was still practised, although disguised, in several forms. Access was highlighted by the fact that everyone could take the test. But the technology of testing reinforced selection rather than access. Bourne (1994) indicated that selection in earlier stages of education, in the form of streaming, was also a technology of social selection. Children were still streamed by ability into different classes: A, the brightest group, B, the middle ability group and C, the lowest ability group. Streaming had operated for many years in junior and infant schools in the U.K and U.S.A. Streaming was seen to be relevant to students' social background. Moreover, as Gipps and Murphy (1994) declared, streaming was found to be an important determinant of success in the later selection process of secondary school. Although this practise may be less popular, Bourne (1994) throws light on another 'masked' form of social selection which seems to take place in primary and junior schools. Teacher stereotyping of some social groups tends to differentiate children by ability, labelling them, and providing them with a differentiated support.

This implied practice also raises questions about reconciling quality and equality. Attention to these two themes - and their reconciliation - fostered new thinking

about the validity of assessment practices. A search began for an assessment approach which would more generally guarantee individuals' rights to a better education. It became a search for more individualized assessment. It was, as Wood (1987) suggested, a

> search for approaches to measurement which would be reflective of and responsive to what is peculiar to education - learning and instruction and school life - and would also be in the best interests of individuals (p.336).

By the late 1970s, it began to be recognised that the Tyler-Bloom-Glaser movement did not guarantee improvement in schooling. New thinking about assessment and classroom life was also supported by the growing influence of constructivist theories of learning.

Assessment As Pedagogic Process

A new era in assessment began in the late 1970s. There were growing concerns about the quality of education and achievement in schools in the U.S.A, Britain and other countries. Attention was drawn to the unintended outcomes of traditional forms of assessment, and the restricting effects they imposed on instruction and students' learning outcomes (Fredericksen, 1984 and Linn, 1983).

These arguments took place within a wider political and professional debate about the role of assessment in improving the quality of education. Improving the quality

of education can only be realized when education corresponds to the social imperatives and the changing situations of a society. Recently, social imperatives have changed tremendously. Rapid technological change means that the kind of individuals a society needs is becoming unpredictable. Individuals, therefore, need to be equipped with a wide range of skills and qualities to enable them to take a variety of roles in the future society. Further, enhancing intellectual and social qualities becomes more critical than distributing factual knowledge and information.

> [T]he aim of education is to help the individual become a competent intellectual performer, not a passive "selector" of orthodox and prefabricated answers (Wiggins, 1993, p. 202).

These circumstances must also be reflected in assessment practices. The argument, therefore, was that assessment should be revised to perform multiple purposes: to monitor educational standards, to improve curriculum, to guide instruction and to promote learning.

The need for assessment approaches which actively take part in the promotion of teaching and learning has led to a re-examination of assessment in terms of what happens in classrooms. To guarantee high quality education with effective formative assessment, the processes of teaching and learning must be fully understood. Any new approach in assessment must go beyond measurement and testing. It should aim to guide pupils' learning and to improve their performance. Thus, the real problem, as Eisner (1993) pointed out, 'is *not* one of correct policy formation' but 'is *one of practice* ' (p. 224).

Assessment should be classroom-based, dynamic, ongoing, formative and diagnostic. Assessment should support teaching and learning. For these reasons, new models of assessment have come to prominence. International decline in formal assessment has also been associated with increased interest in teacher involvement in assessment which, itself, forms part of the classroom processes of teaching and learning (Broadfoot, 1979a, chapter 3).

The model of classroom assessment has been bolstered by modern constructivist psychological theory. The essential aspect of this theory is that learning involves a personal construction of knowledge. It assumes also active interaction between children and their environment. It believes that children come to school with a great deal of prior knowledge of their own experience. The child begins to construct school learning with their existing knowledge. In this sense, the child's pre-existing knowledge contributes to his/her learning and the resultant knowledge becomes part of the student's intellectual apparatus. The theory gives new explanation to intelligence and the role of environment. Learning in this respect involves personal construction of knowledge and entails active interaction between the learners and teachers. Constructivist theory, including Vygotsky's notion of the 'zone of proximal development', has offered much to learning-oriented approaches to assessment (see Wood, 1987, see also Torrance and Pryor, 1995).

Assessment within a Vygotskian framework looks forward to what children can achieve with the support of an adult. The view is that new knowledge or information must be linked to the student's existing schemata. The support of an

adult is also a key element in this concept. It is a formative approach to assessment which identifies:

> the level of task a child is ready to undertake on the basis of what he can already do, *as long as he receives the best possible help from an adult* (Wood, 1987, p. 242).

In this new model of learning, teachers and students are both taking part in the construction of knowledge. Teachers should first know the existing level of students' understanding and thus, present knowledge and skills which are amenable to the child's present understanding. Students' progress is regularly discussed with them so they are involved in improving their learning. Learners gain control over their learning, and knowledge and facts become part of their personal understanding instead of being distinct and isolated. In this sense, learning is scaffolded (Bruner, 1985) in that tasks are provided for children with support from an adult, and that the objective is clearly identified for both teacher and learner.

This view of teaching, learning and knowledge has implications for assessment. Assessment is intended to foster rather than evaluate the learning-teaching process. The new view requires a variety of methods to look into the quality rather than quantity of student's learning. Terms such as 'scaffolded learning' are extended to produce scaffolded assessment which aims to 'move beyond static assessment of what is known to be a more interactive model looking at learning potential (in Vygotskyian terms, rather than in terms of 'intelligence') ' (Gipps, 1994, p. 27).

At the present, this movement linking assessment and pedagogy is widely

encouraged. A number of assessment methods have been suggested which are compatible with this new philosophy of teaching and learning. Broadfoot (1995) identified five features of the new international culture of assessment:

> • An increasing emphasis on formative, learning-integrated assessment throughout the process of education.
>
> • A commitment to raising the level of teacher understanding and of expertise in assessment procedures associated with the devolution of responsibility for quality assurance in the certification process.
>
> • An increasing emphasis on validity in the assessment process which allows the full range of curriculum objectives including cognitive, psychomotor and even affective domains of learning to be addressed by the use of a wider range of more 'authentic' techniques for gathering evidence of learning outcomes.
>
> • An increasing emphasis on describing learning outcomes in terms of particular standards achieved - often associated with the pre-specification of such outcomes in a way that reflects the integration of curriculum and assessment planning.
>
> • An increasing emphasis on using the assessment of individual pupils' learning outcomes as an indicator of the quality of educational provision, whether this be at the level of the individual classroom, the institution, the state, the nation or for international comparisons (p. 12).

Concepts of performance assessment and authentic assessment underwrite the alternative approach. Performance assessment is concerned with the task presented to students. These tasks, as Wiggins (1993) pointed out, should allow students to produce work of their own, 'using a repertoire of knowledge and skills and being responsive to the particular tasks and contexts at hand' (p. 202). This means that students should practice the 'doing' of a subject and thus be able to perform in future adult life. Authenticity, for Wiggins, is an important feature of performance

assessment. But this also does not mean that every performance assessment is authentic. Authentic assessment involves, as Wiggins suggested, the achievement of fidelity and comprehensiveness with which the validity of assessment is increasingly enhanced. Others present performance assessment in terms of forms. That is, performance assessment is concerned with the assessment of a wide range of qualities which cannot be assessed by traditional paper-and-pencil tests such as: problem solving, communication skills and critical thinking. It uses a variety of techniques such as oral, practical, research projects and experiments. Authentic assessment is also concerned with wider qualities but it is carried out in an authentic context. That is, it remains part of the normal classroom situation (Gipps, 1994, chapter 6).

Nisbet (1993) pointed out that alternative assessment can be seen within two contexts of reform, narrow and wide. In the narrow sense, alternative assessment can be restricted to the introduction of new forms of assessment, that is, portfolios, records of achievement and continuous assessment. The 1988 Education Reform Act in England and Wales has been associated with these new initiatives in assessment such as Records of Achievement (RoA) and Graded Tests (Torrance, 1989 and Gipps, 1994, chapter 6). This type of practice, in the long term, is likely to affect students' motivation and their independency. Thus, alternative assessment, in a wider sense, could be extended to include self-assessment,

> the capacity to judge one's own standards of achievement and to use that judgement diagnostically and formatively to promote further learning, as a prime aim of education (Nisbet, 1993, p. 35).

The insertion of assessment into pedagogy is still a matter of debate. Critical issues still have to be considered. The acceptance of alternative forms of assessment is related to several conditions. As Worthen (1993) points out, these include: teachers' competence to perform quality assessment, technical quality and truthfulness (whether assessment is authentic or merely authentic-looking), the acceptability to stakeholders, and their appropriateness for high-stakes assessment.

Given the impact of performance assessment on instruction and learning outcomes, the validity of such assessment must be taken into account. Messick (1988), for instance, called for an expanded view of validity. He argued that validation involves, in addition to the traditional evidential basis of validity, attention to the consequences of tests' use and test interpretation (see chapter 8 of this thesis). Consideration of validity is also identified by Linn, Baker and Dunbar (1991) as a major issue in performance assessment. Linn *et al* made a similar point to Messick.

> Considering validity in terms of consequences forces our attention on aspects of the assessment process that may not be intended or anticipated by the designers of the instruments. We know from experience that results from standardized tests can be corrupted ... It should not be assumed that new forms of assessment will be immune to such influences (p. 17).

Linn *et al* also pointed out the need to consider the issue of fairness and equity in the new approach. The shift from traditional to performance assessment does not necessarily mean fair assessment. Equity is essential in all forms of assessment and, thus, must be ensured. Six other criteria were also emphasized by Linn *et al*

by which performance assessment could be judged: transfer and generalizability, cognitive complexity, content quality, content coverage, meaningfulness and cost and efficiency. They concluded that alternative assessment models should give primacy to validity. They stated that 'the criteria for judging the assessment must correspond to the purpose, regardless of the nature or form of the assessment' (p. 20). Gipps (1994) also pointed to the same conclusion, that assessment must be guided by the purposes of assessment. That is, 'fitness for purpose' is the guiding concept: 'each assessment must have acceptable levels of reliability and validity for its purpose' (p. 103).

Other points of interest were made by Nisbet (1993). He stated three issues which raise problems in applying alternative assessment. The first point is the conflicting requirements of public accountability and instructional improvement. Using alternative forms of assessment for accountability purposes raises wider issues. They may be useful for classroom assessment but, at the same time, they may be less effective for public accountability purposes. Performance assessment, for example, relies on teachers' professional judgement and it is closely tied to individual students' needs. It does not yield a simple score which can be used for assessing standards. But comparability and generalisability still need to be addressed. 'Problems of comparability in the new forms of assessment and doubts about teachers' subjective judgements load the balance in favour of the status quo' (Nisbet, 1993, p. 130).

The second point that Nisbet made is the cost which should be rendered for

alternative forms of assessment. On the one hand, testing for accountability is a large-scale process in which the application of performance assessment is less efficient in terms of time and cost. In a large-scale assessment, testing should cover a wide range of qualities and skills, that is to ensure content coverage of the domain to be tested. Since alternative forms of assessment are mainly able to sample a small area of skills and qualities, multiple-choice items are relatively more suitable for large scale assessment. On the other hand, the use of alternative forms of assessment in classroom setting is also not without problems. It requires extra time, cost and administration work from teachers. This was also pointed out by Torrance (1995) in his conclusion derived from the first pilot trial of the new Standard Assessment Tasks (SATs) in England and Wales. He stated that

> teachers had enormous difficulty in interpreting, conducting and assessing the task - precisely because they were 'authentic', they were too complicated to communicate easily to teachers and too demanding for teachers to conduct under ordinary classroom conditions (p. 55).

Finally, Nisbet highlighted the important of teachers' professional development for achieving effective assessment and raising standards of education. The importance of enhancing teachers' classroom practices was also highlighted by Torrance (1993, see also Torrance and Pryor, 1995). He accepts that evidence for this approach is indicative of positive possibilities. But he pointed out the need for much more detailed observation-based studies of assessment. He pointed out that assessment for formative purposes is more complicated than it is presented in theoretical terms. He, therefore, emphasized the importance of teacher-pupil interaction and, in particular, the role of language in constructing teaching and learning process in

action, in which assessment for formative purposes could be articulated.

This latter argument leads to the second aspect for understanding assessment. That is, to understand its nature within the context of normal classroom practices, the focus of the next chapter.

3

Understanding Classroom Assessment

The previous chapter reviewed assessment from an historical perspective. It showed that assessment, as a concept and as a process, has undergone substantial change in the twentieth century. This chapter locates assessment within the every day activities of classroom life. It explores the question, 'What is classroom assessment?' by reference to the processes of interpersonal judgement, human interaction, and information-gathering. And it concludes by considering classroom assessment as an intentional activity.

What is classroom assessment?

Despite the extensive literature in the field of assessment, no agreed and satisfying definition for the concept has yet been formulated. The reason for this, probably, is that assessment is not a neutral activity. Assessment, as Rowntree (1987) pointed

out, is:

> awash with hidden assumptions, unstated values, partial truths, confusions of ideas, false distinctions, and irrelevant emphases (p. 4).

It is naive, therefore, to expect that there should be an exclusive definition of assessment. Writings on assessment (e.g. Erwin, 1991; Rowntree, 1987; Satterly, 1989; and Harlen, *et al* 1992) have offered different definitions, each concerned with a particular aspect of assessment. Despite their differences, however, they all make reference to four important aspects of assessment. In this all-embracing sense, assessment is

- concerned with the appraisals of learners;
- based on a comprehensive review of qualities such as intellectual competence, affect and behaviour;
- part of classroom informal activities.
- directed towards taking further actions, mainly concerning the learner.

The above synthesis is, however, also offered by Erwin (1991) in an extended definition of assessment as

> the systematic basis for making inferences about the learning and development of students. More specifically, assessment is the process of defining, selecting, designing, collecting, analysing, interpreting and using information to increase students' learning and development (1991, p. 15).

This definition shows how assessment is part of the whole process of education. It is part of curriculum planning, of defining and selecting the knowledge, skills and qualities that student should possess. It is part of the instruction process, of presenting knowledge and skills in a way that fits students' needs and aspirations.

It is part of any interaction between teacher and students, of collecting, analysing and interpreting information. It is, finally, a purposeful activity which uses such information to improve students' learning.

Such a definition may also mean evaluation. In the American literature, the term 'evaluation' can also mean 'assessment'. This is because evaluation and assessment involve similar acts: defining criteria for making judgments, gathering information, decision-making and taking-action. But there are also differences between the uses of these two terms.

The first distinction between the two terms inheres in their scope. Harlen (1980, p. 56), Rowntree (1987, p. 7) and Choppin (1990, pp. 7,8) suggested that educational evaluation is concerned with broader activities than merely the assessment of learners' qualities. Evaluation can also be applied to courses, teaching, programs, curricula and organisational measures. Any information concerning individual learners is only considered as part of a wider evaluation picture.

The second distinction between assessment and evaluation can be seen in the main interest of the two processes. Assessment is concerned with revealing the reality, or substance of the learners' quality without making judgement about the programme involved; whereas, as discussed in chapter 2, evaluation goes beyond this to use such measures as an indication of the effectiveness of a programme. In these terms, evaluation is carried out as a guide to political and social action (Cronbach, 1980, p.16).

For this reason, the evaluation process emphasizes the objectivity of measurement. Assessment, on the other hand, is a subjective process. Eisner (1993) pointed out that assessment is a new term used in education which responds to society's recent needs and concerns 'to get down to brass tacks, to go back to the basics, to measure, to monitor, to mandate' (p. 224). The joint development, evaluation theory and political pressures, generated the need for a new conception of assessment.

> [A]ssessment (the new term) needed to be more generous, more complex, more closely aligned with life than with individual performance measured in an antiseptic context using sanitized instruments that were untouched by human hands (Eisner, 1993, p. 224).

In this sense, Eisner highlights the importance of viewing assessment as an objective process mainly concerned with individual progress and development. The remaining part of this chapter considers assessment from this perspective.

Assessment As Interpersonal Judgement

The process of judgement involves the recognition of attributes of a person, categorising him/her within a certain group or characteristic (Downey, 1977, chapter. 1). Shavelson (1987) also refers to judgement as 'the process of evaluation or categorizing a person or an object' (p. 486). The ostensible aspect of personal

judgement is evident in our daily life in the use of evaluative or descriptive terms about human behaviour. That is, when describing a person, for example, as 'polite', 'quiet' or 'nervous'.

Judgement is a central element in the assessment process. It informally and intuitively takes place during the ongoing interaction between teacher and students. It takes place when a teacher tries to interact with individual students, to collect and accumulate evidence, to categorize them and to make suitable responses to these categories.

In the same way, teachers' judgements are manifest in the way they label students between negative and positive attributes. Hargreaves (1975) identifies four aspects of students' behaviour in which a teacher differentiates between pupils:

	Positive label	**Negative label**
General:	good lad	nuisance.
	Sound	Pain-in-the-neck
	Promising	Fool
	Nice	Trouble-maker
	Making progress	Going to the dogs
Instructional:	hard worker	idler
	Bright	Thickhead
	Neat	Untidy
Disciplinary:	quiet	chatterbox
	Polite	Cheeky
Peer:	Leader	Ring-leader
	friendly	bully
	Popular	Lone-wolf (p. 128).

The teacher's decision that a student belongs, or does not belong to particular categories depends on the matching process between the behaviour of what would

constitute, in the teacher's perception, an ideal attribute with the actual behaviour of the student. But the process is not as neat as it seems. Teachers' perceptions also play a very important role in considering what is a good or ideal behaviour. When a teacher labels a student positively and calls him, for example, 'promising' s/he at the same time incorporates different attributes such as, disciplinary and peer group behaviour. As Hargreaves (1975) pointed out, judgement of students' personality always incorporates a teacher's judgement of other qualities.

This is so because the process of judgement of each quality involves a selection from a wide range of information such as, the students' appearance, their home background, and their behaviour with others. Abercrombie (1960) pointed out, 'we are continually selecting from the information presented, interpreting it with information received in the past, and making predictions about the future' (p. 14). When students come to the school for the first time, teachers are supplied with initial information about them, such as their social and economic background and their previous academic progress. Such information would be combined with other information that the teacher could get by her/himself, such as the students' behaviour and physical appearance. The teacher makes a selection from this initial information, interprets it and makes an initial judgement about the students. As the teacher interacts with students, s/he gets further information. The teacher combines the new with the old information, and again develops his/her conception of the student.

This process is inevitable in the classroom. As Downey noted, judgement is part of

social life. But judgement could be biased when teachers makes an inappropriate initial judgement about a student's quality such as regarding children from working class families as dull. For this reason, more formal and visible techniques of assessment may be preferred and gain more acceptance since they provide objective information about learners. Some educators, however, view assessment as a human activity and accept judgement as part of the process. Reality is a refraction of human perception and objectivity in human sciences can not be completely achieved. This view point can be clearly seen in the following statement.

> Educational assessment, therefore, must be recognized as being a highly imprecise activity at all but the most basic of levels, and as being judgemental rather than metric in character, as requiring the making of sound professional judgements rather than of objective, mathematical measurements (Kelly, 1992, p. 5).

But judgement for a teacher is a professional and pedagogical responsibility. It should be used, therefore, in efficient ways. Judgement can be more sensitive if a teacher uses description of student's competence and behaviour rather than using value-laden words such as 'good' or 'bad'. The descriptive form can work as feedback to students as much as to teachers. Bias can also be reduced and possibly disappear. To avoid the side-effect of teacher's initial judgment, Downey (1977) pointed out that the teacher needs to realize and comprehend new evidence which might be inconsistent with her/his initial judgement (p. 5). He, for this reason, recognizes the importance of observation and responding to the changes in students' behaviour in making a fair judgement.

> Teachers in particular can learn to sharpen their powers of observation, to be ready to accept new evidence and to perceive changes in their pupils' behaviour. They can avoid making over-hasty assessment, thus labelling pupils and setting in motion a self-fulfilling prophecy, by reserving judgement until they have sufficient evidence to make fair comments on their pupils (p. 91).

This latter point explains why teachers' judgement is very important in determining and shaping the assessment process. The act of judgement in itself, as recognized in Abercrombie's definition, involves prediction of future behaviour, or in other word, expectations. Teachers make inferences about a student's present academic achievement and also about their general behaviour and, accordingly, make predictions about their future academic improvement and behaviour. These inferences form a basis for teacher expectations. Broadfoot (1979a) stressed the importance of teacher expectations as part of an assessment process that continuously takes place in classrooms. A teacher's expectation, as Broadfoot suggested, has an influence on students' subsequent success or failure. She stated that:

> assessment in the classroom is a dynamic cumulative process in which the expectations aroused in the teacher by his or her initial characterization of the pupil's home background are reinforced in the interaction of the classroom to compound failure or success (p. 113).

In general, Broadfoot's comment relates to the phenomenon of the self-fulfilling prophecy referred to above by Downey. Brophy and Good (1974) defined self-fulfilling prophecy as

> an expectation or prediction, *initially false*, which initiates a series of events that cause the original

expectation or prediction to become true (p. 35).

The self-fulfilling prophecy is an example of a teacher's stereotyping. For example: a teacher might hold the initial judgement that pupils from working class families are 'low achievers' or 'badly behaved'. This initial judgement is communicated to students through their continuous interaction with the teacher. Brophy and Good suggested a sequence of relations whereby a teacher's expectations influence student performance and, thus, work as a self-fulfilling prophecy. They explained that this might occur as an outcome of a series of cause-and-effect relationships between a teacher and students. For a valid instance of expectation, Brophy and Good's model outlined this phenomenon on a series of six steps.

a) Initial expectations are being formed using several sources of information. Some of these expectations for some students are inappropriate, 'and some are relatively rigid and resistant to change even in the face of contradictory student behaviour' (p. 39).

b) On the basis of some, or all, of the information available, a teacher might offer different messages to pupils. Some students will receive positive comments as evidence of success and others will receive negative comments as evidence of low achieving or perhaps failure. 'Where teacher expectations are inappropriate and rigid, treatment of the students will be inappropriate' (p. 39).

c) Students respond to the teacher differently as a factor of different personalities and of differentiated treatments.

d) Students' responses and behaviour complement and reinforce the teacher's particular expectations.

e) When this process continues with students of whom the teacher holds

inappropriate and rigid expectations, it will affect students' motivation, self-concepts, opportunities to learn and their relationship with the teacher. The students' behaviour and progress will gradually approximate the teacher's initial expectation.

f) Differential treatments will show differential effects and outcomes. The end of year result will be likely to fulfil the teacher's initial expectation. The initial expectation becomes self-fulfilling.

Research interest in this area was generated by Rosenthal and Jacobson's Pygmalion in the classroom (1968). In their study, the authors carried out an experimental design in one elementary school in an urban lower-class community. Their aim was to test the question whether a teacher's expectation for students' intellectual competence would serve as a self-fulfilling prophecy. At the beginning of the school year, all students in the school were pretested with a nonverbal intelligence test. Teachers did not know the nature of this test. They were led to believe that the test was one that would predict the intellectual performance in the coming academic year. Twenty per cent of students were randomly selected and presented to their teacher as those children who are likely to 'bloom' or 'spurt'. The experimenters retested students after one semester, after a full academic year and after two full academic year. The result suggested positive findings which support Rosenthal and Jacobson's teachers' expectation hypothesis. Rosenthal and Jacobson's study generated further research on the effect of teachers' expectations, not all of which supported their widely-publicised hypothesis.

Their research results, as Rogers (1982, p. 37) pointed out, do not provide an immediately obvious and consistent picture of the phenomenon. It is naive, therefore, to state that any expectation that a teacher holds about his/her students will necessarily bring changes in a student behaviour. As Rogers (1982) pointed out

> There will be many occasions where no suitable expectations will be held by the teacher for a particular pupil, and even if there are, there is no a priori reason to suppose that those expectations will give rise to the sequence of events that will lead eventually to a self-fulfilling prophecy taking place (pp. 43, 44).

The Rosenthal and Jacobson study, however, indicates that a teacher's expectation might work in a positive way. But whatever their conclusions, they did much to draw attention to the complexities of classroom assessment and teacher-pupil interaction.

Assessment As An Interaction Process

Rowntree (1987) regarded assessment as interaction process. He stated that:

> assessment in education can be thought of as occurring whenever one person, in some kind of interaction, direct or indirect, with another, is conscious of obtaining and interpreting information about the knowledge and understanding, or abilities and attitudes of that person. To some extent or other it is an attempt to *know* that person (p. 4).

This definition recalls the original Latin root of assessment, *assidere*, which means to sit beside. To sit beside children means a direct encounter and more often a close relationship with them. This term is also close to Erwin's (1991) meaning of assessment as 'a one-to-one relationship between assessor and student' (p. 14).

Rowntree's definition considers assessment as a social act which occurs in every setting of our life but it involves a conscious action to obtain information about the other person. In classroom interaction, both teacher and students are trying to find out something about each other. Teacher and pupils are observing each other, evaluating their actions and accordingly passing their own judgement. Each has a clear set of goals and roles in which interaction takes place. Interaction in teaching/learning situation, however, embodies a particular pattern of relationships. Teacher and pupils are not in equivalent positions. The teacher has a task or job to undertake. S/he has power and authority to decide a pupil's school life. Pupils are in a, relatively, passive and power-less position. Their interaction with the teacher is dependent on how the teacher defines and executes his/her role. In short, it depends, as Hargreaves (1975) pointed out, on the teachers' definition of situation; that is, the role the teacher plays in the classroom and how s/he thinks such a role should be implemented. Hargreaves (1975) identified two basic sub-roles that a teacher always plays in the classroom: teacher as instructor and teacher as disciplinarian (p. 117). He pointed out that a teacher perceives and evaluates pupils against these two sub-roles. S/he assigns rules to the pupils which conform to his/her own definition. Students' conformity to the teacher's rules form a basis for the teacher's evaluation of his/her students and also his/her relationship with them.

Using a symbolic interactionism argument, Hargreaves (1975) explained how teacher's interactions work as an assessment process.

> On the basis of further interaction with the pupils and repeated perception of them, he develops a conception of individual pupils (and classes) who are evaluated, categorized and labelled according to the degree to which they support his definition of the situation. He then responds to pupils in the light of these evaluative labels (pp. 129 - 130).

Patterns of teacher-pupil interaction have been examined in several classroom research initiatives. A well-known device to examine this issue is Flanders Interaction Analysis Categories (FIAC). Flanders and Amidon (1965) designed a system for classifying teachers' verbal interaction in the classroom. Their interaction system contains three main categories, teacher's talk, pupil's talk and silence or confusion. Patterns of teacher's interaction were also identified in another 10 sub-categories. These are:

Teacher talk	1. Accepts feeling. 2. Praises or encourages. 3. Accepts or use idea of students. 4. Asks questions. 5. Lectures. 6. Gives directions. 7. Criticizes or justifies authority.
Student talk	8. Student talk-response. 9. Student talk-initiation.
Silence	10. Silence or confusion (p. 12).

Looking for implications in such analysis for teachers' assessment, these categories still do not tell much about the nature of interaction. They, therefore, require deeper consideration. In a typical classroom situation the situation is complex and

dynamic. The teacher deals with 30 to 36 students of different personalities. The teacher is required to balance the needs of individual students together with covering the content of the subject. This is constrained by the limited time and other resources available. In reality, therefore, a teacher can only interact with each student for few minutes or even few seconds. Galton *et al* (1980) called this the 'asymmetry' of teacher-pupil classroom interaction.

> (W)hile the typical teacher spends most of the lesson time interacting with pupils (either individually, as a member of a group, or of the class), each individual pupil, by contrast, interacts with the teacher for only a small proportion of his time (p. 60).

In such situations, it is important to have some kind of perception of what are the important aspects of teacher-pupil interaction and how these are going to be illuminated in classroom observation. There are several important aspects in which teachers' interaction with students could be further examined. The aspect of the teacher's questions in the classroom is related to the teacher's approaches in assessment. For example, what types of questions the teacher asks and to whom and what type of answers s/he accepts from students. Such aspects reveal the convergent or divergent approach of assessment the teacher adopts in the classroom (see, for example Torrance and Pryor, 1995). Another important aspect in studying the pattern of teacher-pupil interaction is differences in the form of teacher responses. For example: what kind of support, feedback and praises characterizes a teacher's interaction with students.

The differentiated pattern of interaction with individual students, with groups and with the entire class is also relevant to assessment practices. Assessment in the

classroom can mainly be discerned in the one-to-one interaction. In this sense, individual differences are an important fact of human interaction which is also recognized in several classroom studies. Mortimore *et al* (1994) identified several dimensions of difference among learners. These are, age, social class, sex, race, ability and behaviour. They stated:

> Any pupil appearing in the less positive category of all or more of these dimensions may be more likely to be treated differentially (p. 107).

Brophy and Good (1974) also reviewed several studies that focus on teacher interaction with individual students or with separate sub-groups of students. They reported many students' attributes which have been found to influence a teacher's perception of students and their patterns of interaction. These are:

> Group differences: Social class, race and sex.
>
> Individual differences: Student achievement, students' personality, physical attractiveness, seating location, writing neatness and speech characteristics (chapter. 1).

One study which followed a deeper approach and threw some light on this issue is Keddie's 'Classroom knowledge' (1971). In this study, the author used a classroom observation technique, together with tape-recordings and questionnaires, to study teachers' ideological conceptions of knowledge and how this conception is linked to students' ability. The study was conducted in a comprehensive secondary school where students were streamed across three ability brands A, B and C. The course under concern was an undifferentiated programme in humanities. The author found that although teachers were committed to the integrated curriculum, they tend to

hold different conceptions of what counted as suitable knowledge for A, B and C pupils. Keddie showed how teachers hold different ideological conceptions about A and C pupils. For example, she explored two different patterns of teacher's response to the pupils' questions. Questions from A pupils are approved whereas questions from C pupils are dismissed. She argued that this practice was a result of teachers accepting pre-existing categories relating to educational ability.

> It seems likely that the hierarchical categories of ability and knowledge may well persist in unstreamed classrooms and lead to the differentiation of undifferentiated curricula, because teachers differentiate in selection of content and pedagogy between pupils perceived as of high and low ability (p. 156).

Individual differences are, therefore, of importance in shaping the pattern of teacher-pupil interactions and, thus, the conduct of classroom assessment.

Assessment As An Information-Gathering Process

Definitions usually agree that assessment is a process concerned with gathering information. But this process relates to two distinct areas of assessment - the what and how of assessment.

What to assess?

Rowntree (1987) points out that the question of what to assess must proceed the

questions of how to assess. Wood (1991) also asserts that 'only by spelling out what people were supposed to know and be able to do would it be possible first to test and then to report in those terms' (p. 84). This consideration highlights the issue of determining clear sets of objectives for teaching. Specification of objectives is essential for several reasons. It facilitates the identification of goals and objectives for planning the curriculum. It also permits the selection and organization of the content and instruction. Finally, it allows the assessment and evaluation of outcomes to take place. Objectives, for this reason, need to be clearly specified. Popham (1969) associated the clarity of objectives with evaluation process. He noted that:

> Clearly stated instructional objectives permit us to discern more readily whether an instructional program's goals have been realized, thus, permitting an evaluation of the program's worth in promoting those goals (p. 48).

Objectives provide teachers with criteria on which they can base their own judgement. It makes teachers interested in students' progress in relation to predetermined criteria rather than in relation to the achievement of other students. Determining objectives in advance also helps both teachers and students to have a clear idea of what are the goals of each task. Thus, objectives can be discussed with students in advance, so they can orient their efforts towards achieving them. By such a process, therefore, a common ground can be established between teachers and students.

Such a criterion-referenced approach may be intuitive in the classroom. Every teacher has a set of objectives to achieve for his/her class. They orient their

teaching and assessment towards achieving these objectives. But the process also involves other considerations. Two important aspects of the specification of objectives are worth considering. The first is the issue of how the objectives of education can be specified. The second is, what are the impact of such specification on learning, teaching and assessment.

The desired outcomes of education are so broad and so numerous as to defy full specification. There remains, therefore, the question of whether it is possible or realistic to articulate objectives for all the desired outcomes of education. It has been argued, however, that such a thing can, indeed, be achieved for many educational outcomes. As Ebel (1969) pointed out:

> A weakness of many statements of educational objectives is their abstractness, generality and ambiguity. One way of overcoming this weakness is to define the goals in terms of the overt behaviour which indicates achievement of the objective. This helps to make the statements concrete, specific and definite (p. 485).

If such a task is possible, as Ebel and others agreed, the impact of specification of objectives on teaching and learning remain to be considered. Formulating objectives for the desired outcomes can result in a large number of objectives that are impossible to assess in their entirety. Teachers can not pursue all these objectives, either in their instruction or in assessment. Some objectives, when they are highly specified, might seem too insignificant to be assessed. That is, they are treated as unimportant. Teachers may trivialize and disregard them. Some objectives, therefore, would be sacrificed for others.

On the other hand, some objectives of education, as Eisner (1985) pointed out, are not amenable to measurement (e.g. creativeness and novelty). In such cases, assessment might sacrifice these objectives because they require personal judgement rather than measurement. Thus important objectives would be neglected. In Wood's terms the use of highly specified objectives in terms of observable behaviour which can be measured and tested, is to run the risk of assessment driving the curriculum (Wood, 1991, p. 3). It narrows the curriculum and instruction to what could be tested, ignoring other objectives which could not be specified in such narrow terms.

The chief value of the pre-specification of objectives of instruction is, however, that it provides teachers with an explicit characterisation of what should be achieved. At the same time, it is important to recognise the limitation of specifying objectives in advance. Specified, short-term educational objectives may be a necessary, but not a sufficient, condition for effective assessment. By becoming sophisticated in pre-specification, teachers may be inhibited from assessing unexpected outcomes or the broader objectives of education.

How to assess?

This is concerned with the context of assessment procedures. Teachers in the classroom engage in informal assessment as much as they engage in teaching. They obtain information; they interpret, and they act accordingly. Assessment is an ongoing process, integrated with teaching.

On the other hand, teachers might put students into formal assessment situations. They will conduct tests, mark them, describe the results in terms of numerical or lettered scales and use norm-referenced criteria for distributing assessment scores. Assessment in this mode may be continuous, but it is, relatively, separate from the teaching and learning process.

In classroom contexts, teachers have a variety of techniques for collecting information (e.g. oral, written test, application activities, classroom observation and students' work). The teacher's choice of assessment techniques depends on the selection of objectives. Thus, the justification of the use of an assessment technique can only be derived from its contribution to assessment and whether it gives valid information about the attainment of particular objectives. Whenever a teacher focuses on narrowly specified objectives, they might be encouraged to choose one technique, such as a test. But when broader objectives are considered, a wider range of techniques might be more appropriate.

The selection of an assessment technique is not, however, an end in itself. There are other technical problems to be considered, notably, the validity and reliability of assessment. It is important, for instance, to ensure that test items match the intended outcomes and not something else. Items which assess students' ability to "identify numbers up to 100" are not the same items when the objective is to "write numbers up to 100". Too often, assessment of this kind of objective (highly specified basic skills) is more reliable than the assessment of general objectives (e.g. problem solving).

On the other hand, the assessment of all specified objectives within a particular construct does not imply the assessment of the construct itself. For instance, teacher's assessment of isolated skills in writing (spelling, constructing different sentences, and recognising vocabulary meaning) does not mean that a student who could perfectly accomplish such an assessment is able to write an extended essay. In this sense, the teacher will need to address other modes of assessment.

The main question arises: how can the reliability of teachers' assessment be maintained without sacrificing their validity? There is an extensive literature in this problem. (see, for example, Harlen (1994) and OECD, (1993). In the context of classroom assessment, then 'how to' questions are not only related to the gathering of specific information about the 'what' of assessment. They are also related to a clear understanding of the superordinate purposes or intentions of assessment. These, too, help to shape the pattern of classroom life.

Assessment As An Intentional Activity

The published literature has tackled this issue from different perspectives. Erwin (1991), for example, grouped purposes into four major areas, political, economical, educational and social (p. 2). But he later grouped them into two major purposes: purposes for improvement (formative) and purposes for accountability (summative) (p. 7). Rowntree (1987) lists six main purposes for assessment. These are:

selection, maintenance of standards, motivation to students, feedback to students, feedback to teachers and preparation for life (chapter. 2). Gipps and Stobart (1993) explored six main purposes of assessment: screening and diagnostic (termed as professional uses), certification and selection (termed as managerial uses), and record keeping and feedback on performance (serving both professional and managerial purposes) (chapter. 2).

Others have judged assessment from a functional view point. Pennycuick (1990, p. 114) devised a map of assessment functions with two dimensions: formative/summative and individuals/group as shown in the table below.

	Formative functions	**Summative functions**
Individuals	Student motivation Monitoring, feedback and guidance Diagnosis and remediation	Selection and social control Certification and qualification Prediction
	Clear recording and reporting of student attainment	
Group	Curriculum evaluation Feedback on teaching Teacher motivation	Curriculum control Accountability Standards (p. 114)

These functions are actually perceived by Erwin, Rowntree and Gipps as purposes. It is important, at this point, to make a distinction between purposes and functions of assessment. Purposes are actually perceived by society and individuals as hopes of the future but they may not always be realized. Functions are the real - perhaps unintended - outcomes of assessment and possibly need to be uncovered. When purposes are achieved or fulfilled, they might be perceived as functions but they still remain purposes. Functions may not always be positive as assessment

sometimes brings undesired outcomes.

Scriven (1967) was the first to draw attention to the distinction between formative and summative evaluation. Although Scriven's distinction was made in the context of curriculum, the two terms are employed in the context of learners' assessment. The main distinction between formative and summative assessment is between the assessment of the on-going process of learning for the purpose of improving such learning and the assessment of the product of learning for the purpose of evaluating the merit of the completed process.

At the present, formative purposes are breaking down the traditional practices of assessment. Theoretically, formative assessment seems to be part of the learning and teaching process. It is a continuous, ongoing process and could not easily be realized in isolation. But how can formative assessment be applied in the real teaching /learning context? Harlen *et al* (1992) provided a rationale for formative assessment in classroom practice. Formative assessment, they suggested, includes three stages: First information is gathered about the learner's existing ideas and skills. At this stage, the teacher identifies the competence required in subsequent learning, and the students' initial weaknesses. Second the teacher matches learning experiences with the pupil's abilities. The teacher provides learning tasks and activities which are not too easy and not too difficult. Meanwhile, the teacher also provides guidance and support. Assessment, in this sense, is to help and not to grade students.

The third stage of assessment for formative purpose is that the teacher provides effective and continuous feedback to students. Gipps (1994) pointed out that formative assessment 'involves using assessment information to feed back into the teaching/learning process' (p. 124). Feedback is important for both teacher and students. The teacher needs to know how students are performing their tasks to identify their learning difficulties and to remedy their weaknesses. Students need to know where they stand, not in relation to each others but, in relation to a clear criterion of performance. They need to understand their strengths and weaknesses in order to improve future learning.

Finally, Harlen *et al* (1992) recommended that, for formative assessment to be effective, curricular objectives should be articulated as intermediate goals on a continuum rather than in the form of separate statements of attainment. In this sense, they argued, teachers could relate students' learning to a criterion-referenced description of progress and could, therefore, anticipate the course of further development.

Formative assessment has an appeal in that both teacher and students hold a common understanding of what are the desired outcomes and how they might be achieved. The potential promise of formative assessment, as Torrance (1993) pointed out, is that it takes account of the role of teacher-pupil interaction in the learning process. Implicit in the formative assessment is the constructivist perspective to teaching and learning. What is important in this approach is that the teacher is expected to provide appropriate scaffolding in the learning process,

which might be eliminated in subsequent tasks as students become independent. Torrance, however, argued that studies of constructivist approaches to learning are generated from experimental settings or small-scale classroom work and that issues of assessment are not addressed explicitly. Torrance, therefore, drew attention to the need for much more classroom interaction studies which 'provide a much firmer basis of evidence about the relationship of assessment to learning which can inform policy and practice over the long term' (p. 341).

In contrast to formative assessment, summative assessment is concerned with summarizing information about the achievement of students and reporting to different interest groups, individuals, parents, employers, and policy makers (Harlen *et al*, 1992). That is, summing-up students' work is a key feature of summative assessment. Such a purpose may be important because it provides a summary of manageable and readable information to all parties. It provides a broader view of the performance of a system as a whole. It can be used, therefore, for public, national accountability purposes (see chapter 8 in this thesis).

Summative assessment looks to the products of students' work. It, therefore, disregards the learning process, the ways students organise, interpret, and formulate information. The weakness in such an approach to classroom assessment is that the teacher is prevented from examining students' working practices and from early detection of weaknesses. A great deal of feedback between teacher and students is lost. The students' work is only reflected in their scores. These scores have little value for improvement and also may be interpreted differently by different parties.

They do not report student's progress over time and thus, may distort understanding of the teaching and learning. For example, the work of one student may have progressed from grade C to B and then to A, while the work of another student may have fallen back from A to B to C. Both students would have the same average score at the end of the year, but their different progress is not indicated.

Formative and summative purposes, however, may operate together in teacher assessment. Teaching often includes a blend of formative and summative practices. Teachers use a variety of techniques ranging from the informal to the formal. These could embrace the improvement of student learning and the aggregation of cumulative scores. Harlen *et al* (1992), for example, distinguished between summing up and checking up. The former is the cumulation of formative assessment, whereas the later is the collection of new information for summative purposes.

> The former is some form of summary of information obtained through recording formative assessment during a particular period of time and the latter the collection of new information about what the pupil can do at the end of a period of time, usually through giving some form of test (p. 222).

Summing up attempts to blend summative and formative assessment. But it is not clear how it can provide reliable information for different interested groups. The main issue is, thus, how to achieve quality assessment and reconcile validity and reliability. The answer, as Gipps (1994) pointed out, is consideration of 'fitness for purpose'. When assessment is to be used for summative purposes (e.g.

accountability) reliability must be considered. When it is to be used for formative purposes, then validity should receive priority.

Conflict between reliability and validity and its implications for summative and formative purposes already exists in Bahrain primary schools. Interest is moving towards the improvement of learning and the enhancement of teacher assessment. But this new view of assessment for the improvement of teaching and learning contrasts strongly with the long established view of assessment for public accountability. The next chapter explores how such views have fared in the recent history of Bahrain.

4

An Historical Overview of Primary Schools and Assessment in Bahrain

Schools are linked with the wider society. They are subject to constraints from society in general and government in particular. School and classroom experiences are the expression of the interests and expectations of society as a whole. Thus, to understand classroom practices, it is important to examine their circumstances, their culture and their historical heritage.

Bahrain is part of Arab society. It shares the same culture, language, religion and history. Most Arab societies are muslim and they consist of two broad communities, Sunni and Shia. Bahrain, however, has other distinct features. It is a small country in terms of population and size. In 1991, the population was 516,446 of which 346,019 were Bahainis and 170,427 were non-Bahrainis. Non-Bahraini

inhabitants, mainly in the labour force, come from the Far East, India, Pakistan and some other Arab countries.

This chapter traces the development of the Bahrain educational system in the 20th century with an emphasis on primary education. Four phases of development can be distinguished. The first is the initiative period which started in 1919 and marked the beginning of modern public schools in Bahrain. The second is the configuration period which began around 1940, following the discovery of oil in 1932 and the creation of oil refineries in 1936. The third period is the transitional period which started around 1971 when Bahrain became an independent state. And the final phase is the progressive period which started in 1982, and is mainly characterized by the introduction of the class-teacher system in primary schools in 1983. These four periods will be discussed in turn with particular reference to school assessment. Table 1 indicates the pattern of school enrolment between 1940 and 1991.

Year	Primary students	Total students	Population		School age
			% of total population	% of school age population	
1940/41	1855	1900	2.6%	N.A	6 - 14
1950/51	5422	5555	16.4%	N.A	6 - 14
1960/61	18913	20409	16.6%	66.4%	6 - 14
1964/65	30428	35715	24.8%	82.8%	6 - 16
1970/71	36113	50072	28.1%	93.7%	6 - 17
1980/81	44109	92724	38.8%	101%	6 - 17
1990/91	57612	100658	31.1%	103%	6 - 17

Table 4.1. Enrolment development in Bahrain from 1940/41 to 1990/91.

The Initiative Period

Bahrain is an archipelago of thirty three small islands with a total area of 694.15 sq km. It is located in the centre of the Arabian Gulf. The largest island is Bahrain, where the capital Manama is located. Bahrain is twenty-four kilometres off the east coast of Saudi Arabia and as important, on the trade routes linking Europe, the Far East and Australia.

In the early 20th century, the country was economically poor. Pearl diving was the oldest industry, on which the country had been dependent for a long time. Fishing and trading were the other main sources of revenue. The requirement in these jobs was not for advanced education as much as for honesty, sincerity and rudimentary knowledge in reading, writing and calculation.

At that time, the majority of population in Bahrain was illiterate. Education was non-formal and its ideology was mainly religious. AL-Kuttab was the name of the Koran non-formal schools and they were operated separately by local Mullahs. They operated anywhere, homes, mosques or shop corners. The school focused on memorizing the Koran and inducting children into its moral values. In addition, some tuition in reading, writing and arithmetic might also be offered. Memorization was the main method of teaching, integrated with corporal discipline, rewards and punishment. Mullahs were paid by families on a weekly basis, and they also receive another payment when the child was able to recite the whole Koran.

Besides these schools, mainly attended by the local and ordinary people, a few national and foreign schools existed. They were attended by the children from the families of foreigners and wealthy Bahrainis.

The importance of formal education for Bahrain society was recognized by some privileged citizens who began to think of establishing modern schools. They established the first modern school for boys in 1919. The school was named Al-Hidayah which means 'guidance', and it retains this name to the present day. It was located in Muharraq, the old capital of Bahrain and was run by the Sunni community. The syllabus of the school was adopted from other Arab countries, mainly Syria, Egypt and Lebanon. In the following decade, another five schools for boys were opened. By contrast, female schooling received less support. The first school for girls was not opened until 1928.

These modern schools were financed by many individual merchants, and subsidized by the government. School organization was in the hand of committees mainly chosen for their financial standing. The demand for education was limited. Most children were needed to work with their parents in trading, farming and fishing. The enrolment in these schools did not exceed 500 boys and 100 girls in 1930/31 (The Government of Bahrain, 1930).

Changes in management and organization of the modern schools came when the first Shia school was opened in 1929. This school adopted the Iraqi curriculum while the Sunni schools adopted the Syrian curriculum. The emergence of the two

systems was a reflection of social reality. The two communities have different life styles which would be reflected in their schools' ideology. The government did not approve of such separation. The chance for central control came when members of the Shia community resented an inspector who was appointed by the Sunni committee. The government, therefore, decided to take over active control of schools in 1932 (The Government of Bahrain, 1956).

The central Department of Education was given responsibility for the formation and execution of education policy. In the 1930s, the extension of formal education was slow. The value of such education had not yet been comprehended by many people. By the end of the 1930s, only ten schools had been opened in main towns and villages. Reports for that period, however, are vague about the first age for attending school, the official levels of education, and the number of years for completing schooling. These reports, however, referred to some sort of lower classes, (Atfal/Infant) which provided elementary subjects, and top classes which provided more advanced subjects. These two levels seemed to be considered as primary education, the only level existing at that time.

Two sub-systems were operated within the central Department of Education, the Department of Education for Boys' schools and the Department of Education for Girls' schools. These two sub-systems reflected a social situation which segregated females from males in all aspects of life. This segregation was also reflected in different educational aims and curricula. Education for girls conformed to the usual domestic role for women. Work outside home was only accepted in two fields,

medicine and education. Thus, girls' schools provided different subjects. Arabic, mathematics, religion, sewing and embroidery were the main subjects offered. In boys' schools, in addition to the core subjects, Arabic, mathematics and religion, the curriculum included additional subjects such as, history, geography, algebra, geometry, chemistry, physics and music. English language was also provided for the top classes in boys' schools. This differentiated programme for girls' education was obviously noted - and accepted - in the government annual report (The Government of Bahrain, 1940).

> The chief object of the school is not, as in the case of boys' schools, to train girls to earn their own living, because at present, except in the Education and Medical Departments, there is no employment for women. The aim of the schools is to teach the girls better methods of managing their homes and bringing up their children (p. 33).

Differentiation also existed between village and town schools. Village schools provided an inferior programme compared with town schools. The situation in all schools, however, can be associated with Beeby's (1966) first stage of development "the dame school". They, in this sense, shared much of Beeby's definition; unorganized school, vague syllabus, narrow subject content, memorizing and ill-educated teachers.This situation was highlighted by Adrian Vallance, a British expert who was invited to the country in 1939 to investigate and review the efficiency of schools. He also felt that primary schools were ill-resourced and inefficiently organized:

> These little boys are given no books from which to study, nor are they provided with any kind of manual occupation, so important in the early education of children. They are expected to sit hour after hour, morning and afternoon, watching each other's efforts

> with a piece of chalk on a blackboard which only a few of them can see. At other times they spend long hours copying from printed books words or sentences of which they do not even know the meanings. The amount of education imparted by these methods is practically nil - in fact this is not education at all, and such a curriculum can only succeed in implanting in a child's mind a deep hatred of school and study, and all connected with it (Vallance, 1939, p. 9)

The discovery of oil in 1932 and the emergence of an oil industry in 1936 fostered a new concept of literacy. Arabic reading and writing was not enough to get a job in the labour market. Koran schools, although still existing, did not meet the new requirements. English language became important for working in the new sectors. Technical knowledge emerged as another important requirement. As a result, the first technical school, attached to the Department of Electricity, was opened in 1936. The importance of new ideas began to be realized and the public demand for modern schooling started to increase.

With increasing opportunities for employment, education became more important and a minimum formal schooling became necessary to enter the labour market. The aim of primary education, the only level of education known at that time, was acquiring employment in the labour market. The intrinsic social value attributed to formal education was low.

> I could not find, anywhere in the island, any sign whatsoever of a desire for education for its own sake... education is required in Bahrain solely as a means to an end, the end being a lucrative post with the Oil Company. And when one considers the conditions of grinding poverty under which the majority of the people are living, this is a view of education which cannot be too hastily condemned (Vallance, 1939, p. 31).

As mentioned above, schooling was poorly co-ordinated and inefficient. It could not meet the social and economic demands placed upon it. In his report, Vallance (1939) identified seven causes for the inefficiency of schools at that time: premature leaving, irregular attendance, overcrowded classrooms, lack of textbooks, antiquated methods of teaching, lack of sound grounding and utter neglect of the village schools (Vallance, 1939, pp. 4-12).

The Department of Education realized that the existing schools could not respond to the new situation. The status of schools needed to be reviewed and a scheme for educational development was essential. This recognition marked the beginning of the next period in educational development.

The Configuration Period

In the early 1940s, male literacy was only 8.7 percent, while female literacy was less than 2 percent. The total number of students enrolled in primary schools is reported to be 667 girls and 1188 boys in 1940/41, only 2.6% of the total Bahraini population (see school enrolment in table 1).

The low literacy level and the limited capacity of educational provision were serious obstacles to educational development. A new system was proposed which had two tasks. First, to provide a basic common education for ordinary people to

help them acquire jobs, mainly technical, in the labour market. Secondly, to provide advanced education for a minority of intelligent children. The educational system was established around a new pattern. It consisted of three levels of schooling: infant level from the age of 6 to 8, primary level from age 9 to 11 and final level (secondary) from age 12 to 14. A curriculum programme for each level of education was organized around a subject based syllabus.

This pattern was also associated with the beginnings of secondary education. The first secondary school for boys, called Al-kullyah (The College), was opened in 1940/41. In the first year, it accepted only 33 boys since students who left primary schools were employed directly in the labour market or joined the technical school.

The limit on places available at all levels of schooling led to the introduction of procedures to regulate student progress. Three policy regulations were proposed, and subsequently, adopted:

> 1. No child under the age of 6 is to be accepted in school.
> 2. No child is allowed to spend more than four years to complete a 3-year course.
> 3. Examinations at the end of each stage are to determine those who are allowed to enter the following stage (Vallance, 1939, p. 29)

According to these policies, two types of selection were undertaken. Selection could take place during and at the end of each stage. At the primary level, for example, those students identified by their school as dull students were eliminated

from school before completing primary education; whereas those students who proved, in the final examination, that they were qualified for the next stage were accepted to continue their education.

These changes in primary education were, as noted, influenced by two developments, the expansion of employment chances associated with the oil industry, and the establishment of secondary schools. Primary education, therefore had to fulfil two functions: to provide a rudimentary education for boys to enter the labour market and girls to be more qualified as housekeepers, and a foundational education for the minority of boys who wished to advance to secondary education. Many boys who completed their primary schooling gained work in the labour market as soon as they left school. Some girls were also accepted to work as teachers when they completed their primary school. Secondary education places, on the other hand, were restricted; and admission was based on ability and merit. Al-Kullyah, therefore, aimed to

> provide more advanced education for a limited number of intelligent boys who entered it by competitive examinations from the primary schools (The Government of Bahrain, 1946, p. 36)

These policies of examinations and selection, as in other parts of the world, were justified against the limited places available in both primary and secondary education. As a result, examinations became the responsibility of the central Department of Education, as had been suggested by Vallance: 'examinations themselves should be conducted by the Education Department, away from the schools' (1939, p. 56). Examinations began to be used to legitimate policy, again

with the cooperation of employees in labour market, as recommended by Vallance.

> [W]e must ask the Oil Company, and other employers, to co-operate with us, and to refuse to take into their service any schoolboy who cannot produce a government certificate to the effect that his education is sufficient (1939, p. 5).

A certificate of completion of primary level became necessary for entry into the labour market as well as to acquire a place in secondary school. Examinations enhanced and fostered the credential view of education mentioned earlier. Passing the examination, as noted by Al-Hamer (1969), had 'become an end rather than a means in the educative process' (p. 79).

The examination system followed the same pattern as in Egypt. It was highly bureaucratic and centrally administered by the Department of Education. Periodical and final tests took 75 percent of the total mark while 25 percent was allocated to the daily work (Al-Hamer, 1969, pp. 79, 80).

The system of examination was regarded as innovative at that time. It affected other aspects of schooling, leading to the standardization of teaching and the creation of a common syllabus (The Government of Bahrain, 1942, p. 43). Thus, in addition to monitoring students' progress, it promoted and fostered central control of the education system. But two parallel systems of examinations still operated within the Department of Education, as a natural consequences of the two sub-systems, for boys and girls.

The demand for education in the 1940s increased steadily while Al-Kuttab schools

began to disappear. During that time many other changes took place. The first secondary school for girls was opened in 1951/52, and as a result of expansion of commercial sector, commercial education for boys commenced at the secondary level in 1952/53. The total number of primary students rose from 1,855 (1,188 boys and 667 girls) students in 1940/41 to 18,913 (12,677 boys and 6,236 girls) students in 1950/51 (see table. 1). Secondary school in the same year accommodated 133 boys only (Al-Hamer, 1969, pp. 44, 45).

By the mid 1950s, the relative value of primary education in the labour market decreased as pressure was placed upon the secondary sector. Students in primary schools wanted to continue their secondary schooling. The existing promotion and selective policy did not fit with the changing situation. There were obvious imbalances in schools' enrolment between primary and secondary levels. In 1955/56, for example, there were only 462 boys and 39 girls in the secondary school. The relative number of these students as opposed to those in primary schools was 1:15 for boys and 1:84 for girls. It was, therefore, necessary to reconsider polices in education to rectify these imbalances.

In the early 1960s, experts from Egypt and Lebanon were invited to Bahrain to review the whole educational system. In regard to primary education, they focused attention on the importance of providing educational opportunities for all children in the country. Such a policy was a heated issue at that time. The high proportion of repeaters in primary schools represented, they believed, the main obstacle to the expansion of education since it obstructed the policy of providing more

opportunities to the new comers (Al-Nahas and Katul, 1960, p. 34). The average percentage of repeat students, they estimated, was 40 percent of the total enrolled students in the primary level. Al-Nahas and Katul were in favour of more flexible system of promotion as well as more tolerant examinations. As they pointed out

> Children in primary education need to pass as smoothly as possible through their first years of schooling. A high repetition rate only reflects that schools failed in their duties to develop the right education and that they also conducted strict examinations, in terms of questions and in marking procedures (1960, p. 34)

Upon Al-Nahas and Katul's recommendation, a new policy of assessment, including automatic promotion, was implemented in the first three years of primary school in 1962. The new assessment policy introduced continuous, school-based assessment. Regular attendance, of not less than 75 percent of the school year, along with teacher judgement of the quality of pupils' performance, were the two main criteria for promotion in years 1 to 3. Automatic promotion was implemented, and remains in place, for primary years 1 to 3.

The second decision which had been taken, in response to the increasing demand for primary education and the policy of providing basic education for all children, was to change the pattern of primary education. The infant level was joined with the primary level to form six years of primary schooling. Central examination at the end of primary school, remained for the purposes of certification. But the selection policy for the secondary level was abolished. that is, all children who acquire the primary certificate could pass automatically to secondary level.

The associated policy, which proposed reduction in central control of assessment, was, therefore, not entirely accomplished. The Department of Education retained tests as the main type of assessment in schools and preserve the central examination at the end of primary level. Retaining tests and examinations in primary schools allowed a degree of supervision over schools' assessment practices. This factor combined with poorly-qualified teachers who did not recognize the value of continuous assessment and/or understand the implications of automatic promotion, reduced the possibility that changes in assessment policies, recommended by Al-Nahas and Katul, would have an effective result on teaching and learning. As Al-Hamer explained that

> the majority of the teachers working then in the primary schools were neither ready nor properly trained to adopt such an innovation ... out of the total number of the teachers in Bahrain 92% female and 87% male teachers were **un-trained,** with an educational level less than the completion of secondary education (1969, pp. 81, 82)

Automatic promotion was not recommended for the upper grades (9 - 11 years) of primary education. The assessment policy, (Act No. 67/72/62 dated 19/2/62), which defined the basic principle for automatic promotion also described the assessment procedures for the upper three years. The Department of Education issued a directive stating that the total score of students should consist of 75% from the end of year exam and the other 25% from daily course work.

In contrast to the lower primary level, less successful children in the upper three grades were given the chance to repeat each class for another year. In the extreme case, students who were identified as slow learners or retarded, could spend a

maximum of 6 years in the upper three years of primary school. The policy suggested, however, that no child over 15 years should remain in the primary school. A leaving certificate, to be distinguished from the Primary Certificate, was issued to help student to get a job or enter the technical school. This policy also meant an expansion in educational opportunities. The risk of being eliminated from schooling because of low ability, was reduced. But the new promotion policy did not eliminate the problem of repetition. It only postponed it until later years in the primary school. On the other hand, the Act did not state clearly that repeating in the first three years should have been completely eliminated. This problem, as will be demonstrate later, remained even with the automatic promotion.

The influence of the dual system for boys and girls on the effectiveness of the school was, for the first time, indicated by Al-Nahas and Katul. There appeared to be a significant dualism not just in terms of organization but also in terms of practice. The first stage towards integrating the two systems seemed to be started when the first unified examination for both girls and boys was implemented in 1966/67. What was notable in this examination was the difference which emerged in the passing rates between boys and girls. The passing rate for boys which had been around 80% in the previous 5 years dropped sharply to 43%. Girls retained their previous passing level of 80%. This result might have raised questions about the differences in the previous assessment practice between boys and girls schooling and the reasons behind the low level of successful students in boys schools. But, there is no indication that these results attracted attention within the Department of Education at that time.

In 1962/63, secondary education was extended to five years. It was divided into two stages: intermediate (2 years) and secondary (3 years), and three lines of specialization: academic, commercial and teacher education. In the late 1960s, the technical school was also upgraded to form another type of secondary education.

The changing pattern of the secondary level was also provided a pathway to higher education. Three higher institutes were founded in response: the Teacher Training Institute for men (1966), The Teacher Training Institute for women (1967) and the Gulf Polytechnical College (1968).

The 1960s was a decade of substantial changes in enrolment. All children who completed primary education were accepted into intermediate schooling and most of those who completed the intermediate level joined the secondary schools. Extending years of schooling created further pressure to increase the available places for education. As shown in table. 1, the total enrolment increased from nearly 20,000 students in 1960/61 to about 50,000 students in 1970/71. Abolishing the selection policy adjusted the balance between primary level and intermediate and secondary levels. In 1971, the relative number of those enrolled in intermediate and secondary schools to those in primary school became 1:4 and 1:7 for boys and girls respectively.

By the end of 1960s, an educational structure had been laid down. The system was divided into primary, intermediate and secondary schools. These were centrally controlled, and followed a common curriculum developed by state-employed

teachers. Nearly 28 percent of the total Bahraini population were in school as opposed to nearly 16 percent in the early 1960s (see table. 1).

The Transitional period

Until 1970, the society as a whole had undergone rapid changes. The economy expanded and diversified, modernization penetrated every aspect of life, such that a wide gap between generations had become evident. On the eve of the 1970s, a turning point in Bahrain modern history was reached. In 1971, the British government agreed to terminate the special treaty relations between the United Kingdom and Bahrain.

Underlying this transition period, were changes in attitudes and motivation which formed the basis for new national policies in education. New aspirations and perceptions for education were cultivated in post independence years. The Bahrain view was that a national system of education and schools must promote, in addition to a universal body of knowledge and skills, a set of local and regional values and aspirations. It was accepted, therefore, that education should act as a tool of citizenship as well as for achieving economic development (Nakhleh, 1976, pp. 14, 31).

At that time, the educational system faced many issues with regard to literacy, economic performance and the formation of the citizenry. Almost 61 per cent of

the total Bahraini population were still illiterate and another quarter could barely read and write. Moreover, at least 6.7% of 10 to 14 years olds were also illiterate. Foreign labour increased until it represented 37 percent of the total employed population. The labour market was also troubled by unemployment, which was estimated to be 3 per cent of the total employed population. Finally, an annual population growth of 3.5 per cent placed additional pressure on school places. By 1970/71, 28% of the total Bahraini population were at school (36,000 in primary, 7,300 in intermediate and 6,100 in secondary education).

A review of the educational situation in Bahrain in the early 1970s was carried out by Nakhleh (1976). He identified several critical issues. The system of education, despite 50 years of modern education and the fact that over 25 per cent of the population was enrolled in schools, had 'failed ... to halt the growth of illiteracy' (p. 17). He also pointed out important differences in literacy rates between rural and urban areas and between women and men. And he noted the imbalance between compulsory schooling and higher education. He pointed out 'only 426 Bahrainis received advanced degrees (college and above) between 1950 and 1972, over 75 percent of whom were men' (p. 22). Overall, Nakhleh felt that a new era should commence if education was also to be advanced.

With the stimulus of independence, together with the rise in oil prices in 1973, several important areas were enhanced. The issue of illiteracy was tackled with a comprehensive programme for adult education. The term, 'basic education' was to include both primary and intermediate, the latter expanded to three years in 1978,

to represent a minimum of 9 years of schooling (3 + 3 + 3). This last change led to the abolition of the General Certificate of Primary education in 1979 and, as a result, the central primary examination. Central examination and certification, however, remained at the end of intermediate level. Despite the replacement of the central examination with school-based examinations, overall assessment policy in primary school remained within the pre-existing framework. Its main characteristics were automatic promotion; a one year retention allowed in the upper three years; and formal mid-year and end-of-term tests and final promotion reports.

As the expansion of school enrolment levelled off, the attention of policy-makers shifted to the efficiency of schools. Schools began to be seen as a productive system with three main phases: inputs, processes and outputs. It was assumed that, as in any other production system, improving the inputs of eduction would necessarily enhance the flow of students within the educational system and improve school productivity. Such an assumption was held world-wide educationally at that time (Simmons, 1980). Bahrain was no exception. Stimulated by the increasing oil revenues in the mid 1970s, measures were taken to enhance several inputs of basic education. These concentrated on establishing modern curricula, constructing objectives and aims for each year of schooling, improving schools' buildings, enhancing teachers' qualifications and providing educational material (eg. visual aids). These changes marked a turning point in attitudes towards primary education. Schools began to be seen as qualifying agents for productive worker and also a good citizens. Policy began to consider the dual aims of education: preparing all children for life as both workers and citizens.

From this point of view, the problem of retention in the primary school also received research attention in the late 1970s. Al-Sulayti, who was Under-Secretary in the Ministry of Education, suggested that:

> retention... hinders the normal flow of students in the educational system, inflates schools' enrolments, and thus lowers to a great extent the input/output capacities of schools, depriving some students of space in crowded classes (1983, p. 113).

Retention was estimated to be 20 percent of total enrolled students in primary school (Al-Sulayti, 1983, p. 107). The average years for completing primary education in the 1970s (normally 6 years) was estimated to be 9.51 years for boys and 8.3 years for girls (Al-Sulayti, 1983, pp. 110, 111). In the early years of primary schooling, automatic promotion seemed to have had less impact than was expected in reducing the retention rate among students. Data indicated that retention in the first three primary years was nearly 20 per cent by the early 1970s and had decreased slightly, to 16 per cent by the late 1970s (Directorate of Planing, 1981, p. 45).

Al-Sulayti recommended a revision and modification of retention practices in primary education in order to

> accelerate the movement of students in the educational cycle and, thus, promote the input/output capacity of schools (p. 115)

Nevertheless, this problem persisted and again generated concern about the quality of education. The retention problem, as shown below, was behind changes in assessment policy in the 1980s.

In secondary education, the obvious development was the initiative of commercial education for girls in 1970/71. A new class of only 38 students was opened to encourage girls for vocational education. The demand for this type of education was high in subsequent years. Total number enrolled increased to 746 in 1978/79, estimated as 18.6 percent of the total female students in secondary education.

A five-year national plan was implemented in 1978. It aimed to integrate secondary education with the requirements of the economy. The policy emphasized the acceptance into secondary education of most students who complete intermediate education. Nevertheless, the main priority was to increase places in technical and commercial schools. New vocational sections were also introduced: catering and tourism, agriculture & animal husbandry, textiles, and printing. Thus, students who completed intermediate school could be divided between several sections of secondary education. General education (Art and sciences) planned to accept only 30% of boys and 40 per cent of girls of the total secondary students. Technical and commercial education planned to take 60% of boys and 50% of girls. The new sections were expected to receive between 5% to 10% while the other students (boys only), mostly with low attainment levels would be transferred to technical vocational training outside the educational system.

By the end of 1980s, it was obvious that the plan of diversifying secondary education failed to meet the required objectives. The demand for the new sections was very limited. Until the late 1980s, total enrolment in these sections did not exceed 2.7 per cent of total secondary students.

In the early 1980s, 90 per cent of children between the ages of 6 and 11 were enrolled in primary schools. Illiteracy rate decreased steadily to 20 per cent of total Bahraini population in 1981 while those who could minimally read and write formed 28 per cent. Foreign labour increased to 81,000 and competed for jobs with the local labour force. Unemployment rate increased to 6 percent in 1981. Holding certificates was no longer enough for gaining access to an appropriate job. There was a need, therefore, for highly creative, efficient and productive workers. At the same time, education began to face budgetary pressures. In the late 1960s and 1970s, it took nearly 20 percent of the total government budget. This share had been reduced, reaching 10% in the early 1980s. While the government continued to support education, education itself was expected to play an active part in the national plan of development.

The educational advances of the mid and late 1979s, encouraged by the oil revenues, slowed down under these new pressures. The increasing demand on education still persisted. Nearly 9000 new students entered primary schools every year and this number was expected to increase in the future. Policy-makers began to consider the participation of the private sector by encouraging private schools in the country. This policy was expected to reduce the pressure on public education. In the wake of this government policy, public schools had to make concessions to the Westernised education provided in private schools. These concessions dominated the next period.

The Progressive Period

After the death of the Minister of Education, Shaik Abdul-Aziz Al-Khalifa in 1981, a new minister took over. Dr. Ali Fakhro believed that radical change in education should start at once and cover the whole process of education including classroom practice. This philosophy required educators to look at the whole cycle of education, teaching and learning, inputs, processes and outputs. This wholesale review first took place in primary education.

With the aspiration to catch up with the educational innovation in the developed world, a child-centred approach seemed to be the best alternative. It was recognized that the class-teacher is the appropriate system for implementing the child-centred approach and also would enhance every aspect of classroom life. Six educational principles were reflected in the innovations of this period. Classroom life should revolve around:

1. The Integration of a child's experience.
2. Child-centredness.
3. The environment as the main source of experience.
4. The development of the whole child.
5. Individualization of learning.
6. Independent self learning (Al-Ameen *et al*, 1993, Chapter. 1).

But the new system also required more resources which have to be planned in advance. The scheme entailed working towards a different classroom size, a new classroom lay-out, new design in school building and new kinds of teaching. To implement an integrated child-centred approach, the Minister convinced the government to allocate special funds for introducing the class-teacher system in the

first three years of primary education. In addition, a new four year class-teacher programme had been initiated in the College of Education in 1982/83. Graduates of this programme were given priority in class-teacher primary schools. In 1983/84, the class-teacher system was first put into practice in two primary schools (boys and girls). By 1994/95, the number of classrooms following the class-teacher system had risen to 690, 68 percent of the total primary classrooms.

Implementing the class-teacher system also prompted developments in the second stage of primary education. In primary years 4 to 6, the integrated-subject teacher had been implemented in 1986/87 to complement the class-teacher system. Five subjects, which previously formed the basic or foundation subjects, were combined in two groups. Humanities group; comprising islamic religion, Arabic and social studies subjects, and sciences and technology group; comprising mathematics and sciences subjects. Each group of subjects is taught by one teacher. On the other hand, English, Art and Physical education remained as autonomous subjects. Subjects boundaries, however, are maintained in the textbooks and in the weekly plan. Part of the curriculum was also organized on a topic basis for teacher to employ integrated approaches. By 1994/95, 420 classrooms followed an integrated subject-timetable, which represents 46 percent of the total primary grades 4 to 6 classrooms.

The integrated -subjects system, however, has not yet acquired a clear identity. The original philosophy of the integrated-subjects system should follow the same philsophy adopted in the class-teacher system. It seems, however, that more

resources have been devoted to the class-teacher system in a way that much of the requirements for the integrated-subjects system have been sacrificed. This appears to be true in terms of teacher training and classroom facilities. Theoretically, it can be said that the system is a short distance from the class-teacher system, but in real terms, it is only another pattern of the subject-teacher system.

The wholesale review of schooling also included assessment. A new assessment Act was introduced in 1983. But Act No. 25/168-1/83 contained organizational rather than philosophical changes. These were:

1. The abolishing of tests and examinations in the first three years, giving teachers complete freedom to choose other means of assessment. Observation and diagnostic assessment appeared for the first time as new modes of assessment.
2. Division of the academic year into two terms (4 months and two weeks each). Each term is treated separately for curriculum and examination purposes; but both are regarded as one academic year for promotion to the next level.
3. Automatic promotion of retained students in the upper three years. Students are not allowed to repeat a year more than once. The new Act, however, obliged schools to provide remedial lessons for these students. This system afforded flexible promotion opportunities for underachievers but at the same time created sub-classes, named remedial, in primary schools.
4. Retention of the testing system in the last three years of primary school. Students' marks at the end of each term are made up of 30% for continuous assessment, 20% for the mid term test and 50% for the end of term test. The overall pass mark for each subject is 50%.

This distribution of marks means that promotion in the upper three years is highly sensitive to the end of term test. Teaching is preferentially directed towards achieving good performance in this test. The weakness of the 1983 Act was that it gave insufficient attention to child-centred ideas and the six new principles of classroom life. It did little to promote child-centred classroom practices.

This problem was quickly recognized and various attempts were made to remedy this shortcoming. Notions of criterion-referenced, diagnostic and formative assessment began to circulate. Assessment criteria were recommended for each subject. Diagnostic tests were prepared in Mathematics and proposed for Arabic (Yosif, 1991). Finally, the Fifth Educational Conference in 1989 was devoted to new approaches in assessment (eg. formative assessment). But these attempts to fill the gap remained rhetorical. Use of diagnostic testing was left to schools and teachers. Teachers were not given assessment guidelines. And they received no technical assistance. The integration of assessment with teaching/learning processes received very little endorsement in practice. Procedures for assessing students continued to be those indicated in the 1983 Act.

On the other hand, the 1983 Act had a direct influence on retention. Retention rates dropped from 12 per cent in 1982/83 to 8 per cent in 1983/84 for the whole primary level. The decrease continue in the 1980s until it reached only 5 percent in 1992/93. Table 4.2 shows the retention rates for 1974/75, 1988/89 and 1992/93.

Year Sex	1974/75	1988/89	1992/93
Females	17.0	7.7	5.4
Males	21.2	8.8	5.6

Table 4.2. Retention rates in primary schools in the 1970s, 1980s and early 1990s.

These changes have also reduced the number of students who are older than grade level (age 6 to 11). As seen in table 4.3, in the mid 1970s these students represented an average of 73 per cent of total students enrolled in primary schools. A steady decrease occurred in the 1980s. By the early 1990s, the proportion of over-age students decreased to 21.5 per cent of total enrolled students.

Year	Age	1974/75	1988/89	1992/93
1st	7 years and above	52.4	17.7	9.6
2nd	8 years and above	67.4	28.6	16.7
3rd	9 years and above	77.0	33.3	19.6
4th	10 years and above	80.4	38.9	24.8
5th	11 years and above	80.9	43.5	30.0
6th	12 years and above	81.2	43.2	28.7
	Average	**73**	**34.3**	**21.5**

Table 4.3. Students above the regular grade age in primary schools.

These results indicate substantial changes in the flow of students within primary education. It reflects what Al-Sulayti (1983), had recommended in his study of students' retention.

Whether these through-put indicators reflected real learning improvement became a subject of investigation. In the early years of this decade, the Educational Research

and Development Centre (ERDC), however, produced a series of highly critical reports on education (Al-Ameen, *et al*, 1993 and Wehbeh *et al*, 1993). Doubt was cast on the quality of teachers' training, teachers' academic standards, teachers' practice and the outcomes of primary education. This research also generated a public debate about falling standards. Moreover, the ERDC's research resonated with public dissatisfaction about the low standard of students promoted to secondary and higher education. In short, classroom life came under scrutiny. Attention returned to input-output considerations, but the main focus of interest became the processes that link inputs and outputs.

Concern over this matter produced two significant changes in the areas of national assessment and formative assessment. First, the formulation of National Testing policy was delegated to a committee for assessing educational output (formed in 11/3/92, Act No. 36/169-1/92). The Committee was charged to:

a) Devise a vision and strategy for implementing national assessment test.

b) Specify an appropriate organisational framework for implementing this task.

c) Study the test results and provide feedback to the Ministry.

Secondly, new guidelines for assessment in basic education were issued (Act No. 357/A.M/94 dated 19 December 1994). The new system embraced three cycles: primary grade 1 to 3, primary grades 4 to 6, and intermediate grades 7 to 9

The intention of this Act, was that assessment should be: a productive, executive,

analytical and constructive process. It should be regarded as an integral part of teaching and learning. And it should focus on higher, as well as lower, ability skills.

Using important ideas circulating internationally at this time, teacher assessment should be naturalistic, authentic, formative and criterion-referenced. It should allow for continuous development, be integrated with the curriculum, and minimize any side-effects. Schools and school teachers should become responsible for ensuring that students master all curriculum competencies (Assessment Act, 1994).

The new assessment policy specifies the following procedures:

1. Assessment of students' performance in the first three grades (first cycle) should be formative, continuous and diagnostic. An assessment portfolio should record, subject by subject, students' actual performance, student' strengths and weaknesses and teacher's proposals improvement.

2. Assessment of students' performance in the primary grades 4 to 6 (second cycle) and intermediate grades 7 to 9 (third cycle) should be continuous and diagnostic during the term and summative at mid term and end of term.

3. Students should master at least 60% of Arabic and Mathematics competencies. Mastering competencies in Arabic and Mathematics are the minimum requirement for the progression of students to higher grades.

4. Schools are responsible for organising remedial programme for under achievers, and to vary their teaching strategies until remedial students reach the required levels of competence.

At present it is difficult to predict the likely effects of these proposals. Their gradual introduction has been managed to establish initial agreement and acceptance among schools and teachers. The timetable for their introduction was agreed as: 1995/96 for grades 1 to 3, 1996/97 for grades 4 and 7, 1997/98 for grades 5 and, and 1998/99 for grades 6 and 9.

At the end of this historical review, a brief description of the education system at the present time needs to be given.

Bahrain schools today:

The present system of education consists of three main levels, primary (6 to 11 years old), intermediate (12 - 14 years old) and secondary (15 -17 years old). Primary and intermediate levels are considered to be the basic minimum level of education that all children are expected to achieve. There is no selection until the end of the intermediate level where children should pass the external examination. Thus, all children from primary education go to intermediate level and nearly all of them continue to secondary schools. They choose the type of secondary education they want but their final marks in the different subjects determine what type of education they receive.

Children in each grade of schooling may be of different ages because of the retention policy that allows children to stay at least two academic years in each grade. In 1990/91, there were 57,612 students in the primary level (of whom 9% were above the normal primary age). In the intermediate level, there were 24,362

students (of whom 25% were above the normal intermediate school age). In the secondary level there were 18,373 students (of whom 27.5% were above the normal secondary school age).

At the primary level, two types of teaching system are implemented: class-teaching from years 1 to 3; and integrated-subject teaching in years 4 to 6. The class-teacher stays with her/his students for three years (year-one to year-three). S/he teaches five basic subjects, Arabic, religion, mathematics, social studies and sciences (20 sessions per week). The other subjects, art, music and physical education (5 sessions), are taught by other teachers.

With the integrated-subject teaching system, teachers have less independence. In effect, there are two class-teachers instead of one. Arabic teachers at these grades are required to teach three subjects (Arabic, religion and social studies, 11 sessions); and mathematics and science teachers are required to teach both subjects (8 sessions). The other subjects (English, art, physical education, music and home studies, 11 sessions) are taught by other teachers. To complete a teacher's official teaching load, integrated-subject teachers are expected to teach several classes (minimum two classes). For each class, there is a fixed timetable. Teachers move between classes according to the class time table and they may teach students for only one academic year.

Classrooms for the class-teacher system should allow a space of at least $2m^2$ for each child. The size of the class is fixed at 30 students per class. Each child should

have his/her own drawer. The classroom should be provided with a small library and visual aids for children's activities and learning. The traditional seating where students sit in rows is discouraged and group seating is the recommended pattern. Classroom furniture in the integrated-subject class is less like those of the class-teacher system. Classroom layout follows the traditional pattern where students' desks and seats are arranged in parallel rows. The average class size ranges from 32 to 35 students, and may be higher in some schools.

The school curriculum is determined by the Ministry of Education. The curriculum of primary education is derived from the general philosophy of education which stated at least 25 long-term general objectives. These objectives are abstract and general. More concrete but medium-term objectives are, therefore, organized for each stage of primary education. There are 95 objectives (competencies) for first stage (year 1 - 3), and 154 objectives (competencies) for the second stage (year 4 - 6). These objectives are descriptive and classified by subjects (see appendix 1 for Arabic and mathematics). In addition, they are cumulative; that is, objectives in the second stage build upon objectives of the earlier one.

For classroom instruction, the medium-term objectives are broken down into yearly short-term objectives. Several devices are used to focus the teacher's attention on the short-term objectives. The use of standard textbooks is central to the classroom teaching/learning process. Textbooks define the scope and the structure of the session and also provide learning experiences that the teacher should use. Beside the common textbooks, there are also guide books for teachers, weekly and termly

time tables, a topics manual time table, inspection by headteachers, and curriculum scrutiny by specialists from the Ministry of Education.

The assessment of both short-term and medium-term objectives is, however, school-based. The official policy suggests a variety of methods for assessing students' learning: for example, oral, practical and written tests, using observation cards and portfolios. Tests have been excluded in the first stage of primary education. Assessment should be continuous and entirely teacher-based. It is, however, used for summative purposes. Schools use it to promote students from grade to grade and the Ministry uses it to monitor patterns of students' promotion.

In the second stage of primary schooling, tests become the dominant method of assessment. The present context of assessment allows schools to make and implement their own tests. Hence, there is considerable variation from one school to another over matters of content and marking. Centralised management of schools' tests is only evident in three policy respects: tests should be conducted at two points of time (mid-term and end of term), 70 percent of the total students' score should be derived from testing, and a summative report should be submitted to the Ministry of Education at the end of each term.

The mid-term test is the full responsibility of teachers. They set questions, organize the test situation, mark students' papers and report results. The end-of-term test is the responsibility of both schools and teachers. In the end of term test, a set of questions is proposed by teachers and sent to the headteacher. The final set of

questions is selected by the headteacher in consultation with teachers. They are then typed and retained in the administration office. This procedure is designed to produce standardized tests which allow for comparison between classes and students in the same school but not between schools. The end of term test is also cumulative, in that it covers the whole syllabus of the course, while the mid-term test is related to content taken in the first half of the term. Thus, the end of term test provides a comprehensive picture of students' performance during the term. Decisions to promote students are also more sensitive to the results of the end-of-term test. They contribute 50% of the marks, with the mid-term test contributing 20 per cent, and classroom work, 30%.

No measures are taken yet by the Ministry of Education to ensure comparability of tests results between schools. Inspection of the test items might be conducted to ensure the coverage of content of the curriculum and give guidance on the construction of test items. These procedures are, however, not uniform. For example, there are at least 80 curriculum specialists who are responsible for about 3,000 primary teachers in different subjects. Criteria for inspection reports may be unified but there is no assurance that these criteria are followed by all curricular specialists. Neither is there any unified procedure for inspection over the assessment procedures in schools, such as test content, test construction, marking and reporting.

Grades are the major tools used by the schools and the Ministry to judge and evaluate the effectiveness of teaching. The letter grades are used in year 1 to 3 and

consist of six categories, excellent, very good, good, fair, accepted and poor (fail). Percentages are used for year 4 to 6 and correspond with the letter grades as follows: 90 - 100 (excellent), 80 - 89 (very good), 70 - 79 (good), 60 - 69 (fair), 50 - 59 (accepted) and less than 50 (poor or fail). There is no obvious or ultimate reason for using the two ways of reporting. Grades may be seen as less sensitive to ranking, a procedure which began to be discouraged in early years of schooling. In reality, teachers in year 1 to 3 use the percentile scores in marking but convert these scores into letters in the students' reports.

Conclusion:

This chapter has explored assessment in Bahrain. It suggests that, in fact, schooling has been dominated by bureaucratic examinations rather than diagnostic or formative assessment. Before 1960, examinations were used as a screening device. Because of the limited places available and the increased demand for education, students were selected at an early age of schooling. Examinations also functioned to unify and equalise the social culture. Central examination, central grading and marking contributed towards unifying the content of education and the methods of instruction.

In the 1960s, the concept of equal opportunities and the policy of providing a minimum level of education brought a variety of changes into education. Thus, formal education had to be expanded and reach every child on the island. A policy dilemma arose. Should investment be made to expand the quantity of education or to raise the quality of education? With the high level of illiteracy and the high

demand for both primary and secondary education, the choice was made to expand the quantity of schooling. Providing places for the increasing number of children in both primary and secondary education was the main concern. Educators concentrated on the task of re-organization. This meant appointing qualified teachers, expanding secondary education, establishing post-secondary institutions, and finding a suitable design for new schools. Within this situation, ensuring high productivity for the educational system was important. Examination results were used, in this sense, as measures of the system's productivity. Students were treated as statistics rather than as learners.

In the early 1970s, when the organization of primary education had been completed and students' growth started to settle down, the issue of quality began to receive attention. Attention shifted from inputs and outputs to school processes. Several innovations were introduced in the organisation of schools to the enhancement of classroom life.

In turn, primary education was released from the pressure of base-level economic development. It became an autonomous institution in which national aims and national curriculum began to be harmonized. The class-teacher system and the structural changes associated with it in the 1980s represented the first wave of attention to the quality of teaching and learning. Changes in assessment, however, were minimal, which hampered changes towards a more child-centred view of classroom life. Only in 1994 was assessment recognised as central to the whole philosophy of education in Bahrain. Such changes are currently under

consideration. Thus, classroom life must include different and progressive practices in assessment. These are challenged, however, by - the subject of next chapter - inherited practices which obstruct school innovation and reform.

5

Classroom Life

This chapter moves from the theoretical conception of assessment and from the policies of Bahrain's education system to the reality of classroom practices. The analysis is mainly derived from small-scale observation in two primary schools. The chapter begins with an account of teachers and teaching in Bahrain primary schools. The second part moves closer to the daily classroom practices. It explores teachers' style, students' behaviour and teacher-pupil relationships.

In studying teachers' practices, two significant landmarks are first worth considering. These are: variation in the provision of teachers in schools and the common context of teaching and classrooms.

Variation in provision of teachers:

This aspect can be understood within the history of teachers training programmes. Before 1940 most teachers in public schools came from other Arab countries mainly Syria and Egypt. Public schools, however, recruited Bahrain teachers when they obtained a primary education qualification. This policy was adopted because of the shortage of qualified teachers and also because education, at that time, was mainly concerned with providing basic skills in reading, writing and arithmetic necessary for jobs in the labour market. In the 1940s, changes took place. Among these changes, public schools begun to provide an evening teacher-training programme. This lasted until 1954 when the first full-time teacher education programme was initiated at secondary school.

The need for better qualified teachers was reviewed in the 1960s. The Teacher Training College (TTC), established in 1966/67 for males and in 1967/68 for females, provided two-year full-time courses in teacher education. Teachers graduating from this college initially worked in primary schools. Some teachers were subsequently moved to teach in intermediate schools.

The college was replaced in 1978 by the University College of Science, Arts & Education (UCSAE). UCSAE provided four-year full-time teacher education courses. In 1986, all the state colleges were integrated with the University of Bahrain. Accordingly, teacher education is now undertaken in the College of Education, which provides four-year full-time courses. Meanwhile, the Ministry of Education has also created a second path of inservice training for teachers who had

earlier joined the profession with less than a BA qualification. They receive one of two modes of training. Special short-term courses organized within the Ministry of Education and/or long-term training programmes leading to Diploma qualifications. Bahraini students with BA's who graduated from universities in other countries (e.g. Kuwait, U.A.E and Egypt) are also recruited as teachers. In 1983, however, The Ministry of Education stipulated a minimum qualification for teachers' recruitment. New teachers should have an educational B.A qualification. Those who have a non-education B.A and wish to enter teaching, are required to take a one year, full-time teacher-training course in the College of Education.

Thus, the present population of teachers in Bahrain comprises an assortment of qualified/unqualified and secondary/diploma/B.A teachers. Table 5.1 summarizes the situation in 1994/95.

Qualification	Women (percent)	Men (percent)	Total
Secondary level and less	4.7	10.8	7.3
Secondary with educational training.	3.2	8.0	5.2
Teacher Training College.	18.5	12.0	15.8
Educational B.A.	57.9	45.0	52.5
Non-education B.A.	5.5	14.1	9.1
Higher diploma.	9.2	8.7	9.0
Master, Ph. D and others.	1.0	1.4	1.1
Total	**100**	**100**	**100**

Table. 5.1 Teachers by qualifications in public schools in Bahrain in 1994/95.

The shortage of qualified men teachers and the shortage of female teachers in some academic areas indicates another variation. Teaching is not an attractive occupation for men. The shortage of male teachers has been covered in two ways. Teachers

are recruited from Arab countries, mainly Egypt, and female teachers are recruited to work in primary boys' schools (age 6 to 10 or 11). In 1994/95, there were 23 boys' schools which were entirely staffed by 566 female teachers.

There is also a shortage of female teachers in some areas such as English language, fine arts, music, sciences and physical education. This shortage is also met by recruiting teachers from other Arab countries.

Sex	Nationality	Total	Primary teachers	primary teachers (Percent)
Female teachers (Girls' schools)	Bahraini	3010	1272	
	Non-Bahraini	328	118	
Female teachers (Boys' schools)	Bahraini	496	496	
	Non-Bahraini	70	70	
Total (Girls' and boys' schools)	Bahraini	3506	1768	90.4
	Non-Bahraini	398	188	9.6
	Total Female	**3904**	**1956**	**100**
Male teachers	Bahraini	1883	644	91.6
	Non-Bahraini	911	59	8.4
	Total Male	**2793**	**703**	**100**

Table 5.2. Total number of teachers by sex and nationality in 1994/95.

As shown in table 5.2 the total number of teachers in public primary schools comprises 1956 females and 703 males. This number does not include primary teachers (300) who work in 21 mixed-stage, primary-intermediate, schools.

The pattern of job mobility, age and qualification for women and men marks

another variation. The average number of years spent by men is normally longer than women. This is so because some women leave after marrying while others resign before the retirement age. Thus, generally speaking, girls' schools experience greater staff mobility than boys' schools. This undoubtedly affects the age, years of experiences and qualification patterns of male and female teachers. In 1994/95, 16% of men teachers had less than 5 years experience, compared to 24% among women. On the other hand, 10% of men teachers have over 25 years teaching experience while this proportion accounts for only 3 percent among women.

In a study of teacher' training in Bahrain, Wehbeh (1988) pointed to these variations in Bahrain teaching provision and highlighted their influence on educational practices and, particularly, on assessment practices. He pointed out that at least 50 percent of teachers in Bahrain schools derive their pedagogy from 'unspecified, different or even unknown educational backgrounds' (p. 4). He went on to explore how these variations led schools to share the same view of school effectiveness. The 'successfulness of the school' (pp. 3, 4) was indicated by the number of students who succeeded and passed to the next grade. He explained that, as a result, unqualified teachers, Bahraini or non Bahraini, tried to pass as many as possible students in their classes, to indicates their effectiveness and to confirm that a lack of formal educational qualifications was not a barrier to good teaching.

This view of school effectiveness is valid to an extent but, at the same time, it ignores the role of the Ministry of Education. A high through-put of successful students is also accepted by the Ministry of Education as an indication of the

effectiveness of schools. The Ministry, therefore, tacitly encouraged schools to raise the number of students promoted. Whatever their causes, such practices have been accepted in Bahraini education for many years. They are part of the school culture.

The organizational context:

Despite these variations in teacher provision, the Ministry of Education aims to compensate for them with a common organizational context. Teachers in Bahrain work within a bureaucratic and hierarchical structure of authority. They have low autonomy. All teachers are state employees and have tenure for life. Their status, recruitment, redeployment, training and their work loads are all centrally determined. Teachers are not free to choose their schools. Those who wish to change their schools have to do this through the Ministry of Education. A system of inspections and guidance of individual teachers is carried on by the Directorate of Curriculum. Another type of inspection is carried through by headteachers.

Teacher obligation to the syllabi and textbooks is extreme. It is expected that all primary students at a certain grade, at a particular time of the year are exposed to the same knowledge and skills. Individual differences are not considered. Common textbooks produce common whole-class teaching. Teachers use a whole-class instructional strategy as their main teaching style. The blackboard is central in teaching and learning. Each teacher maintains uniform patterns in his/her class which remain unchanged. This applies to their teaching style, their interaction with students and their assessment practices. Classroom organisation remains fixed and students rarely change their places. Their movement and talk are restrained. To use

Bennett's (1976) typology of teaching style, progressive and traditional, one can conclude that traditional styles of teaching dominate current practices in Bahrain schools. This is not to claim that practices are identical. There is evidence that some class teachers are moving away from traditional styles of teaching (Al-Ameen, 1993). Their practices may be a blend of both styles. But this anecdotal evidence still needs to be confirmed by further research studies.

Although the official assessment policy recommends a comprehensive assessment approach, tests dominate assessment practices in primary schools. Assessment procedures are left to teachers. They devise their own types of tests, items and correction criteria. In the orientation phase of the field work, six schools were asked to indicate, on a specially organised calender, any test which they conducted for students in one term. Three schools cooperated and provided the following information:

School	Isa Town		Al-Quds			Sitra		
Subject	Year 3	Year 6	Year 3	Year 4	Year 6	Year 3	Year 4	Year 6
Arabic	14	18	14	10	8	5	4	n.a
Mathematics	4	4	15	4	11	4	11	4
Social studies	7	7	3	5	3	1	5	2
Sciences	3	2	5	4	4	2	5	5
Total per class	**28**	**31**	**37**	**22**	**26**	**12**	**25**	**9**

Table 5.3. Number of paper and pencil tests conducted in three primary schools in one term in 1993/94.

The table shows that testing is extensively used in primary school. This use is also

evident in year three where, officially, teachers are expected to use continuous assessment rather than special test session. Testing procedures are the responsibility of teachers. According to the data in table 5.3, class teachers in Sitra school (in year three) have to correct at least 360 papers in that term. In Al-Quds school, this number increases to 1110 papers. From the scrutiny of tests in several schools, it appears that testing is based mainly on syllabuses and the content of the textbooks. Objective tests are preferred but a mix of both essay and multiple choice items are used.

Small-Scale Observation in Two Primary Schools

The analysis in this part is derived from classroom observation in four primary classrooms (year-three and year-six) in two schools (girls and boys). Observation was carried out for 17 weeks from 25th September 1994 and ended on 25th January 1995. Six teachers in these classes were observed for at least 23, 50-minute sessions, a total of about 19 hours for each teacher. Table 5.4 shows the distribution of these sessions by year group and sex. A full account of how observation data were collected is provided in the first chapter.

Sex	Year three	Year six	
	Class teacher	Arabic	Mathematics
Girls	24	23	26
Boys	25	25	23

Table. 5.4. Total sessions observed in four primary classrooms.

The two schools:

The two schools observed are situated in two towns. The girls' school opened in 1980 and the boys' school in 1985. Both schools' buildings are of modern design. The school buildings, like most public schools, are on two floors, organized in a square format with open spaces in one or two corners. One block is occupied by the staff (the headteacher, assistant headteacher, secretary, social worker, and librarian). Teachers' rooms are situated in the classrooms blocks. The schools also have a library for students and a laboratory room. There is also a multipurpose block which is used for sports, school celebrations and meetings with parents. In 1994, there were 6 staff (headteacher, assistant headteacher, secretary, typewriter, social worker and librarian), 25 teachers and 574 pupils in the boys' school; and 7 staff (headteacher, assistant headteacher, secretary, typewriter, social worker, librarian and educational resources specialist), 37 teachers and 707 students in the girls' school.

The school year is organized in two terms. It starts on 25th of September and ends on 15th of June with two weeks holiday starting on the 1st February. The teachers and the staff, however, start 10 days earlier and stay two weeks more with 10 days holidays between the two terms. Students go to school 5 days a week from

Saturday to Wednesday and have 21 (grade 1 to 3) to 25 (grade 4 to 6) hours in their working week. The school day starts at 7.10 a.m and finishes at around 12.15 p.m for younger students (year 1 to 3) and 1.10 p.m for older ones. Ten minutes is allowed for the morning assembly, and there is a 30-minute break at 10 a.m.

Both schools follow the class and integrated-subjects teacher system. The boys' school, however, altered the main principle in the class-teacher system. The entire class-teacher policy where students stay together with their teacher for three years, is not applied in this school. Students of the same class are distributed at the beginning of every year, when they have new class-teacher. Mr B, the class teacher, explained this policy in terms of expanding students' friendships and promoting their learning if some students do not benefit from their teacher. Thus, the school is still sticking to the old system.

Mrs A (class-teacher):

The classroom observed is year three. The classroom is quite large (about 60 m^2) and is situated on the ground floor near the administration block. Inside the classroom, in one corner there are three set of divided shelves, an open-front cupboard, two cupboards with drawers and a fixed sink. Two wall-mounted notice/pin-boards are hung on the wall in two different areas. The teacher's desk and a table with a small divided shelf occupy a corner near the mounted board. A book & magazine stand with another cupboard occupies the third corner to form the classroom library. Pupils' desks and chairs were arranged in six groups. This arrangement lasted for one month (October). In the remainder of the term

(November to January), the class seating was organised in a horseshoe shape. Decorations also hang on the wall. The class is full of teacher's drawings which mainly relate to the syllabi. Students' work and handiwork are also displayed in the class.

There are 30 students in the class. All students are eight years old. According to school records, most students are from middle class families. Table 5.5 shows students' end of year scores in Arabic and mathematics in 1993/94 (year-two).

Subjects	**Arabic**	**Mathematics**
Scores		
91 - 100	28	28
81 - 90	1	-
71 - 80	-	1
61 - 70	-	1
51 - 60	1	-
50	-	-
less	-	-
Total number of student	**30**	**30**

Table 5.5. Girls' end of year-two scores in Arabic and mathematics.

The above data in table 5.5 show that nearly all students have high scores in both subjects in the previous year. A few students demonstrate more competence than others, for example, fluent reading. The observation notes show that most students appear at the same level of competence. A few students, however, need some help and require the teacher's attention in learning. These were pointed out to me by Mrs A. Mrs A also differentiates between these students and another four students who are slow in working.

Students are responsible for cleaning and tidying the classroom. Each girl has her tissue box, a glass for drinking, a towel and a tooth brush. Students seem to have good relationships with each other. They seemed to cooperate together and act as one group. Disruptive behaviour never occurred.

Mrs A, the class teacher took over the responsibility of the class in year-two when their class-teacher moved to another school. She has class-teacher qualification and had three years teaching experience. She also taught subnormal children for three months. Mrs A actually knows a great deal about students' home background and keeps contact with parents. Parents can also contact her at home if there is any need.

> My contact with parents (she told me) is consistent, as I need to know their home background.

Mrs A established a personal relationship with students and tries to keep an informal atmosphere in her lessons. She approaches students when talking with them, keeps eye-contact and is insistent in her requests. She seems sometimes to refer to her relationship with students as a technique to keep control and order.

> I will be upset. You make me feel ashamed of you.
> I am not talking with you if you are making such noise.

Sometimes, she uses it to praise and encourage:

> There you are! My lovely children. How wonderful you are today.
> If you behave well today I will show you something interesting.

Children are attracted to her and express their feeling in a variety of ways (for

example: bring flowers from their garden, write some emotional notes and bring her some apparatus which she uses frequently in the class). Using Hargreaves' (1975) definition of the teacher's role in the classroom (see chapter 3), Mrs A, plays the two main roles, disciplinarian and instructor. As a disciplinarian, Mrs A seems to use an individualized strategy. She is very firm and combines her comments with warnings. She keeps her eyes on students and never disregards any misbehaviour. Students have become accustomed to her rules and she has few disruptions. She is able to leave them on their own. During her absence, students stay at their own desk and only talk with the nearest partner in low voices. Nevertheless, in learning situations, students seem anxious to please her and to avoid incorrect answers.

As an instructor, Mrs A uses a whole-class teaching strategy and no allowance is made for individual differences. The written activities are always part of each lesson. Students have their own writing and practising books for each subject. Homework is always given but students can complete it in the class if they have time. She insists that each student should try to accomplish tasks with perfection. The way she achieves this is by extensive exercises and homework, constant marking and correction, continuous diagnosis of any weakness, however small, and finally, incentives in the form of rewards and grades.

Mrs A encourages competition, so that children are rewarded stars for good work and tidiness. Her interest is in making children happy and she motivates their curiosity in learning. She sometimes switches to fun as motivation to students. Mrs

A emphasizes good performance, neat handwriting, cleanliness, tidiness and doing homework. Memorizing is emphasized and students' answers should follow what is written in the textbooks. Hand raising is the expected way of participating in answering teacher questions. Mrs A can, accordingly, delineate those who don't raise their hands as requiring further attention.

> I want those who do not raise their hands, this is only a revision lesson.

Mrs A conducts regular tests with her class which follow the official pattern for grade 4 to 6. That is, monthly, mid-term and end of term tests. She, however, called these 'assessments' rather than 'tests'. At the test time, seats and desks are rearranged in 5 separated rows. She usually reads questions to students and may spend some time in explaining questions. She then leaves students to work on their own while she is doing other work. Mrs A corrects papers as soon as students finish. Each student will be called to check their answer if there are any errors. Those who complete model answers, are announced to the whole class and are given applause.

Mr B's class (class teacher):

The classroom is situated in the first floor. Its area is about the same size as in girls school. It is also provided with similar furniture. The layout of the classroom follows the class-teacher system pattern. There are three filing cupboards for students apparatus, organized in one corner of the class. A small library occupies the opposite corner while the sink cupboard is located along the wall between the students' filing cupboards and the library. Two other cupboards, teacher's desk and

chair take the third corner. Students' desks are organised in 5 groups, three at the front and two at the back. The classroom is less decorated than Mrs A's class.

There are 29 students in the class. All students are at age eight except Yosif who is ten years old. Yosif had repeated two years in the previous years. This student, with another three, attend special remedial lessons. School reports show that most students are from middle class families, with the remainder from working class families. Table 5.6 gives the students' end of year scores in Arabic and mathematics for grade two (1993/94).

Subject	Arabic	Mathematics
Marks		
91- 100	13	16
81- 90	6	3
71- 80	6	9
61 -70	2	1
51 -60	1	-
50	-	-
less	1	-
Total number of students	**29**	**29**

Table 5.6 Boys' end of year-two scores in Arabic and mathematics

Students developed contingency friendships, but unlike the girls' class, peer groups are more evident in this class. Incidents of disruption occur from time to time. Occasional complaints about others' infringements are also given to the teacher. Students seem not to be responsible for cleaning the classroom as in the girls' class. The classroom remains uncleaned all the term. Students' shelves and files are covered with dust. They do not have the same awareness about health and cleanliness. They share one glass to drink from the sink and one towel to wipe

their hands.

Mr B, the class teacher has a BA qualification and has taught for 18 years as a primary teacher. As a result of the school alteration in class-teacher policy, he teaches the class for the first time this year. Mr B does not have enough information about students' home background. His initial encounter with them, therefore, revolved around informing students about rules and eliciting some information from the students. He asked students to give their full names and give information about themselves such as, their fathers' occupation, their previous teacher, and their marks from the previous year.

Mr B does not have so many disciplinary rules as in Mrs A's class. He does not take much notice of students' movement and talk in the class while he is teaching. His relationship with students is informal. Students don't hesitate to break the lesson progress, contact him or even interrupt him while teaching. Fidgeting behaviour also prevailed among a number of students. Some students may put their feet on the desk, lean their heads on the desk or fiddle with other things (their apparatus, for example). Mr B appears less able to control this behaviour. 'Keep quiet' and 'Those in the back' are frequently repeated phrases. Some students seem sleepy in the class. In each lesson, several students proceed to the sink to wash their face. No comment is made by teacher in regard to this matter.

Mr B adopts a whole-class teaching strategy. No allowances are made for individual differences. Mr B stays in the front of the class when teaching He tries

to generate students' interests when exposing a new topic and relates topics to students' individual experience. Although students seem to enjoy his sessions, their interaction with the teacher is disorganized. Several students talk at the same time. Every minute Mr B had to repeat to students to sit back on their desk and/or to listen to what he and the others were saying. He interacts with the class as a whole when teaching or individually at work time. Although students are seated in groups, children work on their own. Mr B sometimes arranges group work, but it appeared that one student did the work while the others watched.

Mr B also finds difficulties in controlling the class when students have to work on their own. Before class activities, Mr B gives instructions on how students should deal with the task, without sustaining the students' attention. Many students do not appear to pay attention to what he says. This seems to create a general confusion among children when they start working and a degree of anger and accusations from the teacher. He goes around the class. He checks, marks and corrects students' work. But much of his time is also given to blaming and criticizing.

Mr B does not conduct tests with the class. He depends on the day-to-day interaction with students. He encourages his students to evaluate each other, and to engage with him to discover what is wrong. But he puts less emphasis than Mrs A on providing individualized diagnosis and remedy. Grades and rewards also seem to be absent.

Mrs C and Mrs D's class (integrated-subject teachers):

The class observed is year six. The classroom is smaller than those of class-teachers (about 52 m^2). In addition to students' and teacher's seats and desks, the classroom is furnished with three filing cupboards divided into shelves and pigeon holes, one cupboard for students' files and textbooks, a book and magazine stand and a wall-mounted notice board. During the term, students are encouraged to bring paintings and illustrations which are displayed on the wall.

There are 32 students in the class. All students are at age 11. Students are paired off in four rows, A, B, C and D (4 ranks each). School records show that students are mainly from middle and upper-middle class families, with three to four students from working class homes. Students display politeness, shyness and respect to adults. Discipline is maintained when the teacher is not in the class. Some students undertake certain roles. One is responsible for maintaining order and control when the teacher is not present, two are responsible for students' files and textbooks, one is responsible for the classroom's library and one prepares the blackboard for the next session. According to school's records, students' end of year scores in Arabic and mathematics in grade five (1993/94) are shown in table 5.7.

Subjects	Arabic	Mathematics
Marks' distribution		
91 -100	13	11
81 - 90	6	6
71 - 80	6	5
61 - 70	3	2
51 - 60	2	4
50 (the passing mark)	-	-
Less than 50	-	2
Unknown	2	2
Total number of students	**32**	**32**

Table 5.7. Girls' end of year-five scores in Arabic and Mathematics in 1993/94.

The classroom observation shows no discrepancies with the above data. Many students are competent in both subjects and their motivation to learn is quite evident. They are active participants in the class and, to a large extent, they retain responsibility for their learning. Some students appear more confident than others and show competent participation in the class. On the other hand, some students, particularly those of lower achievement level, look indecisive when answering teacher's questions.

Mrs C teaches the class 11 sessions per week (Arabic, religion and social studies). Mrs D teaches them mathematics and sciences for 8 sessions per week.

Mrs C has 12 years teaching experience. She has the Teachers' Training College qualification (see part one of this chapter). She spent six years as a teacher in the middle (intermediate) school and another six years in the primary school. Mrs C teaches two classes in the school for 22 sessions per week (11 sessions per class).

It seems that Mrs C is a popular teacher. It is common to see many students rushing after her, seeking her advice or consulting her after the lesson has finished. She pays a lots of respect to her students. She called their names in full (e.g Fatima Hassan) as she thinks this enhances the student's self-image. She approaches her class with a regular smile which rarely disappeared. Blaming students, criticising them or giving them critical notices is not within her pattern. Mrs C seems proud of her students. She always admires them and continually pointed out to me that they are still children and still need the care and attention of adults.

Mrs C's sessions are dominated by oral discussion. Written activities where students work on their own is entirely absent in her class. Individual work is done at home. Mrs C gives regular homework, but she rarely checks whether students had done it or not. She seemed to trust her students and assumes that they have done their homework. Nevertheless, she regularly collects students' books for checking and correcting in her own time. Mrs C values neatness and well organized books. She draws her students' attention to good hand-writing, double line spacing and leave the next one, so that she can comment and correct their work and not overuse the correcting fluid.

Mrs. C always stays in the front of the class and rarely moves around the class. She adopts a whole class teaching strategy and consolidates this with reciprocating individualized discourse. Thus, her teaching appears as conversation rather than lecture. She executes a short discourse with one student. During this time all other

girls listen to this dialogue until teacher points to another one. In all cases, she first asks questions to provoke students' thought and stimulate participation from them. Raising hands is not a pre-condition for an opportunity to answer questions. She calls on non-volunteers to participate in the class

> Who has not answered any question today?
> No. I want someone else, those whose voices we did not hear today.

Mrs C seems to be keen to give students an opportunity to participate in her lessons. In any lesson most students have a chance to interact with her. The observation notes in 2nd and 9th of October, for example, recorded 20 and 27 students respectively who had been given this opportunity. Mrs C allows time for students to answer her questions. She also offers some support for those who seem less confident about their answer. She may rephrase and simplify the question. She also never puts off, criticises or blames any student who is not able to answer her question correctly, or who forget to do homework.

Mrs C also involves pupils in correcting answers. In many cases an oral answer is combined by writing answers on the board. Students are also encouraged to work with dictionaries. In several instances, students compete to find out meanings of words from their own lexicons.

In contrast to Mrs C, Mrs D has a different interaction pattern with the class. Mrs.D obtained the Teachers' Training College qualification (see the first part in this chapter). She has 15 years' teaching experience in Mathematics and sciences. Mrs D keeps a firm and formal relationship with the class. She firmly comments

on any disapproved behaviour. She, however, encourages students and instantly praises their work. 'Well done', 'Very good' and 'Excellent' are frequently used. In contrast, she addresses any evidence of laziness or inattentiveness with a sharp comment and warning.

A whole-class teaching strategy is adopted. Her lessons normally consist of two main parts: lecturing and practising. These two parts, teaching and practising, are interweaved in each lesson. They cannot easily be differentiated. Sustained attention is emphasized when teaching, while practising is the time for errors to be shown, discussed and corrected. Each lesson, Mrs C gives some students the chance to practice on the board while others watch and check their answer against what is shown on the blackboard. Practising on the board is given to about 7 students each lesson. In this sense, students share some responsibility with the teacher as well as engaging with their own learning. When students who work on the board are competent, Mrs D has more time to go around the class, check over other students' work and to see how the low achievers are working. On the other hand, if the students are incompetent, then the class could see what kind of errors the students, including themselves, commonly make. Students seem always attentive to problems and work out how they could be remedied. A few students in the class (three to six) need teacher's guidance and help. Deficiencies appear in recognizing place value, borrowing rules and sometimes not memorizing the multiplication tables. Homework is always given. Mrs C assumes that students have done their homework but she sometimes make inspections. Mrs D reiterates insistently how students should display their work. She gives them instructions as

to how they should organise their exercise books.

Mr E and Mr F's class (integrated-subject teachers):

The classroom is very small as it was originally built to serve the old system. The classroom is provided with students' seats and desks arranged in four rows, a teacher's seat and table in the front, one cabinet for the class and teacher's work in one corner, and a sink in the other corner. By contrast with the girls' class, there is no library in the classroom. In addition, students never bring any work of their own to display in the class. The classroom looks plain.

There are 34 students in the class. Students are seated in four rows, A, B (5 pairs), C (4 pairs) and D (3 pairs). The active and energetic ones are sitting in the front particularly in C and D. The age range in the class is from 10 to 14. There are 24 at the age of eleven (the regular year age), 3 at age ten who started school at an earlier age, 5 at age twelve who had repeated one year in their school career, one at age thirteen and another at age fourteen. This range of ages seemed to have an influential effect on students' relationships. There appear to be several peer groups in the same class. These groups have different definitions of the classroom situation. One group might be interested in accomplishing some learning tasks, another might be interested in making amusement or distracting behaviour and another might prefer to withdraw from both situations. Some older students also tried to dictate orders to the younger ones while the younger ones, on the other hand, reject such orders.

The class is generally noisy when the teacher is absent. Some students take advantage of the five minutes between sessions to fight, kick each other, move around or do homework. Some students sit quietly, but movement and noise in the classroom is the dominant feature. Two students are responsible for the classroom discipline. They register those students who make too much noise or fight with other students. The list is given to the teacher as soon as he comes to the class, but not much is done by teacher in this case. Apart from blaming or criticizing students, the list would or would not go to the school administration for further action. According to school reports, students' end of year scores in Arabic and Mathematics in the previous grade (1993/94) are shown in table 5.8.

Subject	**Arabic**	**Mathematics**
Marks' distribution		
91 - 100	10	9
81 - 90	10	4
71 - 80	4	8
61 - 70	2	2
51 - 60	5	6
50 (the passing mark)	2	4
Unknown	1	1
Total number of students	**34**	**34**

Table 5.8. Boys' end of year five scores in Arabic and Mathematics in 1993/94 (boys)

These figures run counter to the achievement I had observed in the classroom. Incidents show that students are deficient in many basic skills in both subjects (Arabic and mathematics). Three to four students show an acceptable performance. While some try to catch up, the rest are far behind the required academic level. The analysis will show later how teachers in both subjects cope with this situation.

Students of higher performance are from upper-middle and middle-class families. Lower performance students come from middle- and working-class families. The observation also shows that some students in the class, particularly those of lower performance, are passive and restless. They do not share or participate in answering teachers' questions and give no evidence of motivation to learn. At practice time, for example in mathematics sessions, some students are distracted from tasks on hand. They don't show any interest in finishing tasks and normally involve themselves in other thing (e.g. chatting with other students, asking permission to go outside the classroom).

Mr E teaches the class 11 sessions per week (Arabic, religion and social studies). Mr F teaches them 8 sessions per week (mathematics and sciences).

Mr. E has 21 years teaching experience as an Arabic teacher in primary school. With the application of the new teaching approach for the primary grade 4 to 6, he is obligated to teach the three subjects. He primarily had obtained the secondary school certificate and later he had been qualified through in-service training. He obtained two diplomas; the educational qualifying diploma and the primary education diploma.

In his first encounter with students Mr E started to dictate rules and prescription to students. He makes no attempt to know students personally, for example, their names and their previous class. Later in the term, he still did not know some students' names. If he wants to obtain an answer from any student he points to

them or calls him 'you'. Mr E holds a stick. In his first lesson with students, he pointed to the stick as a tool for other things than punishment, such as teaching at the blackboard, but indicated that it might be used as well for any nonsense in the classroom. In my presence, however, he used it once to strike two students on their hands. Mr E also established a firm rule against students' requests to go outside the classroom. It seems that this technique, going outside, is used widely by students to escape from lessons they do not enjoy, as happened with Mr F. During the term, therefore, no student requested such a thing from Mr E.

Mr E's relationship with students is passive and lacking in emotion. Many students in his class are restless and seem reluctant to participate. He tries to stimulate students' participation. Many students have no motivation to participate in answering his questions unless he asks them to answer. He tends to ask those who are not paying attention or seem confused. Some students seem to follow a strategy of escaping from such position by raising their hands. But Mr E is also aware of such a strategy. Each lesson has a few incidents of deviance and inattention and Mr E pursues these cases instantly. His reaction to these cases, however, is inefficient. Apart from attracting their attention or blaming them, no radical solution is being taken to alleviate or correct this situation.

Mr E adopts a whole-class teaching strategy and uses a questioning strategy to stimulate students' participation. His questions revolve around the abstract rules of Arabic grammar. He insists that grammar should be recited by students. His questioning strategy yields many opportunities for teacher-student interaction.

These chances, however, last only for few seconds until the correct answer is obtained.

As students appear deficient in many basic skills, Mr E is educationally constrained. He tries to promote students' learning but with no apparent sense of motivation and progress among students. A typical example of such constraint occurred when Mr E organized textbook exercises for the class. He starts by asking one student to read the question from the text book. Reading competence for a number of students is poor and thus reading for them is slow and full of errors. Mr E stops the student at every error, and takes extra time to explain phonetic and grammatical rules. His way seems to cause confusion because he keeps asking a word which the student can hardly pronounce. The whole class at that time are not actually concentrating on the content and the meaning of the question as much as watching the interaction between teacher and student. When reading the question is over, another similar situation starts as soon as the teacher asks the whole class about the answer to the question. It appears then that students are not actually concentrating on the question as much as watching what happens between Mr E and the student. Thus, Mr E's class has several features: passive learning milieu, students deficient in basics, inactive teaching and a dissatisfied teacher. Mr E's complaint about pupils is not so much with students' behaviour as with their learning attitude and progress.

Mr F has different encounters with the same students. He had 6 years teaching experience in Mathematics and sciences in primary school. He has a university

qualification in science education. His teaching load is 21 sessions per week. He teaches mathematics to the three (grade six) classes in the school.

Mr F keeps an informal relationship with his students. He made no firm disciplinary rules. Students seem happy and relaxed in his class. He never shows any annoyance or punishes any student. He accepts a few jokes but tries to limit them. He sometimes expresses this informal relationship by sitting next to students on their desks. Mr F allows freedom of movement in his sessions. He also allows students to go outside the classroom. Students seem to take advantage of his lenience. They frequently ask to go outside the classroom. In each session, between 5 to 6 students are allowed to go outside the classroom. He also allowed group conversation at the practising time.

Mr F's sessions follow the same pattern; checking student's homework (2-3 minutes), explaining and solving some exercises on the board (10 - 15 minutes), checking and correcting students' work (25 - 30 minutes) and giving them homework (2 - 3 minutes). He uses a questioning strategy in teaching. Some students offered themselves for answering by raising their hands, standing, voicing their desire or sometimes answering without seeking the permission of the teacher. The teacher sometimes gives the opportunity to the other students. If they do not know the answer, they either keep quiet or admit their ignorance with a smile (and probably jokes from the class). The whole class seem to have fun at this time. Nobody seems to be offended by such behaviour as every one appears to make errors. When a student needs time to think, he would ask other students or gives

the answer himself. At the practice time, he gives students some tasks to work upon individually. During the practice time Mr F instantly asks them to hurry up. A competitive approach is quite obvious. Mr F keeps announcing those who have finished more tasks and praising those who try to catch up. The most striking event in this part is watching students working individually, but also moving with their exercise books around the classroom. For Mr F, it seems that his attention is not to deal with these trivial aspects of the classroom as much as to correct and check as many students' work as possible.

Conclusion:

This chapter describes the overall pattern of teachers' styles, their relationship with students and students' behaviour. Within this pattern, there are notable gender differences among teachers and among students. Despite differences in teachers' style, it seems that female teachers, unlike male teachers, have succeeded in establishing a working consensus with students. It is not the interest of this study, however, to carry out an analysis of how and why this difference occurs. But it might be important to say that such differences are becoming recognised among educators in Bahrain. It is acknowledged, for example, that girls' schools are better organised than boys' schools. Underlying such a fact is the social culture in Arab societies, where girls and boys are brought up differently.

In boys' schools, student behaviour seems to be beyond the responsibility of teachers. This situation is becoming more obvious with older students (year six). But this situation may also be traced back to their earlier school life. There, as

appears in the observations, is a difference between Mrs A and Mr B in the way they accomplish their role as class-teacher. Mrs A places emphasis on students' behaviour as much as on their academic achievement. The two aspects of students' learning (academic and behaviour) interweave in her lessons. On the other hand, Mr B seems to play his role mainly as instructor. Behaviour is only commented upon if it interferes with instruction. Other behaviour, for example, cleanliness and attitude towards the others is disregarded as beyond the teacher's concern.

In year six, teacher-pupil interaction for Mrs C and Mrs D is highly task-related. The role of the teacher as an instructor is obvious while their role as disciplinarian seems to be dwindling. There is still, however, evidence of the teacher as disciplinarian, which appears in teacher summative assessment records. This is so probably because teachers did not need to adopt such roles with students since they hold the same definition of the situation. Students and teachers have common knowledge about the classroom situations. They act as a group. Nevertheless, it appears that teachers did not completely abdicate the disciplinarian role. They still interfered in such situations, as, for example, when Mrs D blamed Auhood for rudely answering another student.

Other important differences, which also related to gender differences in performance and motivation, are examined in the next two chapters.

6

Teachers' Judgement and Expectation

Chapter three reviewed assessment in empirical and conceptual terms. It suggested that teacher assessment is associated with teachers' judgement and expectation. Judgement, as identified in chapter three, is related to teachers' categorization of students in relation to social criteria. At the same time, it involves anticipation of students' future progress, or in other words teacher expectation. These aspects of teacher assessment are the concern of this chapter.

Observation in the four primary classrooms suggests that teachers base their assessment on two aspects of students' quality: cognitive achievement and behaviour. These two will be discussed in turn.

Cognitive Achievement

Judgement of cognitive achievement dominates teachers' practices. Students are promoted to the next grade according to a level or standard of achievement. As seen in chapter five, six categories of grading system are officially distinguished (Excellent, very good, good, medium, acceptable and poor). But, in practice, these are actually reduced into two categories, success (excellent to accepted), and failure (poor). The syllabus is the main source used by teachers to assess students achievement.

Mrs A:

As a class-teacher, Mrs A should steer her students according to curriculum objectives (competencies, as identified by the curriculum framwork). Among these, 17 are in Arabic and 19 in Mathematics (appendix. 1a). These competencies are mid-term objectives to be achieved at the end of stage one (grade 1 to 3). They are, therefore, not explicitly observed in the daily practice. Instead, textbooks steer the classroom practices. Writing and expression are also obvious. To achieve such objectives, Mrs A builds up high expectation of the students' cognitive achievement. She explicitly acknowledges her expectation in her comments upon student progress.

> Mrs A: What did you do at home, Hind? You were good and actively participated and answered. Now you are inert. Where is your excellence? One goes for good not for bad. Your scores are now only very good, or good. Now, what are you going to be? Excellent?
> Hind: (No answer).

Mrs A: You say, I will be what?.
Hind: I will be excellent.
Mrs A: I will work hard.
Hind: I will work hard.

Although Hind still gets good and very good marks, Mrs A categorises her as 'inert' in comparison with herself in the past. Answering teacher questions, participating and being active in the class contribute towards successful achievement. She makes this expectation explicit and insists that Hind understands what is required from her (the excellence grade) and how this aim could be achieved (hard work). Mrs A appears to give an open message to the class as a whole that excellence is expected and that hard work is required for such a goal. Nevertheless, judgement is not, in fact, being made against students' cognitive abilities. Mrs A' s explicit judgement is concerned, instead, with the overt behaviour of students which, she feels, has an unfavourable influence on achievement (e.g. inattentiveness and non-participation in classroom oral questions).

Mrs A's initial judgement is also flexible as she makes allowances for changes and improvement. In the beginning of the term, she talked to me about a new student in the class identified by Mrs A as below the grade level. She built up this judgement on the student's previous schooling. This student had started first grade in a private school and in her second year she moved outside the country where the school there kept her in first grade rather than promote her to the second. By accepting the students in grade three this year before completing grade two, Mrs A thought that the student would be behind her classmates. Later, however, Mrs A categorized the girl as slow but also showing improvement. At the end of the year,

this girl was promoted to year four, with good marks in all subjects.

On the other hand, norm-referenced criteria appear to be used as a basis for competition. In her day-to-day practices, Mrs A ranks the top ten in the class using two criteria, speed and correct performance. She encourages competition by awarding stars to students and she hangs these stars, opposite the student's name and picture, on a specially organized board. In addition, Mrs A labels her students in four categories: Distinctive achievers (11 students), good and very good achievers (10 students), acceptable but slow achievers (5 students) and slow achievers who need attention and focusing (4 students).

Mr B:

Mr B, as class-teacher, has the same curriculum objectives as Mrs A. Nevertheless, the most apparent aspect of Mr B's cognitive objectives in Arabic are reading comprehension and reading aloud. Free writing is nearly missing. His definition of mastery achievement is moderate.

> Excellent students should have good reading. Students should also be able to construct sentences and be well organized, with probably few spelling mistakes. Some spelling mistakes may even stay until later years of schooling.

But for poor performance his definition is less specific in terms of curriculum qualities. As he pointed out, students with poor performance may not master the competencies necessary for the first cycle. More specifically, he pointed out that students with less than 50 marks out of 100 in the Arabic subject should repeat the year (this is also the policy regulation). Such a statement did not identify

achievement thresholds. The teacher still has no explicit definition of what is the kind of achievement he should expect for all students.

Nevertheless, Mr B seems to accept students as they are, and does not anticipate changes in the class situation. He makes his judgement about the class as a whole as 'acceptable': 'A number of students show an overall good reading quality'. In the classroom observation, I did not notice encouragement, rewards, praise or any indication of pushing students towards short-term goals. He avoids making any explicit judgement of students' academic progress. Students also receive no messages about their progress. In his class, however, he makes comparison between students with regard to their behaviour in classroom tasks. Two criteria, speed and tidiness, are used.

> (To a less competent student, after failing in doing a task while another competent one has done it perfectly):
> Do you see this? It only takes seconds for him to do it.
>
> (To Jafar, when the teacher examined another student's book): Bring your book, Jafar.
>
> (To Noah showing him Jafar's book): You see, which is the best? A clean one like this or one like yours? Jafar is sitting next to you. Why don't you do like him?

Mr B appears to draw out an explanation for good and poor achievement, that is, attention is an important factor. For example:

> (Before asking question): Now, we will see who was paying attention and who was not.
>
> (To student who appears inattentive): If you don't pay attention, you will stay here next year. Do you want

to stay? This is not my problem.

In this respect, teaching and learning contexts are clearly separated. The teacher has the responsibility to teach but learning is the students' responsibility. By this explanation Mr B is trying to protect his professional identity as teacher when there may be indications of confusion or poor achievement. Students are frequently blamed for not paying attention to the teacher and not working hard enough at home. In addition, parents at home are expected to take their responsibility for children learning.

> I told you to ask them at home to teach you this. Why didn't you? When you go home today, ask them to help you in doing all these.
>
> And on another occasion to the whole class:
> Those who don't understand should ask their parents at home.

Despite his exhortations, parents did not seem to share the same view of their responsibilities. Mr B experienced a disagreement with some parents in regard to their children's final marks of mid-term. In the parents' meeting organized by the school (which I attended), I observed that several mothers became angry when they saw their children's scores. These mothers argued with Mr B that their children deserved better scores. Mr B tried to explain the results to them but they disputed his assessment. A few mothers took further action, and went to discuss their concern with the headteacher.

In Mr B's class, four students were identified as having difficulties in Arabic. They had been categorized as 'slow learners'. They, therefore, attend special 'extra' sessions with Mr S, the special teacher in the school. Mr S gives additional help to

the slow learners in the school for year 2 and year 3. In my interview with Mr S, he made two main comments: First, the fact that some of slow learners do not need, he feels, special remedial lessons but rather more attention from their teacher in their normal lessons. Secondly, that much remedial action for the real 'slow learners' should also be given by their teacher in normal sessions. He pointed out that this situation is not always the case, as some teachers think that the special teacher should do remedial work for other teachers.

These comments made by Mr S seems true for Mr B. Mr B seems not to give much attention to these students. When I asked Mr B about these students, he could point to two students and had to ask the class about the other two. Two students raised their hands. This incidents means that Mr B may not be concerned about the slow learners in his class, despite the fact that they have been defined as slow in order to receive more attention from their teacher.

Mrs C:

Mrs C's definition of cognitive achievement follows, to a large extent, the official framework of curriculum objectives (appendix. 1b). It includes the four primary constructs, listening, speaking, reading and writing. These four are manifest in Mrs C's classroom assessment. But reading comprehension still predominates other qualities. This is probably because the textbook originally revolves around reading texts. Free writing is given outside the formal sessions. Mrs C plans group and individual writing. In the group work, students are required to widen their knowledge and experiences by using different sources of information (library,

museum and newspapers). Speaking is also given attention. Mrs C alerts her students to use standard Arabic and avoid the local dialect. Students, in turn, talk with her in the same language. This type of linguistic approach seems effective in enhancing students' ability in oral discussions.

To achieve the above qualities, Mrs C had several methods. Encouragement and praise are frequently repeated, "Well done, your reading is excellent, God bless you, and very good'. These remarks announce the teachers' informal judgement of students' performance. It seems that Mrs C placed high priority on students' self-esteem. She shows a high respect for them, gives them as many opportunities to participate in answering her questions, and consolidates their confidence.

Students in Mrs C's class are generally at the same level of achievement. Very few show high competence, a few show low competence, but the majority are at the same level. For this reason, probably, differentiated treatment or interaction is hardly explicit. She makes no explicit judgement about students' capabilities nor makes comparison between students. Implicitly, however, Mrs A differentiates in her practice between high and low achieving students. She seems to address small questions, which require short answers to low achieving students. She also appears to give chances to high achievers to participate in classroom discussion first as a strategy to give the low achievers a chance to understand the context of the session and thus, later, to give correct answers.

For Mrs C, it seems that students' participation in the class discussion and

obtaining correct answers are two basic elements in encouraging students to improve their learning. Blaming and criticizing would, in her opinion, yield opposite results. In my interview with Mrs C, she seems not to accept failure in her class. She said that for the last five years, she has not had any failures. She relates this to her professional experience: 'In five years, thank God, I have had no failures in my classes'. Students' high performance levels seems to encourage her to build up this expectation. Nevertheless, she pointed out that she still has worries in regard to two students, Hella and Wadha.

Mrs D:

In Mrs D's class, the overall achievement level of the class, in her opinion, is generally good compared with the previous year. Mrs D's definition of good performance is a combination of different basic skills (basic principles in the four operations, memorization of multiplication table) as well as the ability to solve mathematical problems. She categorizes seven students as less productive and, probably, holding negative attitudes towards learning mathematics. Among these, however, she labelled Hella as a slow learner. Hella had been transferred from a private school. Mrs D said that Hella may have failed in mathematics and sciences in her previous grade. Mrs D arranges for Hella to sit in a front seat, so she can observe her work and give her more attention.

She also addresses more attention to the other less competent students. These students are clearly requested to pay attention and are repeatedly asked whether they can follow and understand:

> Rasha, Hella, Eyman do you understand? Wadha, Dahna, are you with me?

Mrs D seems to base her judgement of performance in relation to students themselves. Effort and improved performance are encouraged and praised.

> (To Rasha): Well done Rasha. You are very good this month. I hope this will be always the case.

She also affirms the students' responsibility to work hard. As she noticed that Hella was not doing her homework and did not seem to be making progress, she gave her sharp comments and threatened that she would report such laziness to the administration office. Mrs D expects also that all student should, at least, reach the minimum standards, as measured by their marks in tests.

> (After distributing students' test papers, Mrs D asks each student, in turn, to give her their scores while at the same time, she registers this in her marking book. The score given is out of 30).
> Mrs D: Shaika?
> Shaika: 15.
> Mrs D: Why, Shaika? Last time you had a better mark. You did not work hard this time. Mariam?
> Mariam: 29
> Mrs D: Very good Mariam. Hella?
> Hella: 18
> Mrs D: Still good, but I want a better mark next time.

The message is quite clear for those students. Students become aware that Mrs D is watching progress or improvement of each student. This awareness seems to encourage students to put more efforts into their learning.

Mr E:

Mr E builds up his judgement on his own experience with the class. This

judgement is, however, in conflict with students' previous year's records which show good and very good marks. As he pointed out:

> Most students have been progressed from last year having achieved, as records say, a very good mark. But their performance, in fact, is not so. So how can I give lower marks to them this year? And if I did so, how could I explain this result to their parents?

Classroom observation, reveals that students' cognitive achievement is below their actual grade level. There are also seven students who, in their earlier years of schooling, have been identified as 'slow learners'. They attended remedial sessions with Mr S. Mr S gave me these students' reports from the previous years (grade three), and it seems that they have very serious deficiencies.

In frequent cases, Mr E finds himself exposing very basic skills of the subject, such as, the characteristic of the simple sentences, the verb and the noun and other simple grammatical rules. Reading aloud is not a routine practice which can be fluently accomplished. It always goes with difficulties and the need to explain some basic rules in reading. Free writing is overlooked. In such a situation, there is a real tension between the actual performance of the class and the syllabus. On several occasions, Mr E notes that knowledge and skills he exposes are actually from previous levels of schooling.

> (After recognizing that many students did not understand his question): I know that this is difficult for you, but you have to practice this.
>
> Mr E: A year-four student can answer this question.

In turn, Mr E faces two practical problems. First: how can he make progress with a

prescribed syllabus which is above his students' achievement level. Secondly: how can he restore a reasonable balance between achievement levels and assessment categories. Mr E's resolution focuses on progression with the syllabus. On the other hand, Mr E gives an implicit definition of what is a good achievement. Mr E makes frequent comparison between Mahmood, the highly competent student in the class and the other students. He probably hopes that students could see an obvious example which they could copy. But, on the other hand, the distance between Mahmood's performance and the other students' performance is quite large, such that the comparison may yield an opposite effect on students' motivation. He, however, gives an ambiguous definition of what is poor performance. He defines students' reading progress, in general, as 'slow'.

> Mr E (after listening to one boy's reading):
> How can we call this reading? Slow?
> Student M: He can read but his reading is slow.
> Mr E: How can this problem be resolved
> Student R: Reading more
> Mr E to the student : Right, by practising my boy,
> give your self 20 minutes every day for reading and
> your reading will improve.

This definition is not clear for students because it defines no particular problem. A good reader also can be slow in reading while being fast is not evidence of good reading. The student still does not know what is the problem. In addition, the solution Mr E gives could not be effective unless the student's basic reading deficiencies are overcome.

Small progress in students' performance is, however, praised and encouraged.

> Student (After teacher had announced the right
> answer): This is the first time I am getting the right
> answer.

> Mr E: This is only the beginning and I hope you always do the same. I noticed that you had made an effort.

In my discussion with Mr E, he attributed poor performance to several factors: The time for each lesson is not enough, students do not work hard at home as girls usually do, and attention of the boys' parents is not enough to make them work or assist them in learning. To overcome this problem, he suggested two solutions: continuous practice by the children and increased parental attention. By this argument, he seemed to dissociate himself from underachievement. The problem is not his responsibility; the cause lies elsewhere.

Mr F:

Mr F is in the same position. Although priority is given to progress with the syllabus, there seems to be a tension between what is expected in the syllabus and the actual performance of the class. As a result, for example, problem solving is rarely given and complicated mathematical operations required in the textbook are overlooked or only given to a few students who are more capable.

Mr F has already built up his judgement based on previous experience with the same students.

> I know all the students in the class as I taught them in year five. For me, it is fair enough if a student understands and works in the class. If he does not, I just leave him.

In such a stance, Mr F separates himself from the responsibility of students' learning. He, sometimes, motivates students to work by giving marks for how

much they work, however this work had been performed. The students' performance is then differentiated by how much work they complete.

> (To the whole class): Mahmood is excellent, He finished five exercises.
> (To Talal): You did not work today.
> (To Bader): Bader one step forward, and another step backward.
> (At the end of the session): Let's see those who did not do any work today.

For some students, this strategy seems to work as they compete for scores. For others, it does not. In this latter case, the students' motivation to learn becomes their responsibility.

Mr F makes inappropriate judgements that some students are hopeless and therefore, he could not do much about them. His judgements and expectations of these students are made explicit. He comments:

> You are hopeless.
> You are still slow.
> Mansor, listen and concentrate, maybe something will get into your head.
> Don't you ever understand? This is a revision session. Revision is important. I think you will never understand.

Mr F showed me Mansor's book, probably to explain the above comments he frequently makes. He commented that Mansor does not make any effort at all . Mansor's book was nearly empty. The teacher's explanation that Mansor is not making any effort may be true to a certain extent. As I observed this student in subsequent sessions, he was distracted in most lessons. In the beginning of the term, this student showed several deficiencies in mathematics. He could not, for

example, give the answer of how to divide 32521 by 9. He also couldn't give a correct answer in the multiplication table that Mr E revises with the class. But Mansor still makes attempts to learn. I watched him in a geometry lesson, trying to draw a triangle using a ruler and compasses. The three sides of the triangle were given. Although his attempt was consistent, it was also unsuccessful. He, in the first place, was not paying attention to the teacher while he was explaining this task to the whole class. Mr F, although he passed him twice, did not pay attention to what Mansor was doing. Mansor spent time trying on his own but he seemed confused about the role of the compasses in this task. Finally, he drew out the triangle by the ruler only and drew out two small curves in the top of the triangle as an indication that he had used the compasses. Mr F marked Mansor's work as correct and asked him to do the next one.

Interpretation and solutions:

Generally speaking, it seems that explicit judgement is postponed until the first month or mid-term test is undertaken and marked. Those who have not passed (get less than 50 per cent of total score in the subject) are then identified as failures. Teachers then interpret the students' failure. Mrs D, Mr E and Mr F perceive lower performance as a result of inattention.

> Mrs D: You did this on your own Wadha?
> Wadha: Yes
> Mrs D: Very good. You see the result when you pay attention.
>
> Mr E: You do not understand because you do not pay attention.
>
> Mr F: We have been two weeks in this part (long division), but there are still some who don't

understand. The reason is, inattention.

But more important than interpretation is to find ways to moderate students' scores and improve performance. Moderation takes priority in Mr E's and Mr F's class. One solution, in this respect, is for teachers to compensate for students' low scores through the day-to-day activities. That is, to use students' work in the classroom as another type of test, to mark it and subsequently to moderate student's scores.

> Mr E (To the class as a whole): Your results in religion and social studies are not better than Arabic. I will give you the papers, and bring them back with your parents' signature.
>
> (Later in the same session): Saturday, will be a reading session. Lets hope that your reading will compensate for the low scores. You have Thursday and Friday for practising.

In Mr F's class, 10 students failed in Mathematics in the mid-term test.

> Mr F: I will give you some exercises to practice more. Those who failed should try to compensate their scores.

By the end of the session, 10 students gave Mr F their books to mark. This type of solution could not have an immediate effect on students' total marks because students' classroom work was not better than what they have done in the test. Summative assessment seems to be central in the teacher's intention but this summative assessment is distorted. Real improvement of learning is, therefore, disregarded.

With regard to female teachers, it seems that they are in a different situation. Girls are apparently more competent than boys and are motivated to learn. Thus, those

who have been labelled as failures accounted for only 4 students in Mrs D's class. Mrs D tries to find some explanation for students' failures and looks for co-operation with parents. Attention, success and performance are explicitly linked.

> Mrs D: Dahana, you got 29 out of 100, Wadha 22, Hella 30 and Ouhood 49. (All these students stand as soon as the teacher announced their names) Why? Don't you revise at home? Don't you pay attention in the class? (No response from any student)
> Mrs D: I will have to call your parents.

From the above observation, it seems that drawing a boundary between success and failure has an impact on the teachers' practice. Teachers base their judgement in accordance with this criterion. Students who obtain 50 and more are placed beyond further consideration. Despite the fact that one who gets 51 is likely to be at the same position of achievement of that who, for example, gets 49 per cent. The boundary between success and failure is, however, not always clear cut. Students who get, for example, 49 can be squeezed through. So the actual passing mark becomes 49. This probably can be applied for the next lower mark 48 and so on. In addition, although curriculum objectives transferred via the textbooks, are the main criteria for assessment, these criteria seem to overlap with the norm-referenced criteria, that are used to give the students final marks at the end of the term.

Behaviour

The second aspect of teacher assessment is student non-cognitive behaviour. I had the opportunity to examine students' records for the four observed classrooms. It was striking to see that while students' behaviour was fully documented in the girls' school, it was entirely absent in the boys' school.

These records are of two types: official records and school records. Official records are uniform among all schools in the country. They are organized, issued and imposed by the Ministry of Education. These records must be kept in school and should be transferred if a student moves to another school. The record contains end of year reports with students' scores in all subjects together with teachers' comments on their progress and behaviour. School records, on the other hand, are mainly organised by schools themselves for reference and for informing parents of students' performance and behaviour. They contain students' mid-term marks in all subjects and also comments upon students' behaviour.

In the official records, Mrs A gives a brief description of students' behaviour and progress; for example:

- polite. quiet and co-operative;
- Polite, quiet and needs prompting;
- Very good but seems worrying and fearful.

In the school's reports Mrs A organizes a uniform record which consists a list of behaviour. It includes: co-operation, quietness, cleanliness, participating in

activities, participating in classroom discussion, caring, understanding, tidiness, class work and finally homework. Each student is positioned against these criteria on the basis of whether the student conforms to this behaviour or not.

It can be said that the above behaviour is confirmed in the classroom situation. It is part of the teacher's teaching strategy and part of the students' learning situation. Teaching the behaviour is actually part of the students' learning.

> Tidy your work. Go and repeat this again.
> Your nails are long, cut them please.
> Wash your hands before the break.

In contrast to Mrs A, Mr B's records have no place for any comments on students' behaviour. Students' official records also show no comments by their previous teachers. My classroom observation also supports the view that teachers pay little attention to boys' behaviour. In Mr B's class, students' general behaviour such as politeness, tidiness and cleanliness are overlooked. He only emphasizes instructional behaviour (e.g. pay attention to teacher and keep quiet). At the same time, however, he seems unable to control the class as a whole. Conventional rules were broken every minute, but Mr B took no further action, other than to issue orders

> Those in the back, listen.
> Ali, you will be out if you don't keep quiet
> Keep quiet, this is enough.

In year six, records of girls' behaviour describe the extent to which a student conforms to the instructional rules of the class. The following list, for example, is

derived from students' official records. It shows the types of behaviour commented upon which had been recorded by the teacher in the previous year (1993/94).

Favourable	Number of students	Unfavourable	Number of students
Quiet	16	Needs care	2
Polite	15	Carelessness	1
Hard working	9	Slow	1
Participating	2	Stubborn	1
Energetic	3	Needs attention	1
Does homework	1	Oblivious	2
Excellent	2	Inattentiveness	1
Distinctive	1	Withdrawn	1
Clever	1	Selfish	1
Cooperative	1		
Curious	1		

It appears that quietness, politeness and effort are seen by teachers as positive and expected behaviour for girls. Excellence, distinctiveness and curiosity is only noted in exceptional cases. On the other hand, unfavourable behaviour is also exceptional. Students seem to conform to the teachers' favourable expectations. They, for example, prepare for each session at home, do their homework, show politeness and comply with the teachers' rule in the class.

School reports offer more extensive descriptions of student' behaviour. These reports are written with parents in mind. Descriptions of girls' behaviour are similar in tone and context:

> **Mrs C:** Very good in all subjects. She maintains her attainment level. Does all her homework, participates, animated, polite, quiet, confident and responsible.
>
> **Mrs D**: She is polite, quiet and excellent in every thing.

On the other hand, low achieving students, are characterised using a different set of

descriptors:

> **Mrs C:** Inactive in the class, slow in writing, needs more attention and prompting. Her progress is slow in religion and social studies but satisfying in Arabic.
>
> **Mrs D:** Inactive, does not try hard.

It seems that learning progress is at the centre of teacher attention. Assessment of other characteristics of students' behaviour comes later to confirm the assessment of learning.

These observations imply that students are categorised against forms of ideal behaviour. Some behaviours are stressed such as, politeness, quietness and hard working. These in fact, identify attributes of socially approved behaviour. Other behaviour is entirely excluded. For example: creativeness, open mindedness and independence. The implication of such differences is that socially beneficial behaviour is given much more attention than that which renders benefits to individuals. This is actually not surprising in Arab society. Individuals derive their identification from their families and their behaviour has to conform to what the family approves and accepts. As Halim Barakat (1991) noted:' Affiliation is the prevalent value in the Arab societies (p. 348).

The importance of being polite and quiet promotes obedience on the part of the individual and permits control on the part of the family. This, inevitably, is transferred to the school, as Sharabi (1991) pointed out in his comment over the Arab societies and education:

> Didactics is the most organized way of controlling and domination....This way of education and its relation with the social reality, consolidates the power, prevents questioning and suspends creativeness (pp. 39, 40).

It seems, however, that such control is mainly imposed, as evident by schools, on girls. Boys seem to be free from such control or superintendence over their behaviour. In this study, the behaviour of boys is disregarded. Although the official record allocates space to students' behaviour, these remain empty. None of the boys' school reports, which I examined, gave attention to classroom or school behaviour. Teachers appear to pay less attention to students' general behaviour - whether they lean on desks, slump in seats or are distracted by other things while teaching.

In my interview with Mr E, he defines students' behaviour as 'immature'. He thinks that this is part of the class-teacher system applied in the earlier years. He thinks that the class-teacher system gives much freedom to children as part of its philosophy. As a result, he thinks, students have less control over their behaviour. Discipline is therefore, missing. He also thinks that punishment should be imposed upon students who misbehave in the class. He emphasizes instructional behaviour only and, therefore, expects students to pay attention, to bring their books, to do their home work and finally, to revise and prepare work at home. Mr F, on the other hand, applies no disciplinary sanctions in his class. Approved and disapproved behaviour are not clarified for the benefit of the students. Students' inappropriate behaviour in the class such as inattention and distraction, seem not to aggravate Mr F.

Conclusion:

Assessment practices need to be understood in relation to classroom context. Teachers, in this study, although committed to curriculum, find students an important constraint. Students' characteristics influence, and interfere with, any classroom procedures that teachers make. Within these two main constraints, students' characteristics and curriculum requirements, teachers follow reconciliation procedures. When students' characteristics are not appropriate to accomplishing the curriculum, the constraints over teachers increase. In a girls' school, such a situation does not exist because the students' cognitive achievement matches the curriculum objectives. This is, however, not the same in the boys' school because there is a great distance between students' achievement levels and the curriculum. Such conflict is less apparent in year three because students are still at the early years of schooling and thus their weaknesses may not be sharp and clear.

Critical in all situations is the teacher's perception. In this chapter, it is apparent that the female teachers have a more comprehensive perception of their role as teachers. In the earlier years Mrs A acts as a second mother toward children. Students' whole personalities are always considered. Behaviour and cognitive qualities are interrelated. In the boys' school, the teacher's role in both years is confined to instruction. In addition, teachers seem to put themselves outside the causes of underachievement. This is important in terms of defining teachers' responsibility and practice. Such a perception, held in the teacher's mind and also clearly declared, is likely to limit teachers' practices. The same phenomenon has been observed elsewhere. King (1978), in his study of three English primary

schools, pointed out that teachers in the U.K also adopted a similar distanced view of their responsibilities. They

> were absolved from responsibility for the undesirable changes; a child's bad behaviour or poor progress was not their fault. This protected their professional identity and left unquestioned their practices and the ideology they were based upon (p. 65)

Thus, teachers' practice and ideology are unquestioned and improvement of learning may lie somewhere else outside the province of teaching. In this sense, the teacher is unwilling to question his/her practices or to alter his/her teaching tactics.

As a result, teachers' assessment practices are affected. Although cognitive achievement dominates teachers' assessments, students' behaviour still seems to be an important aspect of assessment in girls' schools. Teachers' definition of students' ideal behaviour is, however, derived from the outside community. The values of society as a whole is the main source of such assessment. On the other hand, in the boys' schools observed in this study, assessment of behaviour is entirely absent. This is actually not to say that the assessment of cognitive achievement is independent. The non-cognitive aspect of students' qualities such as motivation, attention, enthusiasm and disruptiveness seem to be influential in the assessment of students' cognitive qualities. Although students non-cognitive qualities are overlooked, they actually interfere to distort the validity and reliability of teachers' assessment of cognitive achievement. It is evident that teachers' assessment in the boys' school is far less comprehensive than was apparent in the girls' school. For example, extended writing and listening qualities are not a

subject of teachers' instruction and assessment. This absence is also evident for problem solving in mathematics, a daily practice in the girls' school. This approach to assessment also incorporates the teachers' intervention to moderate the actual achievement results of their formal paper and pencil tests as will be revealed in chapter 8.

The next chapter provides further analysis of teachers' practices in the classroom but with particular attention to teacher-pupil interaction.

7

Teacher-Pupil Interaction

Classroom assessment involves using several modes of information gathering: questioning, discussions, observation, correction and marking. Much of these modes are reflected in interactional situations. Teachers are frequently interacting with students to check their understanding. They ask questions, conduct classroom discussion, organise tasks, observe children and accordingly, collect much information. This information conveys evidence about students' learning and progress. The quality of such evidence depends on the quality of teachers' and pupils' interactions. This chapter is concerned with this aspect. The focus is on the cognitive-related interaction since it dominates classrooms life. Such interaction is broken down into two types: question-answer and task-related interactions.

The following investigation of classroom interaction is derived from all sessions observed. Nevertheless, the statistical data is basically obtained from 17 sessions (15 hours) for each teacher. The classroom observation in Bahrain confirms Galton

et al's (1980) 'asymmetry' of the interaction, mentioned in chapter 3. Teachers seemed to interact with many students in one session, but, in contrast, each student interacted with the teacher for only a few seconds. These interactions, however, differed among teachers as follows:

	Average number of students' Interaction in one session
Mrs A	8
Mr B	11
Mrs C	15
Mrs D	15
Mr E	10
Mr F	20

Teachers interacted, on average, with at least a third of the students in each session. This number increased to two-third in Mr F's sessions. The observation notes, however, revealed that some students had no interaction at all in the sessions observed. This was observed in Mrs A, Mr B, Mr E and Mr F's classes.

Two types of interaction seemed to shape classroom life:

1. Question-answer interaction: The teachers' approach in this mode is whole-class interaction which gives little attention to individual students. The teacher asks a question and allows students a few seconds to think. S/he scans the whole class to see students' faces, and whether they appear confused. If not, the teacher notes how many students have raised their hands. This initial information gives the teacher an opportunity to assess the students' general understanding and to evaluate her/his teaching. Each student who is given a chance to answer is regarded as a member of the class; or as a sample from the class. But only lightly-sampled, fragmented knowledge and skills are assessed for each student.

2. Task-related interaction: In this type of interaction, the teacher's attention to the whole class is decreased. Teacher's interaction is with individual students in relation to a particular task (e.g. problem solving or reading aloud). While the students are working on the task, the teacher is able to interact with a large proportion of the class. Thus, the total number of students the teacher interacts with is obviously increased.

Question-Answer Interaction

In this study, a question is defined as 'any statement intended to evoke a verbal response' (Brown and Edmondson, 1984, p. 99). This definition includes statements which might come in question form, that is, using " 'Why?', 'What?', 'How?' and 'when?'". And it also includes any statement which contains an implicit question, that is, using the command form. A question is also counted as one even if it evokes answers from more than one student. For example: 'Name animals that live in icy areas'. Several students could give answers to this kind of question.

Within this type of interaction two patterns can be distinguished: single question answer interaction and reciprocating question-answer interaction. This first pattern occurs when the teacher asks an autonomous question which is not followed up with other questions to the same student. The student's interaction with the teacher, in this case, ends as soon as s/he give the required answer. In the second pattern, the teacher asks a question which is followed by other questions. In this pattern, the teacher's attention to individual students becomes more apparent. The time

given to each individual student is extended as further skills and knowledge are assessed.

Single question-answer pattern:

This pattern dominates teachers' classroom interaction. It is used by all the teachers observed, albeit with differences. Table 7.1 shows how this pattern was distributed among students. One opportunity represents one question-answer interaction with a student whether the student has answered the question correctly or not. If a student is asked two questions in one session, these are recorded as two opportunities. Students' opportunities are categorised in groups of five in the first column of the table. Columns 2 to 7 show the number of students who have these opportunities.

Questions - answer interaction (opportunities)	Third-year class-teacher		Sixth-year girls		Sixth-year boys	
	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
Zero	-	-	-	3	-	11
1 - 5	13	20	1	23	22	22
6 - 10	15	7	5	6	9	1
11 - 15	2	2	15	-	3	-
16 - 20	-	-	6	-	-	-
21 - 25	-	-	3	-	-	-
26 - 30	-	-	1	-	-	-
31 - 35	-	-	1	-	-	-
Total number of students	**30**	**29**	**32**	**32**	**34**	**34**
Total oral questions	**186**	**136**	**514**	**150**	**237**	**117**

Table 7.1 Students' opportunities in questions-answer interaction[1].

Table 7.1 reflects two general points: First, a questioning strategy allowed teachers to interact with many students, if not all of them. All teachers interacted,

[1] All data of this chapter is derived from the researcher's own observation of this study

individually, with a large number of students with the exception of Mr F. Secondly, students' opportunities to interact with the teacher varied. There were 3 and 11 students in Mrs D and Mr F's classes respectively who had no opportunity of question-answer interaction. On the other hand, there were some students who had a higher number of opportunities than others. But in Mrs C's class, variations among students were more apparent. Those who had modest opportunities (6 to 20) accounted for 26 students. One student had between 31 to 35 opportunities while another one had only 1 - 5 chances. These variations meant discrimination by the teachers.

This discrimination, as far as assessment is concerned, was investigated in relation to the students' achievement levels. In the observed classes, 15 students in each class were chosen from different performance levels (5 high, 5 medium and 5 low achievers). Their opportunities in answering teacher's questions in 17 sessions are shown in table 7.2. For comparability purpose, the data are represented in percentage form.

Students' performance level	Third year class-teacher		Sixth year girls		Sixth year boys	
	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
High achievers	33.0	41.5	40.0	39.1	53.3	47.4
Medium achievers	32.0	24.0	33.6	19.6	30.7	15.8
Low Achievers	35.0	34.5	26.4	41.3	16.0	36.8
Total opportunities	100	100	100	100	100	100

Table 7.2 Students' opportunities in terms of achievement level.

The table suggests that differentiation in relation to students' achievement levels varied among the observed teachers. Mrs A did not seem to make any

discrimination between students while Mrs D addressed more questions to low achievers than to the other two groups of students in the class. The other teachers seemed to favour the high achievers. Variations in favour of the high achievers are evident in both Mr E's and Mrs C's classes (where 53% and 40% of the total opportunities were offered to high achievers as opposed to 16% and 26% to low achievers). It also seems that medium achiever students were given less attention in Mrs C, Mr B and Mr E's classes. These practices and variations were investigated further, with respect to the types of questions asked by the teachers.

Questions can be differentiated in several ways. Some research has used Bloom's taxonomy to classify questions into six categories (see Sanders, 1966). In this study, however, the classification devised by Walker and Adelman (1975) was adopted. Walker and Adelman named two kinds of classroom content control, open and closed. Questions in this study were, therefore, classified in two main types, **open questions** where the teacher is prepared to accept a range of answers; and **closed questions** where the teacher is prepared to accept one correct answer only.

This distinction relates to the mode of thinking that the teacher wishes to encourage in students. Closed questions restrict students' thinking to what the teacher wants. In this case, students are passive learners. Open questions encourage students' creativity and independent thinking. Students answering open questions are assumed to be active learners and able to think of alternative solutions. This distinction also relates to convergent or divergent thinking, terms popularized by Hudson (1966). More recently, Torrance and Pryor (1995) have used the same

terms, convergent and divergent, with respect to teacher assessment. They suggest that convergent teacher assessment is a behaviourist approach. An 'assessment oriented to the accomplishment of particular attainment targets, rather than the pursuit of what one might characterise as more the more general development of student understanding of the task at hand' (p. 315). Divergent assessment, on the other hand, is associated with constructivism. Such assessment approach puts 'the emphasis on the learner's understanding rather than on the agenda of the assessor' (p. 316).

Table 7.3 gives the number of (open and closed) questions asked in the classrooms observed.

Teachers	Year-three		Year-six (Girls)		Year-six (Boys)	
Types of question	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
Open	12	13	23	-	7	-
Closed	174	123	491	150	230	117
Total questions	186	136	514	150	237	117

Table 7.3. Types of questions in the observed classrooms.

As indicated in the table, closed questions predominated. Open questions were entirely absent in mathematics sessions, and they seem exceptional in other classes. This result supports the view that the context of the teaching/learning in primary school is didactic. It seems that primary teachers restricted their students' intellectual thinking.

Closed questions enabled teachers to assess precisely what their students had

learned since the answers were clearly specified. They look back to what had been learned but this kind of review does not allow teachers to look ahead to what students might be able to do in new situations. The advantage of closed questions is that they give teachers an overview of students' general understanding, and allows them to compare students. Each student's answer is assessed against the other students' answers. In these types of questions, student individuality is denied and the teacher is prevented from seeing their idiosyncrasies. As Rowntree pointed out:

> [I]f we rely too heavily on highly structured, standardized assessment tasks in pursuit of comparability, measurability and fairness to 'students in general' we may learn more about the similarities among students (and help create them?) than about ways in which they differ. We may lose touch with the creative, perhaps idiosyncratic qualities that are so productive of new, unanticipated answers and up-ended questions (Rowntree, 1987, p. 158).

The form of teachers' responses is another variant in teachers' interactions. Teachers' responses constitute the feedback from teacher assessment. Such feedback is an important element of formative assessment. It, therefore, indicates the teacher's intention for the assessment. The teachers' responses were, therefore, classified into four types: a) Collaborative responses, where the teacher guides students' thinking, elaborates ideas for example, rephrases the question, provides some clues or asks another question (eg, why?); b) Positive responses, where the teacher accepts students' answers and gives positive comments (e.g. 'well done', 'very good', 'excellent'; c) Neutral responses, where the teacher takes a neutral position, makes no comments, repeats the student's answer, or poses another

question; and finally d) Personalised responses, where the teacher reacts negatively to students' answers by blaming students and commenting on their ignorance.

Collaborative responses may have formative benefits for student learning. they maintain the value of learning itself. Thus, both the teacher and the student show a concern for learning rather than, for example, success or failure against predetermined criteria. Positive responses can be regarded as the teacher's reward to students. They carry the teacher's implicit confirmation of the students' success. They provide students with immediate feedback about their answers and may also enhance their motivation. Neutral responses may only have the value of confirming the correct answer for students, and personalized responses publicise the teacher's private judgement of student failure. Details of these types of reactions are given in table 7.4 in percentages.

Years	Year three		Year six		Year six	
Teachers' responses	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
Collaborative	12.9	19.0	21.1	18.6	10.2	9.1
Positive	5.5	6.5	20.8	6.0	0.6	9.1
Neutral	73.0	70.8	55.3	70.0	65.4	70.3
Personalised	8.6	3.7	2.8	5.4	23.8	11.5

Table 7.4. Teachers' responses in question-answer interaction.

As shown in the table above, teachers' responses were predominantly neutral (55 to 73 percent). Collaborative responses were more frequently given by Mrs C, Mrs D and Mr B and personalised responses appeared high in Mr E's class and to a lesser degree in Mrs A's and Mr F's classes.

The evidence gathered suggests that, although questions were the teachers' main instrument of classroom interaction, they were not of a kind which might be expected to assess students' cognitive learning. Rather, questions were closed, focusing on factual knowledge and textbook skills. Such teacher assessment is only weakly formative. It gives little attention to students' individual differences. Likewise, individual differences did not seem to be incorporated into teacher assessment. Students who seemed more in need of the teacher's help were denied diagnostic interaction opportunities. The teachers' assessment interventions for improvement were minimal. Teachers seemed to have other intentions. Their questions were directed more to revision, to obtaining general information, to gaining students' attention and to the overall management of the classroom life.

Reciprocating question-answer pattern:

This mode of interaction takes longer (at least half a minute). It is based on question-answer interaction. Its main feature is that the teacher asks several questions to one student at a time. Their questioning revolves around one topic. One basic question is followed by sub-questions, as in the following example:

> **Topic: The Gulf States Co-operation Council (GSCC).**
>
> Mrs C: Let us talk about the GSCC. What are the aims of the council?
> Several students raise their hands.
> Mrs C: Asma?
> Asma: To consolidate connections between the members and citizens of Gulf States.
> Mrs A: Yes. This is right. But what is the meaning of connections?
> Asma: It is ... relationships.
> Mrs C: And how about consolidate? What is the meaning of consolidate? We consolidate our

relationship? (at the same time, Mrs C presses her two arms together in an attempt to help Asma extract the meaning).
Asma: Unite or incorporate?
Mrs C: Yes. That is right. Good. Now are there other aims for GSCC?
Asma: Co-operation and co-ordination in the main areas of development.
Mrs C: And a third one?
Asma: To unify regulations for the benefit of their citizens.
Mrs C: Very good Asma. Thank you.

In such cases, the teacher assesses a student's understanding in one area of knowledge or skill. To pursue this understanding, the teacher asks the student provoking questions. The teacher's action is more individualized. S/he consolidates and assesses student learning while other students listen to the dialogue. This approach was only used by Mrs C and Mrs D. But even with these two teachers, it cannot be considered the teacher's normal practice. As seen in table 7.5, it was only used occasionally. One explanation for this is that it requires more of the teacher's time. This time is not always available with the large number of students in the class and also with the purpose of accomplishing the syllabus in mind. Table 7.5 shows teachers' interactions with students according to their achievement level.

Students	Number of students	Mrs C		Mrs D	
		Opportunities	Opportunities per student	Opportunities	Opportunities per student
High achievers	7	13	1.8	4	0.6
Medium achievers	17	25	1.5	9	0.5
Low achievers	8	3	0.4	2	0.2
Total	**32**	**48**		**15**	

Table 7.5. Reciprocating interaction in Mrs C's and Mrs D's classes.

Low achieving students seem to have still been neglected in this type of

interaction. Of eight low achieving students, only three and two students, respectively, had the opportunity with Mrs C and Mrs D. It seems that reciprocating interaction was used more frequently with medium achievers but proportionally more with high achievers.

Task-Related Interaction

This mode of interaction is important for assessment because the teacher's attention is on the work of one student. Weaknesses are identified and feedback is provided immediately. In the session observed three kinds of classroom task were used: doing a task on the board, reading aloud and classroom work.

Doing a task on the board:

This task was mainly a whole class task. Although one (sometimes two students) did the task on her/his own, the whole class followed what the student was doing. The teacher followed the work and corrected students' errors at the same time.

Table 7.6 gives the number of occasions offered to students for this type of task.

Doinging a task on the board (opportunities)	Third-year class teacher		Sixth-year girls		Sixth-year boys	
	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
Zero	15	25	4	-	34	16
1	10	4	12	3	-	13
2	4	-	6	8	-	5
3	1	-	6	7	-	-
4	-	-	3	5	-	-
5	-	-	1	4	-	-
6	-	-	-	4	-	-
7	-	-	-	-	-	-
8	-	-	-	1	-	-
Total number of students	**30**	**29**	**32**	**32**	**34**	**34**

Table 7.6. Students' opportunities for doing a task on the board.

The table suggests variations among teachers. Mr E's class had no interaction of this kind. Mr B gave less emphasis to this kind of work, as these opportunities were given to four children which only lasted for a few seconds. In Mrs A's and Mr F's classes, opportunities arose for nearly half the class. But, in Mr F's class there were 16 students who were offered no opportunities over the sessions observed (15 hours).

Mrs A addressed tasks on the board if she noticed that a student had some difficulties in doing a particular task. She invited the student to the board and asked her to do the task while she observed. Her interaction was obviously individualized. Mrs A might also give individual instruction and then allow the student the chance to repeat the work while still observing what she is doing. This interaction ended when Mrs A was sure that the student reached a point to practice on her own.

It seems, however, that the female teachers in year six devoted more time to this kind of work. In Mrs C's class, however, this task was combined with question and answer interaction. A chance to answer the question was combined with another chance to write the answer on the board. In the process, she gave attention to the student, checked what she did, pointed to any errors and asked the student to correct them. The main benefit of this task for Mrs C was, however, for the class as a whole. All students were asked to copy the correct answers from the board. It also drew the class's attention to the common errors the students usually make, for example, in spelling and punctuation. This was also confirmed by the fact that these opportunities are offered to high achieving students.

	Total opportunities	Number of students	Opportunities per student
High achievers	17	7	2.4
Medium achievers	32	17	1.9
Low achievers	10	8	1.2

Mrs D, on the other hand, is the only teacher who offered this task to all students. She also seemed to give more opportunities to low and medium achievers than to high achievers.

	Total	Number of students	Opportunities per student
High achievers	19	7	2.7
Medium achievers	63	17	3.7
Low achievers	30	8	3.7

Mrs D asked students to explain their work. While she listened and observed the student, she identified weaknesses and provided immediate remedies. Nevertheless, although this process may give valuable information about students' difficulties in

mathematics, Mrs D sometimes took the opportunity to go around the class and check the work of other students. On other occasions, if the student seemed incapable of performing the task, the teacher transferred it to another student. In such cases, the original student could not discover what kind of errors had been made and how they could be corrected.

In Mr F's class, task-related interactions rarely occurred. Twenty three opportunities were given to 18 students in the 17 observed sessions. While students worked on the board, Mr F stood by them and guided them through each step. Several factors, however, appeared to contribute to reducing the effectiveness of Mr F's help. First, Mr F dictated the steps as if he himself was performing the task. When a student made any error the teacher sought a correct answer from the class as a whole or sometimes gave the answer himself. The student, therefore, was given no opportunity to think independently or to discover his errors. Secondly, the teacher also gave attention to stimulating the class to guide the student on the board and, by the same means, keeping the class under control. Thirdly, Mr F entrusted each task to more than one student, which prevented individual students from tackling the problem as a whole or to correct their errors before another student took over. Overall, it seems that assessing students' learning was not Mr F's main intention.

Reading aloud:

Reading, as pointed out earlier, is practised only in Arabic. Thus, the information below covers four teachers only.

Reading Opportunities	Third-year class		Sixth-year girls		Sixth-year boys	
	Mrs A	Mr B	Mrs C	Mrs D	Mr E	Mr F
Zero	11	12	16	-	7	-
1	7	6	9	-	16	-
2	8	5	4	-	8	-
3	3	4	3	-	1	-
4	-	1	-	-	2	-
5	-	1	-	-	-	-
6	1	-	-	-	-	-
Total number of students	30	29	30		34	

Table 7.7. Students' opportunities in reading interaction.

Table 7.7 shows that opportunities varied among the teachers. Many students had only one chance to read, and between quarter to half in each class were offered no opportunities at all.

Mrs A gave the student at least one paragraph to read and approached the student when she read. The reading was followed by all students in the class, holding a pencil to follow the text. Mrs A offered help to students when reading. She corrected mispronounced words and asked students to repeat the word. If a word was mispronounced by more than one student, she asked the whole class to repeat the word together several times. The following example, from field notes, illustrates Mrs A's approach in interacting with students:

> Eyman is reading aloud from her textbook. The whole class follow her from their own books. Eyman misspelled the word {Nomowan}. She read it as {Nawman}. Mrs A stops her.
> Mrs A: Nomowan.
> Eyman repeats the word: Nomowan.
> Mrs A (to the whole class): Nomowan.
> The whole class: Nomowan.

> Mrs A explains to the whole class that there is a difference between the two words. She writes the two words on the blackboard and invites students to identify the differences in spelling and meaning. Students offer several differences. Mrs A asks them also to look for syntax cues. They start to think again and one student identifies the critical cue.

Reading, in this sense, provided students with practice in becoming competent readers; and it helped them to recognise weaknesses. Reading aloud also served other objectives, such as word recognition, writing and comprehension. After accomplishing the reading task, students were given the opportunity to ask teacher the meaning of new words. Each word was written in the board, explained, and students had to try to put these new words in sentences of their own.

Mrs A did not forget to give positive rewards for good reading (e. g. thank you, your reading is excellent). Mrs A's practice also does not discriminate between students in relation to their achievement level as shown below.

	Number of students	Opportunities	Opportunities per student
High achievers	11	15	1.4
Medium achievers	10	9	0.9
Low achievers	9	14	1.5

In Mr B's class, the focus of objectives was the same as in Mrs A's, but the practice differed. The process of reading aloud involved Mr B selecting one student to read aloud at least one paragraph while the class followed from their own textbooks. Once the child completed the paragraph, his work was done. The turn

was given to another student. Students' mispronunciations were corrected immediately by the teacher while the student paused for a few seconds and then continued reading. Sometimes the student repeated the word after the teacher, but not always. The teacher seemed to offer more opportunities to students who read in a clear voice.

> In a reading session, Mr B gives four students the chance to read aloud. Each accomplishes reading one paragraph with few errors. Several students still raise their hand asking permission to read. Mr B asks Fadel who also raises his hand: You want to read Fadel?
> Fadel nods his head.
> Mr B: You will read 'aloud' and raise your voice?
> Fadel also nods his head.
> Mr B: You sure? Fadel again nods his head. The teacher gives him the chance to read and Fadel starts reading.
> Some students: We do not hear.
> Mr B: You keep quiet.
> Mr B is standing in the front of the class while Fadel's seat is in the back (I suspect that Mr B could not hear Fadel's reading because I could barely hear his voice although he was a shorter distance from me). No other comments are made by the teacher.
> Fadel completes reading one paragraph.
> The teacher asks: Who also wants to read. I want a clear voice.

Mr B offered more opportunities to high achievers compared with opportunities for low achievers. The observation notes in 17 sessions recorded the following.

	Total number of students	Opportunities	Opportunities per student
High achievers	7	13	1.8
Medium achievers	15	21	1.4
Low achievers	7	3	0.4

Mr B's practice in correcting mispronounced words was unclear. Students' competence in reading, for Mr B, was merely a matter of practice and home responsibility.

> Sa'aid? I asked you to keep reading at home. Next time I want better reading.
>
> (To Hossain after difficulties in reading): Did your mother read with you at home?
> Hossain: No.
> Mr B : Your father?
> Hossain: No.
> Mr B (To the whole class): I told you to ask your parents to read with you at home. You have to practice reading.

Reading aloud was used as part of classroom routine. It did not serve diagnosis or remediation. Many students in Mr B's class had phonetic problems when reading. Mr B corrected these errors, but gave no attention to their elimination. He merely advised students to do more practice at home.

Reading aloud in the girls' school (year six) was integrated with other activities, such as reading comprehension. Using a dictionary to find out what words mean was also part of reading sessions. The students' competence in reading seemed to allow the teacher to proceed with other objectives of the curriculum.

In Mrs C' class, total reading opportunities accounted for 26 in 17 sessions. In these sessions, the observation notes revealed that 16 students had no opportunities to read aloud. This task had been entrusted to competent readers. Lower achieving students were ignored.

	Total number of students	Opportunities	Opportunities per student
High achievers	7	11	1.6
Medium achievers	17	14	0.8
Low achievers	8	1	0.1

Of total 26 reading task opportunities, only one opportunity was given to a low achieving student. In contrast, eleven opportunities were shared among the seven competent readers.

Reading aloud was different in Mr E' s class. Students in year 6 of schooling should have reasonable levels of competence in reading. Reading should therefore normally be used to serve other purposes. But the situation in Mr E's class was not like this. Students' reading abilities did not match the standard required by the textbook. They could hardly proceed. The following describes an interaction between Mr E and a student.

> Mr E reads aloud a segment (titled, 'Development of industry in Bahrain') from the textbook. While he reads, he also goes around the class. The students follow from their own textbooks but their attention is inconsistent. Mr E finishes reading.
> Mr E: Who wants to read the first paragraph?
> Several students raise their hands. Mr E gives permission to Hossain.
> Hossain: Bahrain is famous..
> Mr E: Bahrain is..? There is no title?
> Hossain: Industry in..
> Mr E: No. Development (Altanmia).
> Hossain: (Could not pronounce the word) Altanwia.
> Mr E: Altanmia. Who can read it for him?
> Another student: Altanmia.

Hossain: Altanwia.
Some students start to laugh.
Mr E: Altanmia. Follow the letters. Keep your eyes on the word.
Hossain: Altanmia (and he completes reading the title).
Hossain also mispronounced two words in one line.
Mr E corrects the words and Hossain continues reading until he reaches the end of the line. Mr E gives the chance to another student.

Hossain, as with other students, tackles the whole word rather than letters. Students' errors seem to be the same as students in year three. These problems could be remedied earlier. Mr E's interaction pattern with students declared, in effect, that poor reading was each student's responsibility. This may be true, to a certain extent. But it ignores the reality that in the past, critical weaknesses had received no remedial attention.

Classroom work interaction:

This kind of interaction was indirect. Teachers mainly interacted with students' work. Although the work was given to the whole class, individualized attention predominated. The context of the interaction was that the teacher assigned classroom tasks to students. Tasks were given to the whole class with no consideration of individual differences. Each student worked on his/her own. The teacher went around the class to check their understanding, gave quick glances to each student's work, and sometimes stopped beside students to observe what they were doing. The teacher could be approached by students who completed their work early. When the teachers realized that some students had finished, s/he started marking the students' books. Teachers usually interacted with all the students in the class, with some students having more than one interaction. It was not possible to

record each interaction of this type, but the classroom observation shows that marking students' books was a recurrent daily activity for four teachers, Mrs A, Mr B (class teachers), Mrs D and Mr F (the mathematics teachers).

Mrs A gave regular classroom work. She used classroom practice to encourage competition between students. She usually announced the top ten who were first to complete their tasks correctly. Mrs A responded to students' errors in three different ways:

a) Individual help: in this case, she checked students' work, marked it, or pointed out mistakes and asked the student to try again.

b) Group help: such incidents occurred only twice during my observation. It originally happened when a number of students failed to accomplish a particular task and made the same errors. Mrs A gave group instruction to these students. She, furthermore, checked over their understanding by observing each one performing a similar task.

c) Class instruction help: in this type the teacher reacted to the common errors. She addressed the attention of the class as a whole to this issue.

Mr B addressed daily time to classroom work. It is not, however, a coincidence that Mr B allowed time for classroom work at the end of each session. This seemed to give him time to leave the classroom for a few minutes while students worked on their own task. In the classroom, he either went around the class or sat while students queued for correction. Mr B seemed to take no responsibility for showing students their errors nor to reiterate what had already been taught.

(To one student after checking his book):
Did you understand or not?
Student: (shaking his head as saying No)
Mr B: A minute ago, I asked, Who does not understand. Why you did not say? Have you not got a tongue to talk.
(To another student): Hossain, explain to him.

It seems that, in the classroom work, Mr B abdicated his responsibility as a teacher. His interaction with students was probably more a routine than an assessment task.

Mrs D gave students class work to practice in the class, so she could check over their work and assess their understanding. This sometimes took only 5 minutes but at other times was extended to 25 minutes. She gave attention to each student's work and sometimes asked the student for an explanation. Nevertheless, Mrs D did not always persist in checking all or most students' work. The time to finish other classroom work seemed also to be important. She, on the other hand, identified the common errors and explained these to the whole class. Mrs D sometimes stopped individualised correction when she noticed an error repeated among students (e.g. in the long division: dividing the remainder for the second time before bringing down the other number). She then demonstrated to the whole class how the error had been made, and how it could be avoided.

Classroom work was a daily practice in Mr F's class. It took between 25 to 30 minutes of each session. He went around the class to check and mark. Some students made their own way to the teacher. Students can be grouped into two types: those who initiated contact with the teacher (mostly more capable in

accomplishing tasks by themselves) and those with whom the teacher initiated contact (mostly incapable of accomplishing task on their own). The first group had, however, more chances to interact with Mr F than the other group. Teacher interaction, however, took the same form for all students. Students' errors were pointed out to each student and these were corrected didactically rather than explained:

> You see your mistakes, you just have to move the decimal point. You should divide the number by 10, 100 and 1000. You just move the decimal point. In division you move it to the left.

Students' common errors were not explained to the whole class. The teacher seemed more interested in marking as many books as he could and in expecting students to finish as many exercises as they could. The end product was the focus of Mr F's attention.

In Khalid's exercise book, for example, Mr F gave 10 out of 10 for 8 division problems. The final answers in these problems were correct but my observation suggested that they could not have been Khalid's own work. Four problems had not been correctly solved, although the final answers were correct. Khalid seemed to have copied the answer from another student. He also seemed aware that Mr F was not interested in checking his work as much as in checking the final answer. It seems that Mr F encouraged students to work for scores rather than for learning itself.

Conclusion:

The investigation which has been made in this chapter has the overall value of highlighting teachers' differences. Each teacher had different perceptions of their roles and the objectives of the school system. Despite differences however, similarities were also evident.

Two general points can be made. The first is the domination of whole class interaction approaches at the expense of individualized approaches. One of the main concerns of the new policy in primary education in Bahrain is individualization. Teachers should recognize differences in their classroom practice and devote time to those who need their attention. When this condition cannot be met, teachers may use a whole class teaching strategy. But when assessment is under consideration, teachers need to adopt individualized practices. Such approaches were not considered. The same knowledge and skills, regardless of students' achievement levels were assessed. Even when teachers interacted individually with students, they often used individual work for the benefit of the whole class. Such approaches have little value for formative assessment.

The second point is the apparent emphasis on teaching low level cognitive thinking skills. The use of a single textbook may drive teachers to be over-concerned with the transmission of the factual knowledge embodied in the textbook. On the other hand, the predominance of closed procedures also raises further questions about teachers' unwillingness or capacity to ask open questions.

This chapter has tried to probe teacher-pupil interaction, with particular reference to assessment. It seems, nevertheless, that assessment concerns were not integrated in classroom teaching and learning. Diagnostic and formative assessment were almost absent from the classrooms investigated. Assessment, that is, was separated from teaching and learning. Pencil and paper tests remained the teachers' main approach to assessment, and are the subject of the next chapter.

8

Discrepancies in Classroom Testing

Educational testing is haunted by the legacy of psychometric thought which, conventionally assumed that all educational testing can be devoid of subjective bias. More recently, however, psychometric assumptions about objectivity have been criticised as being difficult to realize. All data gathering is contaminated. Its products are always sociocultural phenomena:

> data are social constructs developed through the relationship of researcher, research participants, research context (including its historical antecedents), and the means of data collection (Smagorinsky, 1995, p. 192).

This viewpoint, as Smagorinsky pointed out, can also be applied to teaching and the assessment of learning. Decisions about what attributes and qualities tests should measure are critical in determining the worthiness of the test. The selection of these qualities and attributes to be tested not only entails making subjective

judgement about learning and teaching but also about society's expectations of individuals. Question design and the standardization of testing procedures are also a matter of human selection. Subjectivity intervenes in the organization of items, the relative share of each item in total score, the definition of correct answers and in the formative or summative uses of tests. This subjectivity is heightened, too, by classroom factors: for example: teachers' professional backgrounds, classroom contexts, students' backgrounds, parental expectations and system expectations. Subjectivity in classroom testing cannot be avoided. It interferes with every assessment action that a teacher takes.

> Even if the test is administered by machine, the situation is still a social one in that the making and interpretation of the tests are social acts - they embody value judgements emanating from the dominant culture as to what constitutes evidence of 'intelligence' and what constitutes valid realization of 'educational knowledge' (Broadfoot, 1979a, p. 105).

The above argument, however, does not mean that testing should be abandoned. In terms of efficiency, cost and benefits, tests are expected to persist. It, rather, means that testing cannot be longer perceived as a totally technical objective procedure. Technical issues of reliability and validity of tests need to be reconceptualized. The following discussion addresses these two issues in turn.

Test reliability is concerned with consistency of measurement. Underlying the reliability issue was the scientific assumption that the value attached to something is actually consist of a 'true' value and an 'error'. If the measurement is repeated several times, then the average of these repeated measures would be estimated to

be 'truer' than a one single measure. This assumption is also applied in educational measurement. The concern, therefore, was to ensure that the measurement procedures are reliable by minimizing the 'error' and maximizing the 'true value' of the students' scores.

Reliability in educational measurement focuses on two questions: (1) whether a student's work will gain the same score if it is marked by different assessors; and (2) whether a student will gain the same score if s/he resits the same test. The second type of reliability is not always easy to conduct because it is impractical to ask individuals to repeat the same test several times. Change in individuals also creates difficulty in the repeated measures. Two alternative procedures have become popular. First; conducting two presumably equivalent forms of test and studying the correlation between the two sets of scores. Secondly: giving a single test which consists of several sections and studying the consistency of performance over the sections (Thorndike, 1990).

Reliability becomes important when practical decisions are to be taken in accordance with the test results. Results should be comparable. Decisions should be fair and based on unbiased measures. In social reality, however, repeated measures rarely duplicate each other. In addition, there is the problem when the assessed skill is likely to create various interpretations among assessors such as writing. Even with designing criteria for marking, a qualitative judgement is likely to occur. The aim in educational testing is, however, to achieve a high degree of consistency in the measure, but the degree of consistency is always subject to interference from

the personal equation.

Awareness of the above difficulties and the concern about reliability brought up several development by which the construction and the design of the test are formed. Educational researchers, like Miller and Parlett (1974), suggested that test structure and the designation of test items are critical in determining the reliability of the test. Reliability of marking is related to factors like the complexity of the task, the difficulties of marking and the combining of marks (p. 25). Reliability, as they noted, increases when task complexity and marking difficulties are reduced. More structured tests, therefore, have been developed to increase the inter-rater reliability, such as multiple-choice and standardized tests. These tests involve less judgement on the part of markers. Thus, they increase the degree of agreement among the markers and accordingly, reliability.

It is now widely acknowledged, however, that the attainment and the resulting 'score' are affected by context (e.g. test situation) and the learners' characteristics (e.g. motivation, ability and social backgrounds). Reality is not fixed and not independent but rather constructed. As Gipps (1994), therefore, suggested ' In any case we no longer conceive of 'accurate' measurement and 'true scores''. In effect, the achievement of reliability becomes problematic with the recent different assessment approaches (e.g. criterion- and performance-based assessment) and the variety of contexts involved in them. These new approaches emphasize validity over reliability and they move away from standardized procedures. Criterion-referenced assessment is intended for diagnostic and improvement purposes rather

than for comparability purposes. A key feature in performance assessment, on the other hand, is the subjective element in scoring. The new assessment approaches cannot be forced into the traditional model of reliability. There is need, therefore, to extend the concept of reliability and to include a range of criteria which are acceptable to a particular purposes of assessment.

'Quality assurance' and 'quality control' are two approaches that both aim for enhancing consistency in assessment. Quality assurance is concerned with enhancing the process of assessment while quality control is concerned with the outcomes of assessment. These two processes contribute to both reliability and validity. Harlen (1994) provided a review of these practices in the U.K and other countries. These practices are linked to political and social purposes and have been prompted by the intention to move beyond standardized tests.

Gipps (1994) also suggested alternative criteria for assuring the quality of educational assessment and for enhancing reliability in line with recent changes in assessment. Some of these criteria are consonant with the notion of assessment as a subjective activity. She suggested six criteria:

> • *Curriculum fidelity:* this implies that the construct, domain or curriculum is well specified and there is a broad coverage of the curriculum (if not of each domain) in the assessment.
>
> • *Comparability:* this is achieved through *consistency of approach* to the assessment by teachers; a *common understanding of assessment criteria;* and that performance is evaluated fairly, that is, according to the same rubric by all markers. These can be achieved by a combination of training, moderation and provision of exemplars.

> • *Dependability:* this emerges from evidence of curriculum fidelity, consistency and comparability, as will
>
> • *Public credibility.*
>
> • *Context description:* this requires that detailed information about context be available so that we may make informed judgements about transferability.
>
> • *Equity:* this requires that a range of indicators be used in an assessment programme to offer pupils multiple opportunities to achieve (p. 174).

Reliability, however, cannot be studied in isolation for the purposes of assessment and in isolation from the range of contexts involved for achievement of each purpose. The choice of whether reliability is of crucial importance depends on the purposes of assessment and represents a trade-off between costs and benefits.

On the other side of the same argument is the concept of validity. Test validity is concerned with whether a test measures what it purports to measure. Validation is 'the process of examining the accuracy of a specific prediction or inference made from a test score' (Cronbach, 1971, p. 443). Validity is, therefore, related to how a test would be used. Different uses of tests highlight three types of validity (French and Michael, 1968). The first (content-validity) is used to determine the present behaviour of individuals in a desired performance domain reflected in the test content. The second (criterion-validity) is to predict an individual's standing in some variable - or criterion - which is external to the test. The third (construct validity) is to describe the individual traits or qualities presumed to be reflected in the test performance. Typically, content validity is important in achievement tests where the test content (skills or knowledge) should sample the skills or knowledge

of a specific universe of situations or content intended to be assessed. Criterion validity is important for selection and diagnostic purposes (eg. intelligence and aptitude tests) where the intent is to predict future academic performance. And construct validity is important in personality tests which probe psychological traits such as creativity and attitudes.

Any given test can be intended for a particular use and be constructed around a particular type of validity. But the test may still generate information which is sensitive to other types of validity. For example, achievement tests are usually intended to cover the content domain of the course. But they could also be used to predict how students will perform in more advanced courses. Such tests, therefore, could be responsive to both content and criterion validity. Other information may also be provided about how scores in such tests could be interpreted and judged. Tests, that is, may also provide construct-related information about whether the qualities reported by test scores reflect intended qualities.

Such overlap in the operation of validity studies created difficulties for test designers. They began to call for a unitary concept. Recent commentators have proposed construct validity as the central concept of validity, itself embracing content and criterion validity (Messick, 1980, and Anastasi, 1986). Messick suggested that all considerations of test validity entail some type of construct validity. He based his argument on the concept of validity itself; that is, on the general appropriateness, meaningfulness and usefulness of inferences derived from test scores. These attributes, he argued, are inseparable. Construct validity attends

to the meaningfulness and interpretability of test scores. In turn, this 'construct meaning' provides a 'rational basis both for hypothesizing predictive relationships and for judging content relevance and representativeness' (p. 1015).

Concern about the social consequences - or 'usefulness' - of assessment has also fostered the discussion of an extended view of validity, consequential validity, including, for instance, a symposium at the 1996 conference of the American Educational Research Association held in New York. Particular attention is paid to the construct validity of the test itself and to the social consequences of each particular use of the test (Messick, 1980 and 1988). The key issues in validity, as Messick rephrased them in 1988, are the

> interpretability, relevance and utility of scores, the import or value implications of scores as a basis for action, and the functional worth of scores in terms of social consequences of their use (Messick, 1988, p. 33. see also Wiliam, 1996).

To this end, consideration of consequential validity is based on two questions: Whether the test is a good measure of the characteristics it intends to assess and whether the test can be used for the proposed purpose (Messick, 1980, p. 1012). Consequential validity, therefore, takes account of two facets of assessment, the validity of tests and validity of test use. Both the evidential basis of test interpretation and the consequential basis of the test can be combined to illuminate this concept (see below, taken from Messick, 1988, p. 42).

	Test interpretation	Test Use
Evidential Basis	Construct Validity	Construct Validity + Relevance/Utility
Consequential Basis	Value Implication	Social Consequences

Messick's conception of consequential validity derives from the argument that assessment typically has value-laden social consequences. Thus, with different uses of the same test, the positive and the negative (or side-effect) consequences must be considered in the validation process.

> Justification of test use by an appeal to empirical validity is not enough; the potential social consequences of the testing should also be appraised, not only in terms of what it might entail directly as costs and benefits but also in terms of what it makes more likely as possible side effects (Messick, 1980, p. 1012).

This concept is widely debated. It has not yet been fully accepted, however, that consideration of test consequences could be regarded as an aspect of validity. Popham, for instance, has argued that the concept of validity should refer to test-based inferences, and should be held separate from subsequent decisions: 'valid inferences can be the basis of commendable or reprehensible decisions. Such decisions do not alter the validity of those inferences' (1996, p. 7). He accepts that the social consequences of test use should be considered, but he is reluctant to consider these as an aspect of validity.

Expanding the concept of validity has arisen from internal debates among psychologists (like Messick). But such debates have also drawn the attention of theorists interested in using assessment to enhance teaching and learning. Tension

exists between views of testing concerned with validity and the measurement of reliability and new perceptions of tests where concern is focused on the instructional and educational consequences of test use.

This tension has became evident in many countries, including Bahrain, where testing is intended for different purposes: testing for improvement and/or testing for accountability. Testing for improvement is formative: to assess students' performance, to check what they have mastered, to diagnose their weaknesses and to guide them in their subsequent learning. Testing for accountability, on the other hand, is used to examine general standards of education, to render an account of the differential performance of individuals, classes, schools and school systems. Such interest in accountability for policy management has placed national testing high on the political agenda of many countries (OECD, 1993). The main purpose of accountability testing is summative: to provide management information to local, national and international policy communities.

Thus, these two forms of assessment differ in important respects. Testing for improvement is integrated with teaching and learning, and based on interaction between teachers and learners. It is continuous, diagnostic and supportive. Its results are used for formative purposes. Testing for accountability is based on large-scale testing programmes and is disconnected from teaching and learning. It is summative, concentrates on basic skills and/or core subjects and, more importantly, is claimed to be objective. Its results are used mainly to inform decision makers about the efficiency of the system of schooling.

It has been strongly maintained that an emphasis on testing for political accountability is likely to undermine testing for educational purposes, and that no single form of testing can simultaneously serve both purposes. As Nisbet pointed out:

> Accountability testing does not fit easily with instruction: it is an imposed system, usually relying on multiple-choice machine-scored tests to hold down costs, and too often used to make superficial judgements. New performance measures do not fit with accountability: they are closely tied to work done in individual classrooms and with individual students, are more expensive to operate, and do not readily yield simple scores which can be used normatively to assess trends in standards (OECD, 1993, p. 129).

As suggested, these forms of testing cannot be unified. Nor , indeed, can they coexist comfortably. Each undermines the other. Testing for accountability, for instance, may have notable classroom side effects. Fredericksen (1984) suggested that testing for accountability increases the pressure upon teachers to 'teach for tests' and that it motivates students to prepare for such tests (p.194). Accountability tests, he argued, make demands on a teacher's time and effort. Maximizing time for testing limits the time available for teaching and instruction which, in its turn, leads to concentration on skills which conform to the test.

> A possible consequence is that the abilities that are most easily and economically tested become the ones that are most taught. If educational tests fail to represent the spectrum of knowledge and skills that ought to be taught, they may introduce bias against teaching important skills that are not measured (p. 193).

Tests, therefore, narrow instruction to include only learning that can be tested.

Skills which are not easily tested are devalued. Their importance is reduced. For example, the ability to make good conversation, a speaking attribute, may not be assessed because it requires tests which are presumed to be less reliable.

It is also widely believed that classroom testing has a restraining effect on teaching and learning. Critics suggest that testing which emphasises recall and memorization overlooks higher thinking skills. In a review of research on the influence of testing on students' performance, Fredericksen (1984), acknowledges that testing increases retention and improves performance. But the effects are specific to the test content (p. 197). In another paper about the impact of assessment on students' learning, Crooks (1988) reviewed studies which suggested that frequent testing has a desirable effect on students' performance, and that it is better than no test at all. Crooks, however, includes the comment that more testing 'may not help (and may actually inhibit) higher level outcomes, even when the evaluations focus heavily on these outcomes' (1988, p. 449). Likewise, tests tend to emphasize the product rather than the process of learning. Teachers who use tests look for the right answer, paying less attention to the processes of students' thinking. Such tests provide very little information to guide teaching and learning. Finally, Broadfoot (1979b) claimed that testing fosters an 'instrumental' attitude to education. It has deeply deleterious side effects on student learning. A major source of motivation is negated because 'by comparing their performance with other pupils, [students] quite early come to realise that they are not likely to be successful in terms of external examination performance' (p. 4).

Current changes in the philosophy of educational assessment can be represented as a change from testing for (psychometric) measurement to testing for (educational) improvement. Testing in the latter sense must be integrated with teaching and learning. Furthermore, these changes in assessment philosophy resonate with research on children's thinking as a process of structuring knowledge and skills. These changes also recognise that instruction, assessment and testing affect students' thinking and achievement in schools. According to this view, tests should address learning as well as performance. In the words of a teaching theorist, 'tests should be responsive to the structure and processes that develop as individuals move from beginning to advanced learners' (Glaser, 1991. p. 28).

Classroom testing is the main focus of this chapter. Using fieldwork data, it reviews primary school testing in Bahrain. The chapter consists of four parts: preparation for tests, the testing situation, marking and grading, and the interpretation processes.

Preparation For Tests

Teachers in years 4 to 6 are officially required to conduct two sequences of testing, the mid term test and the end of term test. The mid-term and end-of-term tests are, however, not the only tests practised in primary schools. As shown in table 8. 1, all classes in both field work schools used monthly tests with the exception of Mr B's class. And Mrs A still used monthly tests although the Assessment Act (1983) had

removed this testing requirement for the first three years of primary schooling.

Teacher		October	November		December	January
		(Monthly tests)	Monthly	Mid-term	(Monthly test)	(End of term test)
Year 3	Mrs A	Teacher test	Teacher test		Teacher test	Teacher test
	Mr B	-	-	-	-	School test
Year 6	Mrs C	Teacher test	Teacher test	Teacher test	Teacher test	School test
	Mrs D	Teacher test	Teacher test	Teacher test	Teacher test	School test
	Mr E	Teacher test	Teacher test		Teacher test	School test
	Mr F	Teacher test	Teacher test		Teacher test	School test

Table 8. 1. Tests conducted in the four observed primary classrooms (1995).

It seems from the above table, that female teachers in year six conducted five tests in one term while male teachers conducted four tests. The difference comes as a result of conducting two tests in November in girls' schools, the monthly and mid-term test. This means that female teachers are loaded with more test work in November. The following description focuses on the preparation of the end-of-term test since it is conducted by all teachers and is also the most influential test.

Mrs A is responsible for the whole process of test construction. She sets the questions, organizes the time which suits her students and marks the test papers. One week before the end-of-term test, Mrs A provided revision lessons for her students. In an observed two-hour session, Mrs A revised mathematical concepts using different strategies, whole-class teaching, individualized and group teaching. She used this lesson to identify students' weaknesses and to provide individualized teaching for those who needed it. It appears that Mrs A's attention preferentially focused on getting the right answer. This emphasis was obvious when Mrs A

revealed her pleasure with Farah who gave several wrong answers in this lesson but finally achieved correct answers. It is evident, too, when Mrs A threatened to 'kill' Amna if she gave a wrong answer when asked to convert 700,000 grams into kilograms. Amna was reluctant for few seconds, but she eventually gave the right answer. At the end of this session, Mrs A also asked students to prepare for revision in science for the next day.

Unlike Mrs A, Mr B had abandoned monthly tests entirely. The learning milieu in Mr B's class was normal until the last day before the end-of-term test. The school, however, conducted end-of-term Arabic and mathematics tests for students in years 1 to 3. All class-teachers contributed test items. This test, combined with teacher assessment, is used at the beginning of the following academic year to group students into mixed ability classes (see chapter 6). Thus, although Mr B contributes items for this test, it is actually a school test conducted for reporting and decision-making purposes.

In Mrs C's class, the teacher seemed to prepare students for the test content during the term. Mrs C highlighted some important subjects and skills that might come in the test.

> This is an important question (What are the major achievements of the Gulf States Co-operation Council?). You must copy the answer in your book and remember it.

Mrs C also pointed to what might be asked in the tests:

> Girls! Be aware! In either topic (the poems), we are going to ask you about the basic idea behind the poem.

In her class, there were several revision sessions before the end-of-term test. In one lesson I observed, she went through several topics in the syllabus, using a question-answer strategy. Her questions covered: comprehension, meaning of words, forming sentences out of new words, reciting factual knowledge, reciting some poems and some grammatical rules. All students in the class had a chance to participate in answering her questions.

Mrs D, the year-6 mathematics teacher, organized practice lessons as a preparation for the end-of-term test. During two weeks before the test, lessons were devoted to practising topics that covered the whole syllabus. In these lessons, students were asked, by turn, to do a task on the board. During the term, Mrs D also drew students' attention to geometry as an easy subject. She regularly reminded students that all of them could meet the requirement of the subject.

> Geometry is an easy subject. Try to get full marks in this topic. It will increase your total score.

Immediately before the test, Mrs C advised her students to

> practice with the questions. Set a test for yourself and try to work all the problems out. Check over your answer and see where you make errors. With such practice you definitely will perform well in the exam.

The classroom practices of Mrs C and Mrs D reveal the relationship between testing and teaching. Although both teachers covered the prescribed syllabus by the end of the term, particular skills and topics were emphasized. The teachers' style of revision helped students to recognize the whole subjects taught during the term, refreshed their memory and identified what they might revise. On the other hand,

this revision strategy encouraged students to cram for the test and thus discriminated against those students who could not follow the teachers' recommendations.

Mr E, the year-6 Arabic teacher, had a different approach. He distributed a piece of paper which contained 10 definitions and ten areas of information; and pointed out that 50 percent of what was included in this handout was going to be included in the test. No revision lessons had been organised until the last day before the test. The last lesson concentrated on some grammatical concepts, as a new topic.

Mr F, the year-6 mathematics teacher, started revision one week before the test. He distributed blank worksheets to each student and asked them to write down all the problems tackled in the following days and to use it as a revision book.

> From today on, write down all the problems in these papers. Some of what we are going to practise these days will come in the test. Try to practise them at home. If you do not understand ask for assistance from any one in your family.

Mr E and Mr F tried to help students concentrate only on knowledge and skills expected to come in the test. What seemed to be important for both teachers was to find a way to help students pass the test. Mr E's and Mr F's perceptions of students' ability appear to be associated with their procedure. During the term, they gave little attention to the forthcoming tests. They seemed to treat it as a routine task rather than as a device to improve students' performance or as a mechanism for reporting (as will be demonstrated later in this chapter). A significant feature of Mr E and Mr F was their last minute attempts to cram something for the test. They

seemed at this stage, to reward students for rote memorization rather than for understanding.

The Testing Situation

The mid-term test and the-end-of term test period are both determined by the Ministry of Education. The end-of-term test started in 22nd of January 1995 and lasted for four days. The timetable was organized by the school and was known in advance by all students in the school, as well as their families.

The year three and the year six tests are organized differently. In year 3, the class-teacher attends the testing situation. Students can consult her/him at any point. The testing situation in year six is more bureaucratic. Teachers attend the class only to read the questions for students and to clarify any ambiguous items. Teachers without previous knowledge of the students are usually selected as invigilators.

Mrs A's test procedures were a combination of formal and informal. Over a period of two weeks, she conducted tests in the five subjects she teaches, allowing a 2-hour period for each subject and allowing for revision and correction. The Arabic test is organized in three parts: listening and dictation, free writing and grammatical rules and linguistic skills. Mrs A marked the papers at the same time. Students started to hand in their papers during the first hour, and Mrs A started to mark the papers. The second hour of the test situation became a combination of testing, marking, identifying errors and giving students further opportunities to

correct their answers. Mrs A called out each student to discuss errors. Students were allowed time to discover what kind of errors had been made and then were given the chance to correct answers. In such a situation, testing is more personalized and used for formative purposes. Students' performance is, thus, open to improvement, enhancing the consequential validity of Mrs A's procedures.

Mr B contrasts with Mrs A. His class experienced only one testing situation: the end-of-term tests conducted by the school in Arabic and mathematics. The testing situation of Arabic was more formal in Mr B's class, though he attended the test himself. Marking was undertaken after the test. Students, therefore, had no chance to improve their performance nor become aware of their results until the start of the following term.

In year six, two subjects are tested each day. Each paper, usually written work, takes between one and a half to two hours. Some attributes, such as listening and dictation in Arabic and English, are tested in the previous week.

Gender differences among the year-6 students were obvious during the field work. In the girls' schools, the forthcoming end-of-term test created a new classroom atmosphere. Desks and seats were reorganized by students in separate rows. Students were engaged in revising individually, in pairs or in groups. Most students appeared to be engaged in memorization and recalling activities. This apparent engagement with memorizing from the test book seemed to be common among all girls in the class, a side effect of the testing. This impression was supported by

another incident. The day after the mathematics test, I went to Mrs D's class and asked for some students' course books. This was not possible as all students had thrown their books away. It appears that the end of term test marked the end of school learning.

In contrast, the boys' class seemed to act differently at testing time. Apart from three students sitting quietly in their desks the other students were in the yard. Desks and seats remained in their normal places. As the bell rang at 8 o'clock, students began to enter the class laughing and joking with no books or bags in their hands. There was no sign of anxiety or tension. Their behaviour suggested that the test was not taken seriously by students. When they had settled down, the teacher asked them to separate their desks. They needed five minutes to form the desks into five rows.

Differences between girls and boys are also recognized in sitting the test. During the test, all girls engaged individually with their task, and no attempt was made to consult other students. Mrs C attended the class for a short period of time. She went around the students, checked what they were doing and pointed to some answers that needed consideration. Mrs D, also visited her class, but she mainly drew students' attention to some printing errors. She asked low achieving students whether they needed any explanation. Students who finished earlier sat quietly in their desks or revised for the next test while waiting for the whole class to finish.

In the boys' school, students seemed puzzled. Faces were looking around for

assistance. A few students were looking at their papers without trying to write answers. Those who had been previously labelled as low achievers handed in their papers early. Students who finished early could leave the class. These students stayed in the corridor outside the classroom and appeared to disturb students who remained in the class (eg. by waiting by the door and trying to talk to them as soon as the door was opened).

These different behaviours reflected different attitudes and motivation among boys and girls. What is apparent in the two cases examined is that students' motivation seems to be related to two factors: students' actual competency and the assessment practices. There was *prima facie* evidence (indicated below) that the boys' competence was far behind the required syllabus. A wide gap appeared to exist between the performance of the majority of students and the optimal required objectives. It appears that many boys recognized this fact. Tasks included in classroom tests did not match their actual competence as much as it matched the goals of the syllabus which, for them, were unattainable. The importance of matching tasks to students' competency levels has been highly acknowledged as a prime factor for motivation. These male students had already experienced repeated failure in their classroom work and in the formal tests. Their self-esteem was low, as they anticipated that their achievement in this test would be no better than in previous ones. Their motivation to learn was weak.

In the girls' school, the field work revealed that the majority of girls had high achievement scores in their previous tests. Classroom observation also revealed that

their high scores are consistent with their actual performance. The goals set out in the syllabus were thus attainable. Competence provided the girls with intrinsic motivation to increase or maintain their performance standards.

A further factor which fostered the boys' negative attitude towards tests was related to the teachers' previous practices. The impact of assessment practices on students' motivation to learn is quite relevant. Low scores in previous tests, as observed, had been altered and modified. Students knew that their scores would undergo similar modification which would allow them to pass. Thus, their performance in such tests was not going to affect their result. They realized that their performance in the test and their test results are two different things. Students, in this process, devalue the examination system as a whole, an effect also noted by Jackson (1968) in his study of U.S elementary schools.

> [each student devalues] the evaluations to a point where they no longer matter very much...He is neither elated by success nor deflated by failure. He may indeed try to 'stay out of trouble' in the classroom and thus comply with the teacher's minimal expectations, but this is principally because getting into trouble entails further entanglements and involvement with school officials and other adults, a situation that he would prefer to avoid (p. 27).

Scores, therefore, have lost their value as extrinsic rewards. They no longer encourage students to make any effort to improve their performance.

Grading and Marking

This section investigates four aspects of the grading process: the grading system, format of the end-of-term test, constructs tested by the test items and the teacher's marking procedures.

Grading system:

Grading is the teacher's summative response. It is the feedback that the teacher provides. Grades, therefore, give a letter or numerical summary of students' performance in a particular area of knowledge or skill. They come in different forms: numerical scores, percentages, letter grades and pass/fail. Grades can be useful as summative statements about students' performance levels which, as noted earlier, facilitates comparison of students, schools and districts. They can also be regarded as extrinsic rewards, motivating students to learn and to improve their performance.

A student's final score is a combination of mid-term test, end of term test and classroom work. This combination is mandatory and should be applied by all primary schools for years 4 to 6. Coursework should contribute 30 percent, mid-term 20 percent and end of term 50 percent. Scores allocated for coursework, however, are still derived from monthly tests. Thus, monthly tests in October, November and December contribute 30 percent as illustrated in the example in table 8. 2.

Monthly scores for student Z			Mid term test (November)	End of term test (January)
October	November	December		
57	60	79	61	85
Average	65		(20%)	(50%)
(30%)	**(65 x 0.3) = 19.5**		**(61 x 0.2) = 12**	**(85 x 0.5) = 42.5**
Final score: 74				

Table 8.2. Calculating the total score for student Z

This type of combination is not practical for improving students' learning. Total scores overlook the criteria against which the students' achievement has been tested. Total scores, therefore, can only serves administrative purposes. The combination above has benefit for students in that it may encourage them to enhance their performance. But it has less value for students who may get a mixture of low scores and high scores. In the example above, although student Z's scores indicate improvement in performance, the low scores of October and November lower the student's final scores from 85 (the end-of-term score) to 74. Nevertheless, the effect of such reduction is still minimal since passing the grade level is set at only 50%. That is, student progress is not critically affected by such a process. Yet, this low level may have an undesirable effect on students' motivation to improve their performance, and students' efforts to improve are overlooked.

Format of the end-of-term test:

This aspect reveals the complexity of the task, recognized by Miller and Parlett (1974), as an important aspect to be considered in the study of reliability. Table 8. 3 and table 8. 4 reveal the format used in both year 3 and year 6 in the Arabic and

mathematics end-of-term tests respectively.

Year/ sex	Format	Multiple choice	Matching	Arrange ideas	One word	Short answer	Brief answer	Extended essay	Dictation	Hand writing	Total
Year-3 Girls	Items	-	-	-	24	23	-	1	1	1	50
	Scores	-	-	-	43 (43%)	33 (33%)	-	10 (10%)	1 (10%)	4 (4%)	100 (100%)
Year-3 Boys	Items	5	6	1	12	14	-	-	1	-	39
	Scores	5 (8.3%)	3 (5%)	5 (8.3%)	16 (26%)	23 (38.4%)	-	-	8 (13.3%)	-	60 (100%)
Year 6 Girls	Items	2	-	-	-	38	-	1	1	1	43
	Scores	4 (4%)	-	-	-	56 (56%)	-	15 (15%)	15 (15%)	10 (10%)	100 (100%)
Year-6 Boys	Items	3	-	-	-	39	2	-	1	-	45
	Scores	4 (4 7%)	-	-	-	63 (74%)	11 (13%)	-	7 (8.3%)	-	85 (100%)

Table 8 3 Items format in Arabic in year 3 and year 6 (1995)

Year/sex	Format	Multiple choice	False/ true	Short answer	Short task	Problem solving (one step)	Problem solving (two steps)	Total
Year 3 Girls	Items	6	-	54	5	2	-	67
	Scores	6 (6%)	-	81 (81%)	9 (9%)	4 (4%)	-	100 (100%)
Year 3 Boys	Items	3	-	23	4	2	-	32
	Scores	3 (5%)	-	42 (70%)	8 (13.3%)	7 (11.7%)	-	60 (100%)
Year 6 Girls	Items	10	5	-	5	1	1	22
	Scores	17 (34%)	5 (10%)	-	22 (44%)	2 (4%)	4 (8%)	50 (100%)
Year 6 Boys	Items	-	-	7	8	-	1	16
	Scores	-	-	21 (42%)	25 (50%)	-	4 (8%)	50 (100%)

Table 8 4 Items format in mathematics in year 3 and year 6 (1995).

Both table 8.3 and table 8.4 reveal that short answer and short task items prevailed in all tests. This type of question requires very little effort from students. The teacher is the one who does the most work while students have little space to demonstrate their understanding. The format minimizes the task complexity for

teacher's marking procedures, as it specifies the correct answer and the related score for each single item.

Writing an extended essay in Arabic had been abandoned in the boys' school. In year 3, the Arabic test contains more objective items. In year 6, students had only a brief essay task which did not exceed 4 lines. On the other hand, the test in the girls' school contained an extended essay, but the scores allocated to this task formed only a small proportion of the total score (10% in year 3 and 15% in year six). Essay items may be inconvenient for teacher assessment but they have the merit of assessing comprehensive linguistic skills. Students may demonstrate good performance in the short answer items which only assess isolated skills but these oversimplify the language that students use in real life. Competency in language use can only be assessed when comprehensive communication skills are considered, not just the fragmentary ones.

The case above can also be applied in the mathematics test where very little attention had been given to the problem solving tasks. All the problem solving tasks were straightforward so students could easily identify the problem and find the solution. Moreover, scores allocated to problem solving were not sufficiently large to affect the total score that students achieved. In both cases, the Arabic and the mathematics tests, the students' total score is not sensitive to their competence in the broader integrated subject skills.

The conclusion one can draw from the above two tables is that the consistency in

marking and grading such tests is likely to be high. It may serve objectivity. But, at the same time, it provides spurious success for students and for the system because the tests do not allow for the high quality skills. The value of such tests for learning may thus be limited, and their impact could be deleterious.

Constructs of the test items:

Teachers' assessment in the two primary schools was basically about school achievement. Thus, the most visible feature of the tests items is that they are curriculum-related. Tests are constructed on the basis of the main elements within each subject, as shown in table 8.5 and 8.6.

Year sex	Elements	Listening	Writing	Reading	Grammar	Linguistics skills	Factual knowledge	Dictation	Word recognition	Hand writing	Total
Year 3 Girls	Items	5	13	-	-	16	-	1	14	1	50
	Scores	10 (10%)	19 (19%)	-	-	43 (43%)	-	10 (10%)	14 (14%)	4 (4%)	100 (100%)
Year 3 Boys	Items	-	2	10	-	21	-	1	5	-	39
	Scores	-	4 (6 7%)	18 (30%)	-	25 (41.7%)	-	8 (13.3%)	5 (8.3%)	-	60 (100%)
Year 6 Girls	Items	9	4	2	10	7	8	1	1	1	43
	Scores	15 (15%)	18 (18%)	4 (4%)	10 (10%)	5 (5%)	21 (21%)	15 (15%)	2 (2%)	10 (10%)	100 (100%)
Year 6 Boys	Items	8	9	4	14	3	6	1	-	-	45
	Scores	8 (9 5%)	18 (21.2%)	10 (11.8%)	20 (23.5%)	6 (7%)	16 (18.8%)	7 (8.2%)	-	-	85 (100%)

Table 8 5 Tests elements of Arabic in year 3 and year 6 (1995).

Year / Sex	Elements	Numbers	Measurements	Geometry	Algebra	Total
Year 3 Girls	Items	40	12	15	-	67
	Scores	73 (73%)	12 (12%)	15 (15%)	-	100
Year 3 Boys	Items	26	2	4	-	32
	Scores	46 (76.7%)	6 (10%)	8 (13.3)	-	60 (100%)
Year 6 Girls	Items	5	7	3	7	22
	Scores	12 (24%)	11 (22%)	14 (28%)	13 (26%)	50 (100%)
Year 6 Boys	Items	13	1	1	1	16
	Scores	33 (66%)	4 (8%)	7 (14%)	6 (12%)	50 (100%)

Table 8 6. Test elements of mathematics in year 3 and year 6 (1995).

Mrs A's Arabic test covered 6 elements but it gave precedence to linguistic skills, writing and word recognition. There was no test of reading comprehension quality. The field work, however, revealed that reading still comprised a major part of classroom activities (Chapter 7).

Mr B's test covered 5 elements which also gave precedence to linguistic skills. The test included reading comprehension which entailed reading a short story of about 75 words and accordingly answering 9 questions, of which five were multiple-choice. On the other hand, listening quality was not tested. Omitting this quality reduces the content validity of the school test of Mr B which was mainly undertaken for administrative purposes.

Testing elements in year 6 increased to 9 for Mrs A and to 7 for Mr E. It seems that there was an emphasis on writing and grammar skills. Recalling some factual

knowledge was also tested which is unusual in testing language.

The mathematics test was based on three elements in year three and four in year 6. In year 3, both girls' and boys' tests emphasized number work which had also content-related validity (see appendix 1a). Tests conducted in the girls' school differs from that conducted in the boys' school. Although both covered four elements, the test in the boys' school emphasized number work. This element had 66% of the total mark as opposed to only 24% in the girls' test. Boys' tests included items which test only isolated skills, as in the following example.

> Omit the zero which has no value in the following:
> 7.00, 700, 15.340, 15.204, 100.270, 0.007

The boys' school test might have corresponded with the teacher's perception of the students' competence in mathematics. Mr E knew that many students in the class were deficient in mathematical skills. The test items, therefore, needed to be attainable. Emphasizing number work and selecting simple items which test single skills was a decision, I think, taken by the teacher to yield good scores.

The teachers' marking procedures:

In the two classrooms observed, teachers marked the scripts themselves, registered the final scores in their marking books and sent them to the administration for inspection and confirmation. The marked scripts, not checked by another marker, remained with the teacher until the end of the academic year.

Examination of the marked scripts reveals that the main responses of teachers came

in the form of marks rather than words. Papers were either full of ticks or crosses. Teachers did not offer any written comments on students' papers, or define students' errors. Marking, essentially, referred to correct procedures as well as correct answers. For example, if answering an item required a series of intermediate processes, marks were given for each correct operation, even if the student did not achieve the right answer. To some extent, this means that the student's total score is not representative of how many correct answers they obtained. It was, rather, an indication of the number of correctly performed skills. This case was not applied for the essay type task as a single overall mark was given.

It appears that there are two stages of marking: the scoring of students' scripts, and reporting. In the first stage, teachers marked students' responses to each item, gave scores for each correct answer as well as for each correct stage in reaching the correct answer and, finally, teachers combined scores to arrive at raw totals. In the second stage, teachers recorded students' raw scores or moderated them in the process of aggregating classwork, mid-term and end-of-term marks.

Scores, however, have consequent educational impact when teachers interpret and act upon them. Interpretation begins when teachers describe, moderate and compare results with specified criteria.

Interpretation Processes

Interpretation requires teachers to base their assessment on a criterion against which they compare students' work. Two main criteria are used, norm-referenced and criterion-referenced. In norm-referenced tests, teachers compare students against each other. How well the student does depends on how well other students do. Norm-referenced testing is related to the psychometric bell-curve distribution. The consequence of assessment is a set of distributed students' scores so that a few get high marks, a few get low marks and the majority get the medium marks.

In criterion-referenced testing students' performance is compared against a predetermined criterion. It usually assesses how well students demonstrate the desired quality. The main difference between norm-referenced and criterion-referenced tests, as Glaser (1963) pointed out, is that criterion-referencing depends upon an absolute standard of quality, norm-referencing depends on a relative standard (p. 519). Criterion-referenced testing marked the beginning of relating the assessment to curriculum and instruction and, thus, moved away from bureaucratic to educational measurement. It identifies the standard of performance on which teachers can focus their attention and help students to achieve. It requires the teacher to direct all students towards a full mastery of performance which, in its turn, conforms with formative and diagnostic assessment. This standard is external and not related to the other students' achievements.

But criterion-referenced procedures may not in themselves promote high standards

of performance. Two factors are related to the way that criterion-referenced testing has an impact on teaching and learning: the stake of the test and the standard of accepted performance. Stakes are related to the consequences of the test and their impact on students' future as well as to teachers' accountability (Gipps, 1994, p.33). High stakes tests with higher standards have a greater impact on teaching and learning, than low stake tests with lower standards. Gipps indicated how the low impact of low-stakes testing affects teachers' perception of students and student placement decisions: 'teachers, students and parents do not perceive performance in these tests as being directly linked to reward' (p. 34).

Both norm-referenced and criterion-referenced practices, however, may be applied to the same test (Rowntree, 1987; and Gipps, 1994). The key issue here is how teachers interpret the test results and how they use them. Thus, it depends on whether the teacher derives summative results and uses these results to rank students in order, or whether the teacher provides formative feedback to students on their performance for each required skill.

In this study, the process of interpreting students' scores differs between the end-of-term test and the other tests, and between female and male teachers. Both norms and criteria are used in interpreting students' scores in monthly and mid-term tests. The two criteria are used and are expressed in broader terms. For example, students are given a single score in each broader construct (reading, writing, listening and speaking) and a total score for the whole subject (Arabic) which is norm-referenced. As has been recognized earlier, assigning a total score to students'

performance is the ultimate outcome of tests in Bahrain primary schools. Thus students' scores express their relative position in the class, and also reflect their performance level in the broader elements of the entire subject. But, on the other hand, the 100 score or the excellent grade is a combination of several constructs as illustrated in table 8.7.

Mrs C (Arabic)		Mr E (Arabic)		Mrs D (Mathematics)		Mr F (Mathematics)	
Speaking	5	understanding	15	Classroom activities	40	concepts and information	30
Reading (silent and aloud)	10	Expression	20	Tests	60	Operations	40
Reading comprehension	25	Reading	25			Problem solving	30
Writing	15	Writing	25				
Dictation	15						
Grammar	15						
Listening	15						
Total	100		100		100		100

Table 8.7. The allocation of total score (100) in year 6 between the different constructs of Arabic and mathematics.

Teachers in year six reported students' results according to the above criteria. With Mrs A and Mr B (not indicated in table 7), total scores do not correspond to any combination of the above criteria. And no other criterion is clear in their reports. Mrs C uses a more comprehensive set of criteria than Mr E. She assesses students' learning in seven constructs, allocating different weights to each. The reading construct seems to be more crucial in deciding students' achievement. Mr E uses both curriculum criteria and classroom work for interpreting scores. Mr E's combination is less clear than Mrs C's combination. Other constructs are also missing such as listening and dictation.

Mrs D's scores' distribution is not related to curriculum constructs but to the type of assessment she makes. She also allocates a higher proportion to classroom work than the official assessment had allowed. Although this combination may encourage students to put more effort in the classroom than they do for testing, its effect could not be significant since the passing score is only 40%. Mr F uses the curriculum criteria as officially required.

The criteria used are impractical in terms of two aspects: the domain of the criteria and assigning a score to this domain. Teachers referred to the broader domain of the curriculum and neglected the more descriptive skills which formed part of the curriculum framework (see appendix 1). Attention to one broad domain could only inform teachers of the general standards of performance in broader skills. In this sense, teachers cannot derive any inferences from test scores except to judge the overall standards of the class. As a consequence, they are unable to use these test results for diagnostic and formative purposes.

On the other hand, representing these broad domains in scores is inapplicable for formative assessment. Scores become norm-referenced, although teachers relate them to the curriculum criteria. In criterion-referenced tests where the main purpose is formative, grades are less useful and probably have disadvantages. In criterion-referenced assessment, the maximum mark would reflect that a student had mastered the required skill. The lower mark reflects that a student has not yet reached this level. To serve this purpose, grades should be related to one particular attribute at a time. This requirement was not met in the above tests. Although

scores were allocated to each single item in the test, they were aggregated and represented in totals. Once this is done, the potential for interpreting grades for formative purposes is lost.

This result is recognized when teachers prepared the results of the first month's test (October). The test scores were interpreted against the accepted standard of passing which is 50%. In the girls' school, students in Mrs A's and Mrs C's classes had passed the minimum level. The standard of performance was very high and students' raw scores in each test were reported. In Mrs D's class, four students got less than 40%. The standard reached by the class as a whole was still high. Mrs D's reaction to the test results was practical. She devoted one lesson to the test. She concentrated on showing students their errors and how to correct them.

The case in the boys' school, where a high number of students failed to approach the required standard of performance, was different. The distribution of students' scores in October for Mr E and Mr F suggested low levels of performance. Table 8.8 shows the distribution of students' scores in this month (the total score devoted to this test is 30).

Scores distribution (Total raw score is 30	Mr E (Arabic)	Mr F (Mathematics)
26 - 30	1	13
20 - 25	2	3
16 - 19	5	3
10 - 15	7	4
6 - 9	3	4
1 - 5	10	5
zero	6	2
Total students	34	34

Table 8.8. Year-six boys' raw scores in Arabic and mathematics in October 1994.

Mr E informed students that their results were not good but that he hoped they could improve their scores in classwork. But students had no chance to see their papers. Mr F, on the other hand, returned the scripts to students so, as he told them, they could see their errors. He explained how the total mark was distributed.

> The answer to question one is 215. Every correct step has a single score.... Score for problem solving is four, one for explanation, one for the method, one for the correct answer and one for the unit.

Mr F retained the scripts, allowing students only five minutes for checking over their results. He pointed to those students who had not reached the pass score as failures and needing to improve their marks by enhanced classroom work.

It seems that raising scores was more important for Mr E and Mr F than improving real performance. The fact that many students had failed in the first month's test might warn teachers of serious problems. Students' performance in this test could illustrate their weaknesses. Nevertheless these teachers did not seem interested in investigating any further. Drawing upon the field work observations, teachers' perception of students played a part in preventing them from embarking upon identifying the causes of poor performance and looking for ways to solve this problem.

What seemed to be more important was another procedure. The standard for passing has been established at a very low level (50%). As the standard is very low, students who could not meet the standard come to represent a threat to the teachers' position. Teachers' awareness of their accountability in terms of the

number of students who reach the standard, meant that they were at risk of censure. Accordingly, scores were moderated, as seen in table 8.9.

Scores distribution (total score is 100)	Mr E (Arabic)	Mr F (Mathematics)
90 - 100	10	11
80 - 89	6	6
70 - 79	2	6
60 - 69	6	2
50 - 59	7	7
less than 50	3	2
Total number of students	34	34

Table 8. 9. Year-six boys reported scores in Arabic and mathematics in October 1994.

Comparing table 8.8 and table 8.9 reveals differences in the standard of the class. The teacher's report shows that out of 34 students only 3 students in the Arabic test and 2 in the mathematics test had not approached the required level of performance. Such levels of failures are not questioned by headteachers or the Ministry of Education. On the other hand, those in Mr E's report who had the highest scores represented nearly one third whereas only one student in his class had actually achieved such a score in the real test result.

This form of moderation occurred with nearly every test taken in the classes observed in the fieldwork. By the end of term, the differences between students' achievement in test and their reported results become more obvious. Tables 8.10 and 8.11 shows differences between students' raw scores in the end-of-term test and their final scores officially reported for parents and the Ministry of Education

in both girls' and boys' school.

Subject/ sex	Mrs A				Mr B			
	Arabic test		Mathematics test		Arabic test		Mathematics test	
Scores	Raw scores	Final scores	Raw scores	Final scores	Raw scores	Final scores	Raw scores	Final Scores
90 - 100	23	26	24	19	10	12	7	12
80 - 89	4	2	5	11	3	4	5	5
70 - 79	2	1	1	-	1	8	4	7
60 - 69	1	1	-	-	5	2	3	1
50 - 59	-	-	-	-	3	1	3	1
less than 50	-	-	-	-	7	2	7	3
Total number of students	30	30	30	30	29	29	29	29

Table 8.10. Comparison between raw scores and final scores of year 3 in the end of term test (January 1995)

	Distribution of students raw scores in the end of term test				Distribution of students scores in the official reports			
Scores	Mrs C	Mrs D	Mr E	Mr F	Mrs C	Mrs D	Mr E	Mr F
90 - 100	20	18	1	6	20	19	10	11
80 - 89	7	8	1	3	6	7	6	4
70 - 79	3	2	1	2	2	1	2	7
60 - 69	2	1	-	3	4	2	5	4
50 - 59	-	3	3	4	-	2	8	3
less than 50	-	-	28	16	-	1	3	5
Total number of students	32	32	34	34	32	32	34	34

Table 8. 11 Comparison between raw scores and final scores of year 6 in the end of term test (January 1995).

In girls' classes, the distribution of students' scores in the end of term test and their final official results reveal no differences. Most students get the highest grade

(excellent) in both subjects. There was only one student in year 6 who was officially reported to have failed in mathematics. This however would not prevent her being promoted to the next grade providing she passed in Arabic (see chapter 4). This result is also consonant with the classroom observation that most students appear to be competent in the required skills.

In Mr B' class, it seemed that few students (only 6 in Arabic and 7 in mathematics) had average scores (60 to 79). Most students got either high or low raw scores. This result was moderated to reduce the number of low scores. Fewer students remained in the threshold and fail categories. In the field work, Mr B suggested that all students in the class should be promoted regardless of their achievement level. He explained that the class size should be kept fixed at 30 students and repetition raises difficulties in keeping with this policy. Thus, the two students who got less than 50 marks in Arabic had passed by the end of academic year 1994/95.

In year six, the majority of students did not perform successfully in the end of term test. A wide gap opened up between the minority who got high scores and the majority who got low scores. But this gap disappeared in the final report. Scores which were recorded in students' final reports, were pulled closer to each other, so differences were minimized. This process is concealed in the moderation and aggregation processes. Students seemed to be aware of this fact since they recognized that their performance in the tests was low. This result explains why students did not display any anxiety in their test practices.

Conclusion:

Tests in primary school must have an educational justification. Officially, the results are used in promoting students within the system. The criteria for promotion are imposed by the Ministry of Education. These requirements allow the policy makers to judge the system of schooling and compare schools. Tests, in this respect, are used as a device for administrative purposes, and to complement the conventional inspection of schools. The hypothesis is that these tests are valid and reliable and thus, they are good measures of school through-put and of students' actual performance. The result of this perception is that they focus on outcomes, not processes. But the reality, as seen above, is that testing has become a ritual, emphasizing low level skills and isolated qualities, and indicating superficial success rather than meaningful learning.

In the girls' school, as demonstrated, the standard as recorded in the official reports corresponded with the students' actual performance and also to their actual test results. In the boys' school, it appears that there were substantial inconsistencies between the reports and actual achievement. The result is that many students had been promoted regardless of their actual achievement. These inconsistencies may not be significant in year 3, but their existence indicates an emerging problem. Learning in earlier years of schooling is critical for students' subsequent learning. Students who are pushed into the system with deficient writing, reading and arithmetic may not benefit from any type of education in their higher grades. Teachers in years 4 to 6 of primary schools are more committed to the official syllabus than to providing remedial education. Tests in the boys' schools were used

to cover the apparent weakness in performance. Such tests neither help students' learning nor motivate them to improve. Their weaknesses build up and school learning becomes alien to them.

This chapter raises doubts about the value of testing for improving learning. Assessment practices remain technical, summative and separated from teaching and learning processes. The quality of education is undermined. Discrepancies exist between educational philosophy, the policy of the 1990s and classroom assessment practices. The next chapter highlights these discrepancies by re-examining some issues of policy and practice in Bahrain primary schools.

9

From Practices to Policies of Assessment

Classroom observation reveals differences among the teachers studied in the way they perceive and practice their role, and in the way they interact with students. But teachers and students are also members of a wider society. Thus, what happens in the classroom is also a reflection of their society's values, constraints and expectations. Despite differences, therefore, classroom similarities also exist. In Bahrain society, a common heritage of social life and values interacts with a centralized educational system to foster similarities in classroom life. Examination of classroom practices, therefore, offers a window on the educational system as a whole.

This chapter reflects upon the assessment practices which have been reported in previous chapters. It discusses them in the broader context of the policies of the Bahraini educational system. The chapter focuses on three areas of policy concern:

the nature of achievement, the methods of assessment and the purposes of assessment. That is, it addresses issues related to the questions: what, how and why assess? The final part of this chapter suggests further actions which might be taken in the light of the evidence of this research.

The Nature of Achievement

To consider the nature of achievement is valuable. The nature of achievement that schools use to differentiate among pupils underwrites the success and failure of individuals not just in school but also in their future life. It also reflects the knowledge and skills valued by the system. That is, our perceptions of, and our expectations for, the recipients of formal schooling can all be discerned in the processes of assessment.

Achievement in Bahraini primary schools is mainly specified and controlled through a centralized curriculum. Despite new thinking which favours more flexible, child-centred teaching methods, a centrally planned curriculum continues to dominate. Teachers are expected to maintain a commitment to the prescribed content of common textbooks. There is little room allowed for alterations or changes.

Textbooks steer school assessment. They are officially regarded as important tools not only for 'specifying the content and serving the process of learning', but also

because each textbook 'regulates the instruction and the evaluation of its outcomes' (Directorate of Curriculum, 1992, p.22). Thus, textbooks serve to provide common content, to guide instruction and teaching methods, to facilitate the assessment of learning and to guide the instruction process. Textbooks, in this sense, work to maintain a unified culture and to control the system of education as a whole.

To a great extent, what teachers value in students' achievement, therefore, reflects the common curriculum prescribed in textbooks by the Ministry of Education. But, just as important, achievement is also shaped by the assessments used in Bahraini classrooms. Three aspects of assessment are implicated in this latter shaping process; namely, the differences between medium term and short term objectives, cognitive and non-cognitive objectives and, finally, high level and low level skills.

Medium term and short term objectives:

The current provision of primary education in Bahrain has two turning points where careful assessment is expected. Students are transferred in year 3 from the first stage of basic schooling to the second stage and, in year 6, students are transferred from the second to the third stage. The national curriculum document (Directorate of Curriculum, 1992)[1] proposes that assessment should take into account the medium-term objectives at the end of each stage. It emphasizes that achieving these objectives for each stage is a necessary condition for progression to the next stage.

[1] This document was prepared to guide central curriculum policy and planning. Teachers actually know little of its detail.

As far as teachers are obliged to achieve and assess textbook content, a gap may exist, therefore, between the medium term objectives and the short-term (instructional) ones. Thus, achieving short-term objectives may not always lead automatically to the accomplishment of medium-term objectives. The assessment of the latter requires another arrangement which seems absent at present. As noticed in the field work, teachers generally anticipate immediate outcomes of learning. They stress achieving the prescribed content in textbooks. Although the fieldwork provided evidence that some teachers, often female, considered aspects of objectives which render long-term advantage, all teachers appeared more conscious of their assessment of short-term objectives. For example, long-term objectives such as developing intellectual attitudes towards problem solving, and managing independent learning were all emphasized by teachers to achieve a good result in the short-term. Whether these processes were embodied in future learning received little attention.

The gap between the two types of objectives may, however, differ for the class-teacher and for the integrated-subjects teacher. Theoretically, the class-teacher system is more likely to minimize the gap between short-term and medium term objectives. Class-teachers are expected to stay with students over a period of up to 3 years, so that responsibility for both short-term and medium-term objectives remains with the same teacher. This continuity appeared to be realized in the observed girls' school. In the fieldwork, Mrs A, the class teacher, repeated to her class that she did not want teachers in year 4 to criticise her students' competence. She was aware that she would be judged when her students started the second

stage. In arranging her plan, Mrs A told me that in her first year with students, she concentrates on social and learning behaviour; for example, helping students adjust to the school life, getting them to know each other, and building a good relationship with them. In the first year, she said, some qualities are given priority such as reading. In the second and third years, she starts to concentrate on other basic qualities, such as writing and speaking. Being a class teacher, who stays with students for 3 years, Mrs A was aware of medium-term objectives, and, thus, has a medium-term instructional plan as well as a short-term plan. This does not mean, of course, that all class teachers give attention to the medium-term objectives. How far this takes place needs to be investigated in further research.

In the second stage, teachers are mainly responsible for achieving short-term objectives specified for each year. A gap may arise if teachers pay differential attention to the qualities assessed. The gap may be wider in some circumstances, such as in the case of the boys' school in this research. Teachers, like Mr E and Mr F, who started to teach low achieving students in year six, felt that low achievement was the responsibility of the previous years' teachers and/or could be attributed to the students' natural abilities. In either case, they felt these deficiencies fell outside their responsibilities as class teachers. They felt that they could do nothing to remedy students' accumulated weaknesses. Moreover, their subsequent neglect of these deficiencies may even have lowered students' expectations. Responsibility for the medium-term objectives is easily lost among second stage teachers.

This gap was investigated, in the fieldwork schools, by contrasting the medium term objectives of the second stage (years 4 - 6) with the end of term tests (first term and second term) for 1994/95 in both schools. The analysis checked whether the stated medium-objectives were represented, in some form or another, among the end of term test items. It was found, for example, that the medium-term objective in Arabic to 'extract the basic facts from audible text' was reflected in the listening test. Students listened to a small segment read by their teacher and then were asked to answer several questions about what they had heard.

As illustrated below, this comparison suggests that a gap can arise between medium-term objectives and what teachers actually assess in the end of terms tests.

		Medium term objectives	Objectives present in the end of the two terms test
Arabic	Girls	22	10
	Boys		6
Mathematics	Girls	38	13
	Boys		16

Although the above contrast may be overstated because the end of term tests are not expected to assess the medium-term objectives, the results are indicative. They suggests that schools are not paying attention to the achievement of these objectives. The objectives appearing in the end of term tests were all short-term objectives. That is, teachers in the above cases were mainly assessing instructional objectives. In this case, an agenda for assessing medium-term objectives was

missing. So far, the system has no means of knowing whether the educational objectives for each stage have been achieved; nor is there any systematic or valid procedure to assess the overall standard of schools' performance or to account for comparability of performance between schools. It was not until 1992, when the Minister of Education announced the need for such information that a committee was formed to address these problems. This committee has not yet reported.

Cognitive and non-cognitive objectives:

Cognitive and non-cognitive performance is one of the ways in which assessment systems differentiate among pupils. Cognitive objectives are related to acquisition of knowledge and learning of skills, while non-cognitive ones relate to social and moral skills. Parsons (1959) suggested that the cognitive component of education is the 'learning of information, skills, and frames of reference associated with empirical knowledge and technological mastery'. The second component, he identified as 'moral', includes the non-cognitive attributes of behaviour, such as, 'consideration and co-operativeness in relation to fellow-pupils, and good "work habits" ... leading to capacity for "leadership" and "initiative"' (p. 304). In addition, Parsons suggested that non-cognitive qualities are as important as cognitive qualities and cannot be clearly differentiated from each other in the assessment process. Overlap of cognitive and non-cognitive elements of learning is also demonstrated in the work of Wood and Napthali (1975). Their conclusion was that:

> Although consideration of intellectual or cognitive qualities bulks large, teachers are influenced by other factors; in particular, the extent of pupil commitment and interest in the subject appeared to be significant (p. 159).

They found that teachers' differentiation of pupils is also based on all or some of six non-cognitive constructs; namely,

> (a) The involvement of the pupil in the learning situation.
> (b) The ability the pupil has in the subject.
> (c) The overall ability of the pupil.
> (d) The behaviour of the pupil.
> (e) The quality and tidiness of work presented.
> (f) The interest displayed by the pupil in the subject (p.159).

Nevertheless, cognitive objectives have dominated assessment practices. This dominance of cognitive assessment derives from the persistence of the psychometric paradigm in educational practice. It assumes that knowledge is static, separable from other aspects of human life and, therefore, is easily measured. This view looks to individuals as absorbers of knowledge and detaches the human intellect from social experience. Emphasizing formal ways of assessment (such as tests) plays a central role in reinforcing and sustaining the assessment of cognitive development, competence and performance. Qualities such as co-operation, responsibility and independence cannot be easily tested. Thus, they become neglected qualities.

As new views of learning come into circulation, more attention is being given to social qualities. Consideration of both cognitive and non-cognitive qualities is not just important for the development of individuals, it is also necessary for understanding students' intellectual potential. The acquisition of knowledge and skills does not occur apart from the wider, social experiences of individuals.

This integrated view of individual development is clearly reflected in the general aims of education in Bahrain. Primary education focuses on raising each pupil's 'physical, intellectual, spiritual, social, and moral development', to an 'optimum aptitude' (The Ministry of Education, 1985). Such views are, however, not reflected in the objectives of curriculum in primary education. Objectives of education for both the first and the second stages are dominated by cognitive qualities. Among 249 objectives covered in the two stages, only 34 touched on the social and physical development of the child (18% and 11% in the first and second stage respectively).

Some cognitive objectives may enhance the other qualities such as social and spiritual development. For example, the ability to 'specify some of the Islamic morals and practices' (objective No 6 in religion for the first stage), may also enhance students' moral and spiritual development. Nevertheless, using the word 'specify' in the curriculum guide, means that the emphasis on such objectives could focus on recall rather than understanding. This imbalance distorts the philosophy of child-centredness which the Ministry chose to adopt in 1983. The only conception of learning evident in the assessment programme is the view of learners as passive absorbers of quantities of knowledge - a view of learning widely questioned by contemporary educational commentators (e.g. Wood, 1994).

In the field work of this research, it seemed that the acquisition of knowledge and skills was still the teachers' overriding concern. Nevertheless, female teachers still reflected upon other areas. They also gave attention to students' social

development. This was evident in year 3 (in Mrs A's class). Mrs A directed the students' attention to approved and disapproved behaviour. This attention helped to clarify standards for discipline in the classroom. Mrs C and Mrs D, the year-six teachers, gave attention to students' social behaviour in their records. Such teachers also rewarded students for qualities such as involvement in classroom activities, and the tidiness of their work.

Teachers in the boys' school in both stages of schooling confined their attention to the academic domain. Neither Mr B, Mr E nor Mr F gave attention to their students' moral behaviour in the classroom. Lessons in all classes seemed predominantly academic. Indeed, Mr B, the year 3 teacher, condoned behaviour which interfered with lesson progress. But this was never commented upon nor subject to remedial actions by him.

The emphasis on cognitive qualities provided a rationale for teachers and a system for determining success and failure. But it also had social consequences. It served to sharpen the separation of students into successes and failures. The longer students stay in school the wider the gap between successes and failures can become. Schools in Bahrain honour the brightest students in public events attended by all staff in the school and all students and sometimes by parents. For the lowest achievers, remedial lessons may be offered. These lessons have to be paid for if they are organized outside school hours. Although the purpose of these lessons has never been questioned, their effectiveness is not expected to be higher than ordinary classroom lessons. Observed schools seemed to assume that poor

achievement is merely the result of students' intellectual weaknesses, ignoring the context in which such performance is achieved.

The differentiated process in primary schools is a result of understanding achievement in terms of cognitive qualities only and neglecting their relationship with non-cognitive qualities. Schools, however, are also viewed as socialization agencies and for this role to be fulfilled, non-cognitive qualities are of comparable importance to cognitive qualities. There is need, therefore, to shift towards more comprehensive approaches to assessment if all-round improvement in schooling and learning is to be achieved. There is a need to value personal and social qualities as much as academic achievement.

Low and high level skills objectives:

This issue is related to the level and type of thinking that students follow to reach their answers. It has been tackled from different perspectives. Educational measurement conventionally extends this continuum from simple objectives (knowledge) to more complex objectives (evaluation), as mentioned in chapter 2. Other theorists have used more simplified models; for instance, Hudson (1966) made a distinction between convergent and divergent thinking. Convergent thinking occurs when the individual is required to find the right answer and when her/his own choices is restricted by the assessor. Divergent thinking arises when the individual is invited to think fluently and when there is a variety of possible answers to the problem. The key educational assumption in this work is that promotion of higher-order thinking increases the wider generalisability and

applicability of learning.

The same interpretation operates in the Bahrain curriculum. The official curriculum states that assessment should:

> go beyond the assessing of acquisition of knowledge to the assessment of high level qualities such as analysing, synthesis and evaluation. It should consider the other objectives such as application which has relevance to the real life situations. (DoC, 1992, p. 26).

Application of this approach in the classroom depends on several things. The first is how far this view is translated in textbooks since the latter are central in the classroom practices. The only research which tackled this aspect was undertaken by Wehbeh *et al* (1990) when they analyzed the social studies textbooks for years 4 to 6. They concluded that these textbooks inhibit higher level thinking. Textbooks sustain the image of a convergent, passive individual who accepts life as it is, who is hardly creative and who may even become afraid of change. Social studies textbooks are being replaced to achieve more quality learning outcomes. But more research is still needed to see whether textbooks foster or inhibit higher level thinking.

The second element which needs to be investigated, is whether teachers' assessments consider the different levels of thinking. A comment made by a class-teacher whom I interviewed, highlights the overemphasis on assessment of the low level skills.

> We need to make assessment of the higher order skills. Our schools do not assess these skills. We do not know how to test them and to define the level of performance which indicate that students approach those skills (Class teacher in Al-Alah primary boys' school).

The implication of the observation described in chapters 7 and 8 is that teachers concentrated on closed questions which anticipate short answers derived mainly from textbooks. Test items were also highly structured to enable students to produce pre-defined short answers. The appropriate test response was reduced to mere words or ticks.

Assessment was isolated from life-long learning. Learning only occurs in the classroom. It is confined to what the textbooks say. As observed in the field-work, two class-teachers, Mrs A and Mr B, adhered to textbook content, which in itself limited students' chances to do other external activities. A student in Mrs A's class asked whether she could use 'the birds' instead of 'bird' in forming a sentence. Mrs A refused the student's request. She turned to me explaining that she wants students to conform to the rules in the textbook. She thought, perhaps, that she was doing good for the children themselves in the short term; that is, encouraging students to respect the textbooks as sources of learning and enabling them to pass the test. But she probably had not considered what such a practice could produce for the students' real life learning.

In the second stage, the same practice was also evident. Mrs C admired students who presented answers which echoed what the textbook contained. In the boys'

school, Mr E noticed that some students worked on exercises from external books. He seemed to dislike this and asked those students to use only the school's textbooks for doing their exercises.

A likely consequence of such emphasis on closed, highly structured tasks is that it may encourage low quality learning outcomes. It hinders students' ability to think. It confines them to a very limited range of knowledge and activities embodied in the textbooks. It may produce pupils who can read, write, do some mathematics and have some factual knowledge. But it may not foster the responsible, creative and independent person that Bahraini society needs for the future.

The Methods of Assessment

Assessment methods correspond with what teachers assess. Teachers who pay attention to cognitive qualities are more likely use formal ways of assessment such as written tests, whereas teachers who are interested to assess other qualities such as co-operativeness, creativeness and independence, would seek other ways of assessment such as group projects and research work.

The Bahrain curriculum document suggests different methods for assessing students' learning; for example, oral, practical and written tests, using observation cards and portfolios. The Assessment Act (1983) also specifies varied methods of assessment. Further, it prescribes different methods at the first and second stages.

In the first stage, assessment is expected to include systematic observation, classroom daily practice, students' tasks, projects, and diagnostic tests. For the second stage, assessment follows the same procedures, complemented with quizzes and tests (mid-term and end of term). In the second stage, the use of tests for the summative reports seems to take priority over the first stage procedures. Summative reporting, however, may do little to foster improvement in the teaching/learning processes.

The broader view of assessment is that the first stage should have an impact on teachers practices in the classrooms. This opportunity does not seem to have been taken. Mrs A, for instance, was still addicted to tests. She was also constrained to submit summative records at the end of terms. Although Mr B abandoned tests entirely, he seemed not to adopt other alternatives. He proceeded with intuitive assessment procedures and still used norm-referenced criteria at the end of terms.

Attempts to enhance the validity of teachers' assessment in both stages have been made at the curriculum level (Yosif, 1991). The objectives of the curriculum which are considered as learning outcomes are expressed in descriptive terms of performance. They are designed to guide teachers to make criterion-referenced assessments. This form of assessment is accompanied by a criterion-referenced record form. The key principle in the design of these records is that each basic skill is broken down into more elementary skills. Teacher are expected to assess and record students' progress in each of these elementary skills. Whenever students do not master the required skill, teachers are expected to repeat the assessment

procedures after a period of time, until the skill is mastered.

These records provide evidence of the learning progress of each student in the different skills necessary for the mastery of the subject. They, therefore, support models of formative, continuous assessment that allow for individual differences. In my visits to primary schools, three schools appeared to use them in assessing students' learning. In two schools, however, these records were also used summatively to compare students or to measure the overall standard of the class. They provided a comprehensive report for the class as well as for the individual students. It seems that, in these schools, the assessment records were used as much for reporting purposes as for improving learning and teaching.

At the first stage of schooling, teachers seemed prepared to substitute norm-referenced for criterion-referenced assessment, probably prompted by the emphasis placed upon the acquisition of basic skills.

> Assessment, for us, focuses on basic skills which students should master; but not on the content of lessons... We scrutinize the common errors and try to focus on them in the following lessons, and so on (Interview with class-teacher in Al-Alah boys' primary school).

Class-teachers whom I met shared this view of assessment. Classroom assessment allows for a criterion-referenced model to be applied. Assessment questions are more sophisticated and are intended to elicit whether the student possesses a definite skill or knowledge required by the curriculum. Nevertheless, the centralized curriculum became an obstacle in the implementation of a fully

differentiated criterion-referenced approach. Individual differences should be recognized and taken into account in criterion-referenced assessment, a practice which contradicts the teachers' commitment to follow textbooks syllabi and associated normative assumptions. For this reason, criterion-referenced methods were retained for the benefit of the whole class rather than the individual student. For curriculum purposes, students were treated like classes rather than as individuals. For example, all students were given the same tasks and the same time to complete them.

Use of systematic observation and diagnostic tests became a new requirements in 1983. Such methods were deemed important because students' difficulties in learning can be more easily discerned through observation and diagnostic tests. Observation, for example, allows teachers to examine students' work closely and thus to identify errors and difficulties in learning and remedy them as early as possible. Some teachers seemed aware of including observation as part of assessment:

> I rely on my observation in the classroom. As I stay with students 5 lessons (4.5 hours) a day, I could better make observation notes about their progress each month (In an interview with class-teacher in Sitra girls' primary school).

This teacher seemed to rely on memory to record her observation about students' progress. Her style of observation was intuitive, and built on her general experience with students rather than on a deeper, focused observation of an individual child. Sound observation is based on prolonged collection of evidence about learning. But

carrying out systematic observation takes up teachers' time and also requires special training, neither of which is available for most teachers. As teachers enjoyed little training and support for these procedures, it was not surprising that they used simpler methods of assessment (i.e. tests) or abandoned assessment altogether (as in Mr B' case). In the field work, neither observation nor diagnostic testing was observed in classrooms. As students went about their classroom work, teachers took the opportunity to do other work, such as correcting and marking.

The emphasis on tests in the second stage of schooling as a major instrument of assessment segregates assessment from the teaching and learning process. By such emphasis, tests are undertaken ritually, following bureaucratic procedures. They only serve reporting and summative purposes. Tests were mainly used to report students' achievements as required in the policy. But even this information cannot be securely linked to school practice. This study has questioned the relevance of such reports of students' actual achievements. In the girls' school, evidence showed that teachers drew their final assessment reports from students' actual performance in the formal tests. This is because students' performance in these test was highly compatible with mastery performance. But this was not the case in the boys' school. Students' performance was much lower than the system envisaged. For this reason, the actual results were moderated to meet the required level. In this respect, tests are not a reliable measure for making comparisons among classes and schools or to judge the overall standard of performance of the system.

The Purposes of Assessment

The purposes of assessment embrace the philosophical and political context of formal education. Answers to the question 'why assess?' disclose the aims of education. There are many aims which can be ambitiously anticipated for an educational system. But the real achievements are related to the actual outcomes, identified by assessment. The examination of assessment policy and practice, therefore, offers insight into the purposes and consequences of any educational system.

Bahraini educational policy offers a double guarantee. All children are entitled to acquire the basic education they need for a better future life and, secondly, to choose the kind of education they think would fulfil their needs (The Ministry of Education, 1985). In addition, the intention of the assessment policy is to maintain a balanced flow of students through the system and, secondly, to sustain and advance standards as cohorts of children pass through the system. In practice, assessment has been intensively used to achieve the first goal, that is, an effective through-put of school students. But it has not yet operated to achieve the second political task. Monitoring through-put is not the same as improving learning and enhancing standards.

Monitoring students' progress.

At the end of each year, the school decides those who can be upgraded and those

who need to retain their grade for another year. Two important aspects of monitoring students' progress need to be considered: the promotion-retention policy and the standard accepted for that promotion. First: the promotion-retention policy maintains the test as the main device for determining students' promotion through grades 4 to 6. Students in years 1 to 3 may also be retained if their performance is below the required level (see chapter 4). The characterisation of tests as high stakes measures creates pressure on teachers to teach for the test. This study gives evidence of a close match between teaching and testing. Preparation for the test was a common practice for almost all teachers of this study. Students were given some knowledge of what might come in the test; and selection of the test items was also influenced by what students are actually able to do.

Teaching to the tests to raise test scores is a kind of practice that exists also in other countries. Gipps (1994) reviewed some of the studies which documents similar evidence of these practices in the USA. These practices (called test score pollution) include:

> teaching test-taking skill, promoting pupil motivation for the test, developing a curriculum to match the test, preparing teaching objectives to match the test, preparing pupils on items similar to those in the test, using commercial materials specifically designed to test performance, and presenting before the test the actual items to be tested (p. 46).

Shepard and Smith (1989) have argued that teaching to the test is reinforced by retention policies. In the USA, enduring attempts have been made to maintain standards. The policy of automatic promotion has been replaced by merit-based promotion which, in turn, favours retention. Retained students, as Shepard and

Smith pointed out, who fail to meet the standard set for promotion need to be specially tutored. 'Remediation usually involves teaching to the tests rather than pursuing any broader educational goal' (p. 224). Students' motivation is also affected while the repeated testing does not guarantee the improvement of achievement. It is acknowledged that the repeated testing increases performance even if the student had learned nothing. Retention, they argued, is 'a response to the accountability culture and the factory model of schools...the accountability function is largely symbolic, as no real gains in achievement are made by the pupil who is so recycled' (p. 224). Thus, holding teachers accountable for a standard which is evaluated by paper-and-pencil tests affects teachers' practices. The public and professional interpretation of promotion-retention policies shapes, in part, the realities of classrooms.

The second aspect of the promotion policy is the standard accepted for promotion, set at 50% of the possible total score. But even this policy can be breached. Students can be promoted if they have less than 50% in no more than two subjects, provided neither is Arabic. The policy, however, stipulates that schools should provide remedial programmes for repeating students. But these are not available in all schools. Remedial programmes require qualified teachers, more time and additional resources. Thus, many of these students may have no opportunity to remedy their weaknesses and improve their learning.

The system for promotion serves to minimize the retention rate in primary schools as described in chapter 4. Lowering the standard makes it easier for more students

to reach this level, for the teacher to advance students above the promotion threshold, and for schools to rescue students who initially fall below this level. This means that teachers are unlikely to achieve improvement in quality learning for all children since the policy does not expect them to do so. They may not strive to bring all children to mastery learning if a mixture of high, medium and low achievers is the expected distribution for achievement. Teachers will be satisfied with their efforts when pupils advance beyond the minimum level. But low expectations could also influence students. The minimum standard becomes the maximum that all groups - teachers and students - strive for. It may become a goal in itself.

One of the reasons for introducing such a promotion policy in 1983 was to motivate students to stay in school and enhance their learning. But this policy was also intended to reduce costs, by reducing the school population. Further, the proportion of students promoted from one year to the next also became the criterion by which school effectiveness was to be judged. For example, a school with 90 per cent of promoted students is deemed to be more effective than a school with only 80 per cent promoted students. Promotion procedures may reduce costs, increase through-put, and differentiate schools, but their relation to the achievement of national curriculum objectives is unclear.

This attention to the through-put of education ignores the reality of students' achievements in schools and thus, the real quality of education. It does not allow due consideration of performance in the system as a whole. As the central

department devotes so much attention to the internal efficiency of schooling, raising the number of passing students becomes an end in itself. Schools and teachers are diverted to maximizing the proportion of promoted students since performance on this criterion is used to judge their effectiveness. This criterion for judging schools is drowning out other aspirations and practices. The evidence may point to higher promotion ratios and lower retention in primary schools over the last decade. But there are few in-built guarantees that these changes in retention rates indicate comparable improvement in levels of learning.

The impact of such practices on learning outcomes is rarely addressed. Pushing year 6 students who have already failed, is a high risk, high stakes strategy. For instance, in year 6 in the boys' school, there are two extremes of students. Mansor represents those students who could hardly read, write and do any mathematics and Mahommood represents the most competent students. They both were in year 6 and both promoted to the third stage of basic schooling. Quite a number of students, like Mansor, may have been moved through the system without adequate consideration of whether they acquired the minimum basic skills. Long-term consequences are expected of such practice. Future opportunities for students like Mansor are seriously restricted. Their chances of getting a job and, thus, a proper position in society are diminished since the selection for jobs and training programmes depends on qualities they do not possess.

This study, however, does not intend to generalize that all schools in Bahrain are the same. There are of course differences as well as similarities. Further studies are

needed to investigate the impact of assessment policies and the accountability system over teachers and students. Bearing in mind the limitations of this study, it should be noted, nevertheless, that such practices could not be interpreted as a practice of poor teachers. This kind of interpretation ignores the impact of tests and accountability over teaching and learning. The question is whether good teachers under the same circumstances (poor achieving students, low motivation, pressure to raise the number of students promoted and pressure to finish the required syllabi) would adopt different practices. Many teachers in Bahrain face circumstances similar to those faced by the teachers of this study. They are, therefore, under the same pressure. Tests are the most readily available device to reduce this pressure.

This type of compromise is also found in other countries. As noted earlier, Shepard and Smith (1989) showed that this practice is a reflection of promotion-retention policies. They gave evidence of the discrepancies between rhetorical policies and realities. The volume includes a case-study of Ellwein and Glass which demonstrated inconsistencies between public statements about reform and the implementation of retention policies. The Waterford School District (not the real name) was implementing a reform programme which aimed to replace the traditional social promotion system with new testing programme from Kindergarten to seventh grade. Results of these tests were to be used as the single criterion by which students were promoted or retained. The researchers followed the implementation of this reform programme. They matched test performance and subsequent decisions. They found that some students (29% of kindergarten and over 50% of second graders) identified for retention were nevertheless, promoted.

They concluded that ' test performance and promotion decisions were related asymmetrically' (Ellwein and Glass in Shepard and Smith, 1989, p. 171). They acknowledged that tests are corruptible within an accountability system and the outcomes might be detrimental to the improvement of learning if teachers are pressured.

The social consequences of the spread of such practice are costly to a small country, like Bahrain. At the present, Bahrain is highly dependent on imported labour. Foreign workers accounted for nearly 63 per cent of the total labour force in 1991. In a free labour market, Bahraini workers found themselves in competition with highly qualified foreign labour. In such a competitive market, they need to be equipped with high level skills and qualities. This situation is compounded by the fact that Bahrain has a small population, with nearly 50 per cent under the age of 20 years in the 1991 census. Education is becoming a high cost social service. It should render future benefits in terms of the quality of its outputs. The education system should give greater attention to achievement rather than through-put. The problem of low achievement and the consequences of social wastage in the long term, cannot be solved by tinkering with the promotion policy. It requires changes in classroom practices (teaching and assessment). These practices have never been questioned nor have the consequent outcomes of education been investigated.

Improving learning outcomes:

The improvement of learning is indeed what all teachers and the system anticipate

for students. Assessment can significantly contribute to this purpose in different ways. It can, for example, motivate students to learn, diagnose their difficulties in learning and give them feedback on what they already know and with what they need assistance. Assessment, in this sense, is integrated with teaching and learning.

Attempts to use assessment for this purpose have, indeed, been made by the central department of education (Directorate of Curriculum) and within the schools themselves. Initiatives at the central level include the development of criterion-referenced record keeping (mentioned above), tests which take account of the different level of skills, and teaching and assessment materials for remedial and diagnostic purposes (Yosif, 1991). These materials are intended to help teachers assess student learning in basic skills, recognise the achievement standards of their students, and to manage the teaching and learning process.

In my interviews with teachers, I found some acknowledgment of new approaches in assessment; for example, consideration of students as individuals in the teaching/learning process, emphasis on the need for diagnostic assessment, and acceptance of the need for a broader view of assessment, including cognitive and non-cognitive, low and high level objectives. My fieldwork, however, provided no evidence of the impact of such new perceptions on classroom practice. They are likely, I suspect, to remain unadopted until research and development is undertaken with regard to the further harmonisation of classroom practices with national aspirations and policies.

The role of assessment in improving learning has, however, already been questioned in Bahrain. Attempts to measure learning outcomes of primary schools have raised several questions with regard to schools' assessment practices. This research study was carried out in 1989 and reported its results in 1993 (Wehbeh *et al*, 1993). The study used a criterion-referenced test in both Arabic and mathematics for year 3, the end of stage one, and year 6, the end of stage two. It covered most skills required by the curriculum objectives and involved both low and higher order skills. Nearly 10 per cent of students in the third and sixth primary years completed the test at the end of the academic year 1989/90. Student responses were scrutinized in depth and a full analysis of student performance was provided.

The results revealed unexpected deficiencies in students' learning outcomes (see a summary of the results in appendix 2). In the Arabic test; for example:

1. Nearly a quarter of students had not mastered the basic skills and a higher proportion were far below the curriculum criterion (nearly 35% in both years 3 and 6).
2. Nearly all students lacked the ability to construct an extended sentence in the extended essay test.
3. Unreadable papers were submitted for the free extended essay (24 papers in year 3 (7%) and 31 papers in year 6 (8.3%).
4. Nearly all students failed to go beyond the given text in the reading comprehension test.

Similar results were found in mathematics; for example:

1. Nearly one third of year 3 students were deficient in the basic skills in numbers concepts.
2. Nearly one third of year 3 and year 6 students could not give correct answers in measurement.

3. A high proportion in year 6 (from 20% to 80%) were deficient in the geometry test.
4. Nearly one third of students in year 3 and more than half of the students in year 6 were unable to find the correct strategy when solving problems.

The research suggested that the quality of student learning outcomes was not in line with official expectations. The question remains: why has the required quality of learning not been established despite all efforts of the Ministry?

Towards New Policies and Practices

This chapter has reviewed Bahraini assessment practices and policies. It suggests that an out-of-date assessment culture still prevails in primary schools. Bahraini assessment practice, at both governmental and classroom level, lags behind recent international thinking about assessment. Changing social and political conditions around the world have produced new perceptions of assessment. The new view is that assessment should also be integrated with teaching and learning. The search for objective measures of mental ability has become less important than promoting life-long learning outcomes. Accordingly, assessment should not only refer to national standards, it should also be formative, allow for a wide range of abilities and skills, and be used to promote learning outcomes.

In the light of the changing situation in Bahrain, there are also social and economic pressures to change assessment. At the present, there are indications that

unemployment among Bahraini citizens will increase. The labour market demand is more for workers who are creative, personally effective, possess high level thinking skills and have the capacity to adapt to change. In other words, future schooling should offer more than a grounding in reading, writing and mathematics. From the beginning of formative stages of schooling, schools should strive to achieve such advanced aims. The various methods that schools use to assess pupils' achievement and to advance them to higher levels should serve both the needs of individuals and the needs of society. Assessment, in this respect, is a leading factor. With this new perception, assessment has a role to play in improving the quality of learning.

Assessment, at the present, is seen as a powerful means of educational reform in many countries. This role of assessment, for example, is now manifest in the USA, the UK and some other developed countries (Harlen, 1994). Torrance (1995) also pointed out that using assessment in educational reform is an idea whose time has come. He considered that this awareness of the importance of assessment has created opportunities to conceptualize and advocate the different uses of assessment: the use of assessment in selection and its related reliability and validity, system monitoring and management, standard setting and accountability and finally teaching and learning in the classroom. In general, two basic perspectives emerge: using assessment results to influence or manage the system; and using assessment processes to improve teaching and learning. These perspectives have not yet been incorporated into educational policy in Bahrain. Assessment has not yet made an effective contribution to the educational system. Changes in education which have been initiated since the 1970s have neglected

assessment as a factor for enhancing and enforcing such improvement.

The first indication of this negligence is the limited role allowed for assessment within teaching and learning. Attention to classroom practices in the early 1970s (as noted in chapter 4) focused on curriculum, pedagogy, teachers' training, textbooks and instructional materials. No attention, that is, was given to the integration of teaching and assessment. Another initiative was the creation of the Directorate of curriculum in 1975 and the rapid expansion of its staff (to nearly 144 curriculum specialists in 1995). The educational aspiration was to develop new curricula rather than to reform the whole processes of teaching, learning and assessment. Similarly, teacher training programmes have helped since 1983 to foster child-centred approaches to teaching, but without comparable attention to child-centred models of assessment. Finally, although a special unit for Measurement and Assessment was opened in the Directorate of Curriculum in 1984/85, the work of this unit regarded assessment as separate from teaching and learning.

New approaches in assessment which embodied new views of learning circulated among schools, but they only touched the surface features of assessment. Practices did not change. Assessment has remained an isolated and bureaucratic process which mainly serves to monitor student progress and provide summative reports. Thus, the positive influence of assessment on teaching and learning is very limited. To the extent that assessment determines what is taught and what is learned, improvement cannot be achieved without integrating assessment into learning and

teaching.

Recently, this shortcoming in classroom assessment has been politically addressed in a new Assessment Act (1995), mentioned in chapter 4. It calls for the integration of assessment into teaching and learning. In turn, it gives teachers flexibility with respect to curriculum implementation, which seems, as I noticed in this study, to be the main constraint upon teachers. The impact of such changes on classroom practices will depend on the responsiveness or resistance of the present assessment culture. It will depend on how much effort the central departments in the Ministry make to extend teachers' professionalism in assessment. Besides such developments, research is also needed to evaluate such policy initiatives, especially their classroom consequences.

Another indication of the neglect of assessment is the absence of a national assessment programme which could inform the educational policy as a whole. Since the abolition of the General Certificate of Primary Education in 1979, no measures have been taken to ensure the reliability and validity of schools' assessment. Assessment has been entirely left to schools. Although central control over curriculum is practised, this study has cast doubt on the reliability and validity of schools' assessment and also pointed out the absence of any responsibility for ensuring the achievement of the objectives of education at the end of each stage. The only information available to the system about the students' achievement is recorded in the summative official reports. These reports fail to discriminate the different aspects of students' achievement in each subject. Further, research studies

which were conducted in this area do not offer continuous, up to date information by which standards of achievement can be compared in a number of periods.

This brings up the need for national assessment measures (probably high stakes) to be introduced for the two stages of schooling. The Bahraini educational policy is moving at the present towards the fulfilment of such a need. The first political step was actually taken in 1992, when a Ministerial Resolution established the Committee for National Assessment in primary education (see chapter 4). This committee has not yet (in Spring 1996) reported its findings and decisions. The national test carried out on a sample of schools on June 1996 is expected to illuminate much of the realities of schooling and achievement in primary schools. The result could reveal the extent of the inflating scores in the system and what area of improvement is required for assessment reform.

From the above, it appears that two distinct areas of reform are needed: the creation of a valid and reliable national assessment programme and the integration of assessment into classroom teaching and learning. These two kinds of assessment cannot be easily reconciled, a policy issue which already faces other countries (Harlen, 1994 and OECD, 1993).

Assessment in Bahrain is in a similar position. A national assessment policy should answer several questions. What kind of achievement should be assessed? For what subjects and at what stages of schooling? What types of methods should be used? And what role should teachers play in this process? Such questions have already

been posed to the National Assessment Committee. Whatever the answers to the above questions, the consequences of national assessment for classroom practice should also be considered. National assessment introduces new types of control over schools and teachers for the purpose of accountability and raising standards. It may enforce the use of testing, and emphasize the reliability rather than the validity of assessment. It may also mean that teachers will teach for the test.

On the other hand, the results of this study suggest that national assessment policy should be very sensitive to teachers' assessments. A review of assessment should take into account the purpose of promoting individual learning and this is best achieved by enhancing teachers' awareness of assessment. Classroom assessment should provide information about students' learning in a form which can guide teaching and learning, and inform Bahraini society about educational standards and the improvement of results. The present technique of assessment used in school can hardly achieve such an aspiration. Reform requires substantial changes in teachers' attitudes towards learners and learning which, in turn, require new training programmes and additional classroom resources.

Thus, professional development must be a key element in national assessment. The task for educators in Bahrain is to view assessment as central to the process of quality improvement in education and society.

References

Abercrombie, J. (1960) *The Anatomy of Judgement.* London, Hutchinson.

Airasian, P. W. (1994) The impact of the taxonomy on testing and evaluation. In Anderson, L. and Sosniak, L. (eds) *Bloom's Taxonomy, A Forty-year Retrospective.* Chicago: NSSE. pp. 82 - 102.

Al-Ameen, A., Nazem, J., Hamood, A., Al-Mutawa, A. and Al-Shaik, N. (1993) *Evaluation of the Development of Primary Education.* Educational Development and Research Centre, Ministry of Education, Bahrain (In Arabic).

Al-Hamer, A. M. (1969) *Development of Education in Bahrain 1940-65.* Bahrain, Oriental Press.

Al-Nahas, M. and Katul, J. (1960) *Report of Education in Bahrain.* Department of Education, Bahrain (In Arabic).

Al-Saleh, F. (1992) *Summary of Thesis and Dissertations in Education.* Information and Documentation Centre, Ministry of Education, Bahrain (In Arabic).

Al-Sulayti, H. (1983) *Student Retention in Public Elementary Schools in Bahrain and its effect on input-output.* Ph.D thesis, Indiana University.

Anastasi, A. (1986) Evolving concepts of test validation. *Annual Review of Psychology* **37**, 1-15.

Anderson, L. W. (1994) Research on teaching and teacher education. In Anderson, L and Sosniak, L. (eds) *Bloom's Taxonomy, A Forty-year Retrospective.* Chicago: NSSE. pp. 126 - 45.

Barakat, H. (1991) *Contemporary Arab Society: A social exploratory study.* Beirut, Centre for Arab Unity Studies (In Arabic).

Beeby, C. E. (1966) *The Quality of Education in Developing Countries.* Oxford, Oxford University Press.

Bennett, N. (1976) *Teaching Styles and Pupil Progress.* London, Open Books.

Bloom, B (ed.) (1956) *Taxonomy of Educational Objectives, Book 1: Cognitive domain.* London, Longmans.

Bloom, B., Madaus, G. and Hasting, J. (1981) *Evaluation to Improve Learning.* New York: McGraw-Hill.

Bourne, J. (1994) A question of ability. In Bourne, J. (ed.) *Thinking Through Primary Practice.* London, Routledge. pp. 212 - 222.

Broadfoot, P. (1979a) *Assessment, Schools and Society.* London, Methuen.

Broadfoot, P (1979b) Communication in the classroom: a study of the role of assessment in motivation. *Educational Review* **31** (1): 3 - 10.

Broadfoot, P. (1981) Towards a sociology of assessment. In Barton, L. and Walker, S. (eds) *Schools, Teachers & Teaching.* London, Falmer Press. pp. 197 -217.

Broadfoot, P. (1995) Performance assessment in perspective: international trends and current English experience. In Torrance, H. (ed.) *Evaluating Authentic Assessment.* Buckingham, Open University Press. pp. 9 - 43.

Brophy, J. and Good, T. (1974) *Teacher-student Relationships.* New York, Holt, Rinehart and Winston.

Brown, G. A. and Edmondson, R. (1984). Asking questions. In Wragg, E. C. (ed.) *Classroom Teaching Skills.* Beckenham, Croom Helm, pp 97 - 120.

Bruner, J. (1985) Vygotsky: a historical and conceptual perspective. In Wertsch, J. (ed.) *Culture, Communication and Cognition: Vygotskian Perspectives.* Cambridge, Cambridge University Press. pp. 21 - 34.

Carver, R. (1974) Two dimensions of tests, psychometric and edumetric. *American Psychologist* **29** (7), 512-18.

Choppin, B. H. (1990). Evaluation, assessment, and measurement. In Walberg, H. J. and Haertel, G. D. (eds) *The International Encyclopedia of Educational Evaluation.* Oxford, Pergamon Press. pp. 7 - 8.

Cronbach, L. (1971) Test Validation. In Thorndike, R. L. (ed.) *Educational Measurement (2nd ed).* Washington, DC: American Council on Education. pp. 443 - 507.

Cronbach, L. (1980) *Toward Reform of Program Evaluation: aims, methods, and institutional arrangements.* San Francisco, Jossey-Bass.

Crooks, T. (1988) The impact of classroom evaluation practices on students. *Review of Educational Research* **58** (4), 438-81.

Directorate of Curriculum. (1992) *The Framework for Basic Education Curriculum.* Ministry of Education, Bahrain (In Arabic).

Directorate of Planning. (1981) *Comparative Statistical Study of the Last Ten Years (1971/72 to 1980/81).* Ministry of Education, Bahrain (In Arabic).

Directorate of Plans and Programming. (1990) *The Educational Plan for 1989/90 to 1994/95.* Ministry of Education, Bahrain (In Arabic).

Downey, M. (1977) *Interpersonal Judgements in Education.* London. Harper & Row.

Ebel, R. (1969) The relation of tests to educational objectives. In Ingenkamp, K. (ed.), *Developments of Educational Testing.* London, University of London Press. pp. 483 - 92.

Eisner, E. W. (1985) *The Art of Educational Evaluation, a personal view.* London, Flamer Press.

Eisner, E. (1993) Reshaping assessment in education: some criteria in search of practice. *Journal of Curriculum Studies* **25** (3), 219 - 33.

Erwin, D. (1991) *Assessing Student Learning and Development.* San Francisco, Jossey-Bass.

Fakhro, A. M. (1987) Keynote Address. *Proceedings of the Conference Testing in the Arab World.* Management Planning and Research Consultants (MPRC), Bahrain.

Flanders, N and Amidon, E. (1963) *The Role of the Teacher in the Classroom.* St Paul, Minn: Amidon & Associates.

Fredericksen, N. (1984) The real test bias. *American Psychologist* **39** (3), 193-202.

French, J. and Michael, W. (1968) The nature and meaning of validity and reliability. In Gronlund, N. (ed.) *Readings in Measurement and Evaluation.* New York, Macmillan.

Galton, M., Simon, B. and Croll, P. (1980) *Inside the Primary Classroom.* London, Routledge & Kegan Paul.

Gipps, C. (1994) *Beyond Testing: towards a theory of educational assessment.* London, Falmer Press.

Gipps, C. and Murphy, P. (1994) *A Fair Test? Assessment, Achievement and Equity.* Buckingham, Open University Press.

Gipps, C. and Stobart, G. (1993) *Assessment, A Teacher Guide to the Issues.* London: Hodder & Stoughton.

Glaser, R. (1963) Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist* **18**, 519 - 21.

Glaser, R. (1991) Expertise and assessment. In Wittrock, M. C. and Baker, E. L. (eds) *Testing and Cognition.* New Jersey, Prentice Hall. pp. 17 - 30.

Glaser, R. (1977) Evaluation of instruction and changing educational models. In Hamilton, D., MacDonald, B., King, C., Jenkins, D and Parlett, M (eds) *Beyond the Numbers Game: a reader in educational evaluation.* Berkeley, McCutchan Publishing Corporation. pp. 37 - 51.

Hamilton, D. (1980) Educational Research and the shadows of Francis Galton and Ronald Fisher. In Dockrell, W.B. and Hamilton, D. (eds) *Rethinking Educational Research.* London, Hodder and Stoughton. pp. 153 - 68.

Hamilton, D. (1989) *Towards A Theory of Schooling.* London, Falmer Press.

Hargreaves, D. H. (1975) *Interpersonal Relations and Education.* London, Routledge & Kegan Paul.

Harlen, W. (1980) Evaluation in education. In Straughan, R. and Wrigley, J. (eds) *Values and Evaluation in Education.* London, Harper & Row. pp. 52 - 70.

Harlen, W. (ed.) (1994) *Enhancing Quality in Assessment.* London, Paul Chapman.

Harlen, W., Gipps, C., Broadfoot, P. and Nuttall, D. (1992) Assessment and the improvement of education. *The Curriculum Journal* **3** (3), 215-30

Heyneman, S. and Ransom, A. (1992) Using examinations and testing to improve educational quality. In *Eckstein, M. and Noah, H. (eds) Examinations: Comparative and International Studies.* Oxford, Pergamon.

Hudson, L. (1966) *Contrary Imaginations, a psychological study of the English schoolboy.* London, Methuen.

Jackson, P.W. (1968) *Life in Classrooms.* New York, Holt, Rinehart and Winston.

Keddie, N. (1971) Classroom knowledge. In Young, M.F.D. (ed.) *Knowledge and Control.* London, MacMillan. pp. 133 - 60.

Kelly, A. V. (1992) Concepts of assessment: an overview. In Blenkin, G. and Kelly, A. *Assessment in Early Childhood Education.* London, Paul Chapmon. pp. 1 -23.

King, R. (1978) *All Things Bright and Beautiful.* New York, John Wiley & Sons.

Linn, R. (1983) Testing and instruction: links and distinctions. *Journal of Educational Measurement* **20** (2), 179 - 89.

Linn, R., Baker, E. and Dunbar, S. (1991) Complex, performance-based assessment: expectation and validation criteria. Educational Researcher **20** (8), 15 - 21.

Lloyd-Jones, R. and Bray, E. (1986) *Assessment from principles to action.* London, Macmillan.

Messick, S. (1980) Test validity and the Ethics of assessment. *American Psychologist* **35** (11), 1012-27.

Messick, S. (1988) The once and future issues of validity: assessing the meaning and consequences of measurement. In Wainer, H. and Braun, H. (eds.) *Test Validity.* London, LEA. pp. 33 - 45.

Miller, C.M.L and Parlett, M. (1974) *Up to the Mark, a Study of the Examination Game.* London, Society for Research into Higher Education.

Morgenstern, A. (1966) Historical survey of grouping practices in the elementary school. In Morgenstern, A. (ed.) *Grouping in the Elementary School.* New York: Pitman. pp. 3 - 13.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1994) Teacher expectation. In Pollard, A. and Bourne, J. (eds.) *Teaching and Learning in the Primary School.* London, Routledge. pp. 99 - 109.

Nakhleh, E. A. (1976) *Bahrain, political development in a modernizing society.* London, Lexington Books.

Nisbet, J. (1980) Educational Research: The state of the art. In: Dockrell, W.P. and Hamilton, D. *Rethinking Educational Research.* London, Hodder and Stoughton. pp. 1 - 10.

Nisbet, J. (1993) Introduction. In OECD, *Curriculum Reform: assessment in question.* Paris, OECD. pp. 25 - 38.

OECD. (1993) *Curriculum Reform: assessment in question.* Paris, OECD.

Parlett, M. and Hamilton. D. (1977) Evaluation as illumination: a new approach to the study of innovatory programmes. In Hamilton, D., MacDonald, B., King, C., Jenkins, D. and Parlett, M. (eds) *Beyond the Numbers Game, a reader in educational evaluation.* Berkely, McCutchan Publishing Corporation. pp. 6 - 22.

Parsons, T. (1959) The school class as a social system: some of its functions in American society. *Harvard Educational Review* **29** (4), 297 - 318.

Pennycuick, D. (1990) The introduction of continuous assessment system at secondary level in developing countries. In Broadfoot, P., Murphy. R. and Torrance, H (eds) *Changing Educational Assessment, international perspectives and trends.* London, Routledge. pp. 106 - 117.

Popham, W. J. (1969) *Educational Evaluation.* New Jersey, Prentice - Hall.

Popham, W. J. (1996) *Consequential validity: right concern - wrong concept.* Paper presented at the Annual Meeting of the American Educational Research Association, April 8 - 12, New York.

Rogers, C. (1982) *A Social Psychology of Schooling, the expectancy process.* London, Routledge & Kegan Paul.

Rosenthal, R. and Jacobson, L. (1968) *Pygmalion in the classroom; teacher expectation and pupils' intellectual development.* New York, Holt, Rinehart and Winston.

Rowntree, D. (1987) *Assessing Students: how shall we know them?.* London, Kogan Page.

Ryan, J. (1972) The Illusion of Objectivity. In Richardson, K., Spears, D. and Richards, M. *Race, Culture and Intelligence.* Penguin Books. pp. 36 - 55.

Sanders, N. M. (1966) *Classroom Questions, What Kind?* New York, Harper & Row.

Satterly, D. (1989) *Assessment in Schools, Theory and Practice in Education.* Oxford. Basil Blackwell.

Scriven, M. (1967) The Methodology of evaluation. In Tyler, R., Gagné, R. and Scriven, M. *Perspectives of Curriculum Evaluation.* Chicago: Rand McNally. pp. 39 - 83.

Sharabi, H. (1991) *Introduction to the Study of Arab Society.* Beirut, Dar Al-Taliha (In Arabic).

Sharp, S. (1984) Psychologists and intelligence testing in English Education, 1900-1940. In Broadfoot, P. (ed.) *Selection, Certification and Control; social issues in educational assessment.* London, Falmer Press. pp. 35 - 50.

Shavelson, R. J. (1987) Teachers' judgement. In Dunkin, M, J. *International Encyclopedia of Teaching and Teacher Education.* Oxford. Pergamon Press.

Shepard, L. (1991) Psychometricians' belief about learning. *Educational Researcher* **20** (7), 2 - 16.

Shepard, L. and Smith, M. (1989) *Flunking Grades: Research and Policies on Retention.* London, The Falmer Press.

Simmons, J. (ed.) (1980) *The Education Dilemma, Policy Issues for Developing Countries in the 1980s.* Oxford, Pergamon Press.

Simons, H. (ed.) (1980) *Towards a Science of the Singular.* Centre for Applied Research in Education, University of West Anglia.

Smagorinsky, P. (1995) The social construction of data: methodological problems of investigating learning in the zone of proximal development. *Review of Educational Research* **65** (3), 191-212

Stake, R.E. (1995) *The art of case study research.* London, SAGE.

Têng, Ssu-Yü. (1943) Chinese Influence on the Western examination system. *Harvard Journal of Asiatic Studies* **7**, 267 - 312.

The Government of Bahrain. (1930) *Annual report.* Oxford, Redwood Burn.

The Government of Bahrain. (1940) *Annual report.* Oxford, Redwood Burn.

The Government of Bahrain. (1942) *Annual report.* Oxford, Redwood Burn.

The Government of Bahrain. (1946) *Annual report.* Oxford, Redwood Burn.

The Government of Bahrain. (1956) *Annual report.* Oxford, Redwwod Burn.

The Ministry of Education. (1985) *The Philosophy of Education.* Bahrain (In Arabic).

Thorndike, R. L. (1990) Reliability. In *Walberg. H. J and Haertel. G. D* (eds) The International Encyclopedia of Educational Evaluation. Oxford, Pergamon.

Thorndike, R. L. and Hagen, E. (1955) *Measurement and Evaluation (2nd ed).* New York, Wiley.

Torrance, H. (1989) Theory, practice and politics in the development of assessment. *Cambridge Journal of Education* **19** (2), 183 - 91.

Torrance, H. (1993) Formative assessment: some theoretical problems and empirical questions. *Cambridge Journal of Education* **23** (3), 333 - 43.

Torrance, H. (1995) Teacher involvement in new approaches to assessment. In Torrance, H. (ed.) *Evaluating Authentic Assessment.* Buckingham, Open University Press. pp. 44 - 56.

Torrance, H. and Pryor, J. (1995) Investigating teacher assessment in infant classrooms: Methodological problems and emerging issues. *Assessment in Education* **2** (3), 305 - 20.

Tyler, R. (1949) *Basic Principles of Curriculum and Instruction.* Chicago: Open University Set Book.

Tyler, R. (1986) Changing concepts of educational evaluation. *International Journal of Educational Research* **10** part 1 (entire).

Vallance, A. (1939) *Report on Government Education in Bahrain with Proposals for Reform.* Department of Education, Bahrain.

Walker, R and Adelman, C. (1975) *A Guide to Classroom Observation.* London, Methuen.

Wehbeh, N. (1988) *A Preliminary Study of Teacher Training in Bahrain.* Educational Research and Development Centre. Bahrain (In Arabic).

Wehbeh, N. (1995) Bahrain. In *Final report of the international meeting on education reform and educational research: new challenges in linking research, information and decision making.* National Institute for Educational Research, Tokyo. pp. 48 - 50.

Wehbeh, N., Al-Zayani, M., Al-Shaik, N. and Al-Mannai, L. (1990). *The Image of the Bahraini Citizen in Social Studies Textbooks.* Bahrain Centre for Studies and Research. Bahrain (In Arabic).

Wehbeh, N., Al-Mannai, L. and Al-Mutawa, A. (1991) *The Concept of Educational Research: From Expectation to Practices, an analytic study of M.Ed research plans submitted to the College of Education in Bahrain.* Educational Research and Development Centre, Bahrain (In Arabic).

Wehbeh, N., Al-Mutawa, A., Helal, A., Al-Mannai, L and Al-Kayat, N. (1993) *The Educational Outcomes of Primary Schools.* Educational Research and Development Centre, Ministry of Education, Bahrain (In Arabic).

Wiggins, G. (1993) Assessment: authenticity, content and validity. Phi Delta Kappan **75** (3), 200 - 12.

Wiliam, D. (1996) National curriculum assessment and programmes of study: validity and impact. *British Educational Research Journal* **22** (1), 129 - 41.

Wood, D. (1994) *How Children Think & Learn.* Oxford, Blackwell.

Wood, R. (1987) *Measurement and Assessment in Education and Psychology.* London, Falmer Press.

Wood, R. (1991) *Assessment and Testing.* Cambridge, Cambridge University Press.

Wood, R. and Napthali, W. A. (1975) Assessment in the classroom: what do teachers look for? *Educational Studies* **1** (3), 151 - 61.

Worthen, B. (1993) Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan* **74** (6), 444 - 54.

Yosif, E. (1991) *Attempts to Develop Tools and Techniques in Educational Assessment.* Directorate of Curricula, Bahrain (in Arabic).

(The end)

APPENDIX 1

Basic Competencies For Primary Curriculum in Arabic and Mathematics.

A: The First Cycle (Year 1 - 3)
Arabic

Listening:

1. Follows the instruction and guidelines.
2. Extracts the overall meaning of an audible subject.
3. Constructs his/her own opinion from what s/he hears.

Speaking:

1. Expresses his/her own thought, wishes and experiences using correct words and correct syntax.
2. Participates in dialogue or conversation.
3. Uses the appropriate styles in social situations (greeting, acknowledgement, apologizing....).
4. Gives a short story or short news considering the right sequences of events.
5. Expresses his/her opinion in some situations.

Reading:

1. Reads aloud a short segment with the ability to vocalize the phonic symbols and the animation of meaning.
2. Reads and applies the instruction.
3. Extracts, in appropriate speed, the overall meaning and some information from a small segment.
4. Gives her/his opinion of what s/he reads.

Writing:

1. Writes with clear handwriting, correct spelling and aesthetics.
2. Applies correct punctuation (point, colon, comas and question-mark).
3. Writes answers from a segment in correct and meaningful sentences.
4. Writes questions from some subjects and stories.
5. Expresses him/herself and the things around in several integrated sentences.

Mathematics

1. Recognizes the number concepts up to 4 digits with the ability to read and write them.
2. Specifies the place value of numbers up to 4 digits.
3. Compares and ranks numbers up to 4 digits.
4. Recognizes the concepts of fractions with the ability to read and write those with a denominators from 2 to 10.
5. Compares between two fractions with equal denominators.
6. Recognizes the concepts of the four operations ($\oplus$, $\ominus$, $\otimes$ and$\oplus$).
7. Recalls the basic facts of addition and multiplication and also their reversed facts (subtraction and division).
8. Adds three numbers where the product is up to 4 digits.
9. Subtracts two numbers where the minuend is up to 4 digits.
10. Estimates the products in addition and subtraction.
11. Applies logical mathematics in addition and subtraction.
12. Recognizes some geometric concepts, points, straight line, curve and region.
13. Distinguishes the geometric figures (square, triangle, rectangle and circle).
14. Uses the metric units (gram and kilogram).
15. Measures the length of straight line using the centimetre.
16. Finds the perimeter of some geometric regions (without using formulas).
17. Makes estimations of lengths and weights.
18. Recognizes the value of local money with the ability to deal with them.
19. Solve mathematical problems involving two steps of solutions
    - Translates the problem from verbal to numerical expressions and specifies the required information.
    - Plans the solution and choose the suitable steps.
    - Carries out the solution.
    - Checks out the result of the calculations.

B. The Second Cycle (Year 4 -6)

Arabic

Listening:

1. Distinguishes between different emotions by intentions and rhythms.
2. Extracts the basic facts from audible subject.
3. Gives her/his opinion with explanation.

Speaking:

1. Expresses her/him self and describes the external incidents in correct terms.
2. Manages a dialogue or organises a conversation with friends.
3. Applies correctly the linguistic styles (questioning, asking, preferences and stipulation).
4. Uses appropriate speeches styles in different social situations.
5. Justifies his/her opinion in some situations.

Reading:

1. Reads silently different subjects and topics.
2. Reads aloud with clear, correct and fluent language.
3. Extracts facts and information from a reading segment.
4. Concludes meaning of words depending on their contexts, lexicons and her/his personal experience.
5. Understands and explains what s/he reads with the ability to give opinion.
6. Adjusts her/his reading's speed and styles to the purpose and the nature of the subject.
7. Appreciates beauties in texts.

Writing:

1. Writes with clear handwriting considering correct spelling, organization and aesthetics.
2. Applies punctuation in the right position (points, colon, comas, question-mark, and exclamation mark).
3. Writes in several different genres (autobiographies, letters, news, reportage and abstracts).
4. Writes subjects which depend on gathering information.
5. Express her/him self with correct words, moderated sentences and correct compositions.
6. Applies grammar and linguistic genres in writing.

Mathematics

1. Understands number concepts up to nine digits with the ability to read and write them.
2. Specifies the place value of numbers up to nine digits with the ability to compare and rank them.
3. Distinguishes between the even and odd numbers.
4. Analyses numbers into its prime factors (factorization).
5. Finds the highest and the lowest common factors up to three numbers.
6. Finds square root of a square number and cube root of cubic number.
7. Adds and subtracts two numbers (product up to nine digits).
8. Multiplies and divides numbers (product up to nine digits).
9. Verbally applies multiplication and division of 10, 100, 1000.
10. Understands the concepts of improper fractions with the ability to read and write them.
11. Compares fractions and fraction numbers and rank them.
12. Applies the four operations to fractions and fraction numbers.
13. Understands the concepts of fractions and decimals up to 4 decimal places with the ability to read and write them.
14. Expresses money units and decimals measures in decimal context and vice versa.
15. Applies the four operations to decimals.
16. Transfers fractions to decimals and decimals to fractions.
17. Applies approximations to numbers and fractions.
18. Estimates products of the four operations to numbers and fractions.
19. Specifies the fractions value by applying the 0, ½ and 1 as a criterion.
20. Applies some logical mathematics to four operations on numbers and fractions.
21. Understands the concepts of the ratio and the proportion and their application (the scale and proportional division).
22. Understands the basic geometric concepts and its symbols.
23. Identifies the measurement units (mm, km,..) and the time units (seconds, minutes,..).
24. Transfers from one measure to another (length and weight).
25. Measures and draws a line segment using centimetre and millimetre.
26. Measures and draws angles.
27. Understands relationships between two angles in oblique lines and parallel lines.
28. Distinguishes the attributes of quadrangle and triangle.
29. Draws triangle, circle and quadrangles (square, rectangle, rhombus and parallelogram).
30. Understands the areas, volume concepts and their units.
31. Finds the perimeter and area of a plane regions.
32. Finds the volume of cube.
33. Estimates the lengths, areas and volumes.
34. Explains tabulated data.
35. Tabulates and explains a sets of data.
36. Finds the arithmetic average.

37. Solves mathematical sentences of multiple steps.
    - Specifies the necessary data.
    - Specifies the required data.
    - Translates the problem into symbols.
    - Analyses the problem into its main components.
    - Applies the solution.
    - Check out the solution.

APPENDIX 2

The Educational Outcomes Of Primary Schools
'Summary of research result'

Wehbeh. N., Al-Mutawa. A., Helal, A.,
Al-Mannai. L. and Al-Kayat. N

1. Arabic and Mathematics:

Wehbeh and coworkers (1993) investigated the learning outcomes in Arabic and Mathematics for primary grades three and six. A criterion-referenced testS were used to assess how students' performance approached the objectives of the two subjects. The criteria were based on the stated behaviourial objectives of the syllabuses. A sample of students in third and sixth primary grades completed the test at the end of the academic year 1989/90 (see table 1).

Subject	Grade	Females	Males	Total	No. of papers	No. of Questions
Arabic	Third	192	154	346	3	19
Maths	Third	194	166	360	2	22
Arabic	Sixth	215	156	371	3	17
Maths	Sixth	225	160	385	3	19

Table 1. Criterion-referenced test in Arabic and Mathematics.

Arabic test results:

The test included items in reading and spelling and grammar. Composition skills were also examined by asking students to write 50 to 70 words for third grade and 200 words for sixth grade. Tables 2 and 4 show the contents and the results of the tests. The results indicate that:

1. A high proportion of students had not mastered the basic skills specified in the curriculum. Further, some were far below the curriculum criteria.

2. Nearly all students lacked the ability to construct extended sentences. They used only one structure of sentence, often a succession of short sentences (e.g. I went,

she came, I slept).

3. A quarter of the students were unable to present elaborated ideas.

4. Many students made frequent spelling mistakes. For example: no single word had been written correctly in the dictation by all students at both grades.

5. A high proportion of students, concentrated in some schools, gave up particular questions without any attempt to give answer.

6. In the comprehension test, nearly all students failed to go beyond the given text. Their answer reported only the text content.

7. In the free writing there were 31 and 24 unreadable papers in the sixth grade and in the third grade respectively. They contained Arabic letters but with no meaning.

Analysis of the Mathematics test results:

The test was designed to cover knowledge and skills in various mathematical domains including: numbers, measures, fractions and problem solving. Tables 3 and 5 contain the summary of tests results. The results indicate:

1. In third grade, only 54.7 percent could use the correct symbol (<, >, =) to distinguish 50 minutes and half an hour.

2. Only 47.7 percent of third grade students were able to complete the number sequence 3, 5, 7, 9,.., ...

3. Only 60 percent of sixth grade students were able to subtract 357 from 700.

4. More than half of the students were unable to find the correct strategies when solving mathematical practical problems. (see table.3, questions No.18-22 and table 5 questions No.14-19).

5. Nearly third of the students offered incomplete answer. Typically, they omitted the context or/and the unit (e.g oranges, metre).

Many questions are raised by these preliminary findings. What kind of assessment practices are going on in Bahraini schools? Is the low standard of some students related to the assessment practices? And, what types of assessment are used by teachers to promote students from year to year?

Question's aim	No of students with			
	correct sub-answer	full correct answer	full wrong answer	no answer
1. Integrate sentences to form a consistent sequence.	85	85	92	9
2. Complete gaps with appropriate words (10 given words).	277-130	74	10	9
3. Complete gaps in four sentences, with a given words.	218-137	75	46	17
4. Distinguish inappropriate word in group of words.	215-137	87	53	33
5. Give question for a given answer (five questions).	131-5	1	163	31
6. Put punctuation in a given text (ten punctuation).	226-48	2	34	19
7. Translate sentence into another formula (seven sentences).	135-28	3	111	43
8. Rearrange sentences in sequence.	45	45	141	83
9. Answer a question from a given text : what ?	176	176	89	22
10 Answer a question from a given text : why ?	53	53	270	23
11. Answer a question from a given text : why ?	36	36	282	28
12. Answer a question from a given text : How ?	46	46	263	37
13. Give an appropriate title for a given text.[1]				84
14. Extract word from a text to correspond with a given sentence (3 words).	116-92	54	145	39
15. Extract sentence from a text which correspond with a given meaning.	108	108	145	58
16. Put word in sentence (3 words).	150-132	72	84	55
17. Match opposite words.	291-195	132	8	16
18. Arrange words in alphabetical order.	92	70	133	51
19. Dictation of 44 text words	2	2		
Total number of students[2]	346			

Table 1. The result of the Arabic test for third grade (age group 8-9)

[1] No correct answer for this question.

[2] The horizontal total does not necessary correspond with the total number of students

Question's aim	No. of students with			
	correct sub-answer	full correct answer	full wrong answer	no answer
1. Read, write numbers and recognize their value.	120-96	38	136	34
2. Write five numbers between 997 and 1020.		50	100	81
3. Compare numbers, values and define their relations in <, >, or = symbol.	294-176	25	3	1
4. Specify the longer ribbon of two different lengths (300 cm and 5 m)	105	31	54	101
5. Arrange progressively 8950, 9508, 9580, 8905, 8095, 5980.	171	31	90	68
6 Complete two sequences (3, 5, 7, 9,..,..) and(..,..,89,92,95,98,..,..)	172-31	31	141	43
7. Multiple choice question: a- specify the nearest number to 150 b- specify the number equal to five hundred and seven.	152-144	88	107	39
8 Seven items in a simple number operation, four operations among them require to recognize the zero value.	252-176	69	19	14
9. Nine items in vertical number operation (addition, subtraction and multiplication with and without carrying).	280-47	6	22	14
10. Give the appropriate number in the following vertical operations: 576 - ? = 920 and ? - 460=402	81-51	31	188	70
11. a) Add the followings 3546, 294 and 80 b) Subtract 678 from 967	51-36	17	138	142
12 Choose the appropriate measure of: a- man weight b- door hight	102-92	19	102	59
13 Measure two different lengths using the ruler.	197-188	178	110	51
14 Calculate perimeter of pentagons shape with a given side-length.	100	100	101	159
15 False or true in fraction comprehension (two sub-items)	243-134	126	31	78
16 Present the appropriate fraction from a given two shaded shapes.	168-57	51	122	3
17 Choose the right operation implied in a given mathematical problem.	134	134	202	24
18. A woman bought a number of oranges. She divided them between her five children where each child have 4 oranges. How many oranges she bought.	87	97	116	33
19. A man have B.D.70. He bought cassette recorder of B.D.28, watch of B D. 35. How much money he had left?.	63	23	195	28
20. Ebraheim bought a pencil of 75 Fils, a rapper of 125 Fils, and a book of 230 Fils His change was 70 Fils. How much money he had?.	30	10	126	90
21. A workman gain monthly B.D.280. He spends monthly B.D.254. How much money he would save after working for six months?.	14	4	163	60
22. With Adel: one coins of 50 Fils, 3 coins of 10 Fils and 2 coins of 25 Fils. a- How much money he got? b- Can he buy a rubber and book of 120 Fils?, a pencil and book of 140 Fils ? and a ruler of 80 Fils ?.	20	12	88	33
Total number of students[3]	360			

Table 2. The result of the Mathematics test of third grade (age group 8 - 9).

[3] The horizontal total of students does not necessary correspond with the total number of students.

Questions' aim	No of students with			
	correct sub-answer	full correct answer	full wrong answer	no answer
1. Give appropriate title for a given text.[4]				18
2 Answer a question from a given text. Explain.	37	37	247	34
3. Answer a question from a given text. Why ?.	54	54	217	35
4. Answer a question from a given text. How ?.	133	133	191	47
5. Extract the appropriate sentence from a given text which corresponds with a given sentence.	184-127	72	109	27
6 Explain the meaning of two expressions extracted from a given text.	128-103	75	165	41
7 Change a word in a sentence with another appropriate word. 3 items	110-25	6	196	40
8. Give the opposite for a given words. 4 items.	213-101	40	81	27
9. Give the order of a given different thoughts which correspond with the sequences in the text.	122	69	160	20
10. Arrange ten names in alphabetical order.	80	168	88	35
11 Vowelize 11 words in a given text.	145-11	1	18	112
12 Use the appropriate device to change a negative sentence into: a- past tens b- future tense.	84-71	53	213	56
13. Use two particular devices to change three given sentences.	45-7	-	277	8
14 Give the appropriate answer using the numbers in a word form.	60-50	20	265	8
15. Change a given sentence to three different forms: a- feminine singular b- plural (two persons) c- muscular plural	85-62	12	212	15
16. Put punctuation in a given text (12 items).	252-24	-	42	2
17. Dictation of 61 text words.		11		
Total number of students	371			

Table 3. The results of the Arabic test for sixth grade (age group 11-12).

[4] No correct answer for this question.

Question's aim	No. of students with			
	correct sub-answer	full correct answer	full wrong answer	no answer
1. six items in proportion and percentage comprehension.	227-68	23	58	7
2. Recognize the place value of numbers, fractions, %s and proportions (10 multiple choice items).	232-96	3	2	1
3 Compare numbers and fractions and define their relations in <, >, and = symbol (10 items).	315-74		7	3
4. a- Number operations (8 items in addition, substraction, multiplication and division with carrying). b- Subtract 317 45 from 925 And add 41, 500.33 and 9. c- Rewrite the normal fraction 2/5 in percentage.	329-11	2	11	1
5. Fraction and percentage operation (8 items).	238-9	3		8
6. Find and correct error in two presented mathematical operation.	118-83	27		55
7. Knowledge and comprehension in measures : angel, length, space, volume and perimeter (9 items).	146-13	2	115	6
8 Comprehension and explanation in measures (3 items).	121-55	23		11
9. Draw the rectangle ABCD where AB= 4 cm and BC= 6 cm using the appropriate measures.	301	301	78	6
10 Draw the triangle ABC where BC= 5 cm, abc= 45 and acb= 100 using appropriate measures.	55	55	313	17
11 Complete a square following a diagonal 4 cm line.	195	195	170	20
12 Give the third angle of a right-angle triangle ABC where abc= 60.	37	47		66
13. a- Give the area of triangle with a given side measures. b- Give the area of a parallelogram with a given length and hight. c- Give the area of shaded shape.	95-84	5	237	39
14. A library sells stationary books in two ways: a- every 6 books with 360 Fils. b- every 3 books with 210 Fils. If you want to buy 12 books, which way you prefer and why?.	134	43	178	30
15. A man bought a car of B.D 5000. If he payed 30% of the price in advance and payed the remaining amount in 35 instalments. How much will be each instalment?.	104	20	199	62
16. A B.D.2480 distributed by a society between two families in 3:5. How much money each family get?.	147	25	163	50
17. A parallelogram land. Its base= 45 metre and its altitude= 20 metre. The area of this land is equal to another square land. Find the side of the square land.	131	20	175	59
18. A setting rectangle room of 6 metre length and 4 metre width furnished with a carpet. The sides of the carpet were 50 cm at a distance from the room walls. Find the area of the carpet?.	11	1	319	54
19. A meeting hall in school with 10 rows. each raw contains 10 chairs. In a meeting, the students were sitting as follows: - 4th first rows were fully occupied by the students. - 1/2 the next 6th rows were occupied . How many students were there in the hall?.	63	29	226	67
Total number of students	385			

Table 4. The results of the Mathematics test of sixth grade (age group 11-12).