



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

**THE RELATIONSHIP BETWEEN
TEACHER TRAINING IN MEASUREMENT AND
CLASSROOM ASSESSMENT PROCEDURES IN KENYA'S SECONDARY SCHOOLS**

M.A. (EDUCATION) THESIS

BY

Andrew K. Chirchir

**Thesis presented to the School of Graduate Studies of the
University of Ottawa in partial fulfilment of the requirements
for the degree of Master of Arts in Education.**



Andrew Kiptanui Chirchir, Ottawa, Canada, 1995.



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-07793-4

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGEMENTS

The writer wishes to express his sincere gratitude and appreciation to the Canadian International Development Agency (CIDA) and the governments of Kenya and Canada for funding this study; and to the Teachers' Service Commission (and the Ministry of Education) for granting the writer study leave to undertake the study.

A word of gratitude goes to Dr. Marc E. Gessaroli, the writer's advisor, for his invaluable advice, guidance, and encouragement throughout the writer's course of study; to Dr. Bruno D. Zumbo and Dr. Pierre Michaud for their advice and participation as committee members at the proposal stage; to Dr. Bradley J. Cousins and Dr. Marvin W. Boss for their advice and participation as examiners in the thesis defence. A word of gratitude is also extended to Mr. Michel Brabant, the faculty research assistant, for his tremendous help with the SPSS computer Programme for data analysis.

The writer is also indebted to professors Blaine Worthen of Utah State University and Barbara S. Plake of the University of Nebraska at Lincoln (US) for providing copies of the instruments used for developing the research instrument for this study; to the teachers who took part in the pilot phase of the research instruments, to the Kenya Government for granting a permit to undertake the study; to the Provincial Education Officer for his permission to conduct the study in the province; to the District Education Officer (Kericho) for his permission to conduct the study in the district; and to all the teachers who participated in the study for offering their time and cooperation, without which the study would not have been a success.

I do also wish to extend my sincere gratitude to my beloved wife, Jane, and my children

Daisy Cherotich and Gerald Kipkurui for their love, encouragement, and prayerful support during the entire period of my study. Most of all, I am grateful to God for opening the door for me to continue my studies. In a foreign land, He has been to me more than an earthly father can ever be to his beloved child.

DEDICATION

This work is dedicated to my daughter Karen, who because of the delay in my thesis defence, had to be born in my absence. She will grow to know that her second name has something to do with my absence at her time of birth.

TABLE OF CONTENTS

CHAPTER	<u>PAGE</u>
1 INTRODUCTION.....	1
Organization of the Thesis.....	9
2. REVIEW OF LITERATURE.....	10
Importance of Classroom Assessment.....	10
Importance of Measurement Principles in Classroom Assessment as seen by Experts.....	19
Teacher Use of Measurement Principles.....	23
Factors Affecting Use of Measurement Principles.....	26
Training in Measurement.....	26
Teachers' Perceptions of the Importance of Measurement Principles.....	29
Availability of Time.....	31
Availability of Resources.....	32
Teacher Competence in Mathematics.....	33
Class Size.....	33
Recent Thoughts about Educational Measurement.....	34
Summary of the Literature Review.....	38
3. RESEARCH METHODOLOGY.....	40
The Setting of the Study.....	40

	The Target Population and Study Sample.....	41
	The Questionnaire.....	43
	The Interview Guide.....	46
	Data Collection and Analysis Procedures.....	47
	The Mail Survey.....	47
	Semi-structured Interview.....	48
4.	RESULTS.....	50
	Teacher Training in Measurement.....	50
	Importance and Frequency of Teacher Use of	
	Measurement Principles.....	54
	Importance and Frequency of Teacher Use of Student	
	Assessment Results.....	55
	Importance and Frequency of Teacher Use of	
	Test Construction Principles.....	57
	Importance and Frequency of Teacher Use of Different	
	Types of Test Questions and Other Assessment Methods.....	58
	Importance and Frequency of Teacher Use of Test	
	Administration Principles.....	60
	Importance and Frequency of Teacher Use of Test	
	Marking Principles.....	61
	Importance and Frequency of Teacher Use of the	
	Principles for Summarizing and Interpreting Test Scores.....	62

CHAPTER**PAGE**

Importance and Frequency of Teacher Use of the Principles for Evaluating Test Quality.....	63
Importance and Frequency of Teacher Use of the Principles for Reporting Student Assessment Results.....	65
Teacher Use of Measurement Principles by Teacher Curriculum Area.....	66
Teacher Use of Measurement Principles by Teacher Qualification.....	68
Effects of Various Factors on the Use of Measurement Principles in Schools.....	68
Availability of Time.....	69
Availability of Resources.....	72
Ministry of Education Policies on Examinations.....	73
Ministry of Education Policies on Continuous Assessment of Students.....	74
School Policies on Student Assessment.....	75
Teacher's Level of Comfort with Mathematical Concepts.....	76
Teacher Workloads.....	78
Class Size.....	79
Irrelevance of Some Measurement Principles.....	80

	Wide Syllabuses.....	82
	Teacher Assessment Needs.....	83
5.	DISCUSSION OF RESULTS.....	87
	Teacher Training in Measurement.....	87
	Importance of Measurement Principles from the Teachers' point of View.....	90
	Use of Measurement Principles in Schools.....	92
	Factors Affecting Use of Measurement Principles in Kenya.....	99
	Benefits of the Study.....	101
	Limitations of the Study.....	102
6	CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH.....	105
	References.....	111
	Appendices	
	A The Research Permit.....	121
	B The Survey Questionnaire.....	123
	C Teacher Interview Guide.....	138
	D The Covering Letter.....	143
	E Teacher Frequency of Use of Measurement Principles by Teacher Curriculum Area.....	145
	F Teacher Frequency of Use of Measurement Principles by Teacher Qualification.....	154

LIST OF TABLES

		<u>PAGE</u>
Table 1:	Frequency and Percentage Distributions of Teachers	
	Trained in the Use of Measurement Principles.....	51
Table 2A:	Importance and Frequency of Teacher Use of Student	
	Assessment Results.....	56
Table 2B:	Importance and Frequency of Teacher Use of Test Construction	
	Principles.....	57
Table 2C:	Importance and Frequency of Teacher Use of Different Types	
	of Test Questions and Other Assessment Methods.....	59
Table 2D:	Importance and Frequency of Teacher Use of Test	
	Administration Principles.....	60
Table 2E:	Importance and Frequency of Teacher Use of Test	
	Marking Principles.....	61
Table 2F:	Importance and Frequency of Teacher Use of the Principles	
	for Summarizing and interpreting Test Scores.....	63
Table 2G:	Importance and Frequency of Teacher Use of the Principles	
	for Evaluating Test Quality.....	64
Table 2H:	Importance and Frequency of Teacher Use of the Principles	
	for Reporting Student Assessment Results.....	65

Table 3:	Teacher Frequency of Use of Different Types of Test Questions and Other Assessment Methods by Teacher Curriculum Area.....	67
Table 4:	Means and Standard Deviations of Teachers' Perceptions of the Relative Effects of Various Factors on the Use of Measurement Principles in Schools.....	70
Table 5:	The Effect of Teacher Level of Comfort with Mathematical Concepts on the Use of Statistical Principles by Teacher Curriculum Area.....	77
Tables of Appendix E		
Table 6:	Teacher Frequency of Use of Student Assessment Results by Teacher Curriculum Area.....	145
Table 7:	Teacher Frequency of Use of Test Construction Principles by Teacher Curriculum Area.....	146
Table 8:	Teacher Frequency of Use of Different Types of Test Questions and Other Assessment Methods by Teacher Curriculum Area.....	147
Table 9:	Teacher Frequency of Use of the Principles for Test Administration by Teacher Curriculum Area.....	148
Table 10:	Teacher Frequency of Use of Test Marking Principles by Teacher Curriculum Area.....	149
Table 11:	Teacher Frequency of Use of the Principles for Summarizing and Interpreting Test Scores by Teacher Curriculum Area.....	150

Table 12:	Teacher Frequency of Use of the Principles for Evaluating Test Quality by Teacher Curriculum Area.....	151
Table 13:	Teacher Frequency of Use of the Principles for Reporting Student Assessment Results by Teacher Curriculum Area.....	152

Tables of Appendix F

Table 14:	Teacher Frequency of Use of Student Assessment Results by Teacher Qualification.....	154
Table 15:	Teacher Frequency of Use of Test Construction Principles by Teacher Qualification.....	155
Table 16:	Teacher Frequency of Use of Different Types of Test Questions and Other Assessment Methods by Teacher Qualification.....	156
Table 17:	Teacher Frequency of Use of Test Administration Principles by Teacher Qualification.....	157
Table 18:	Teacher Frequency of Use of Test Marking Principles by Teacher Qualification.....	158
Table 19:	Teacher Frequency of Use of the Principles for Summarizing and Interpreting Test Scores by Teacher Qualification.....	159
Table 20:	Teacher Frequency of Use of the Principles for Evaluating Test Quality by Teacher Qualification.....	160
Table 21:	Teacher Frequency of Use of the Principles for Reporting Student Assessment Results by Teacher Qualification.....	161

ABSTRACT

The purpose of this study was to determine teacher use of measurement principles and the factors influencing this use in the assessment of student achievement in the Riftvalley Province (Kenya). Given that most of the assessment in the classroom consists of instruments developed by teachers, a first step in exploring the utility of measurement principles is to investigate the use of these principles in specific assessment areas. This could eventually lead to the determination and possibly the improvement of the fit between measurement training and teacher classroom assessment practices.

The study was specifically designed to provide information on teacher use of measurement principles by considering whether teachers had received training in educational measurement principles, how important they perceived these principles to be, and how often they used the principles in the assessment of student achievement. The study was also designed to determine factors influencing the use of measurement principles in schools.

Data were collected, using questionnaires and personal interviews, from 246 teachers working in public secondary schools in the Riftvalley Province (Kenya). The teachers provided information regarding their training in various measurement principles and the importance they attached to these principles. They also rated their frequency of use of these principles in the assessment of student achievement. In addition, they rated the extent to which various factors affected their use of measurement principles. Later, personal interviews were conducted with some of the respondents in order to clarify the responses and other important issues in greater depth than was possible with the questionnaires. The interviews were also used to determine teacher assessment needs.

Responses on teacher training in measurement were analysed using frequencies and percentages. Those on the importance and frequency of teacher use of measurement principles were analysed using frequencies, percentages, means and standard deviations, and t-tests. The responses from interviews were content-analysed for recurring themes.

With respect to teacher training as a whole, the results show that teachers have been trained in the principles of educational measurement. However, there is some indication that measurement training did not effectively address the assessment concerns of many classroom teachers. Teachers do not feel adequately prepared in test construction, marking, and the reporting of student assessment results.

The results on the importance of measurement principles provide some clear indication that teachers attach much importance to the principles for test construction, test administration, marking, and the reporting of student assessment results. The use of descriptive statistics, item analysis, and test reliability procedures were rated as being comparatively less important in the assessment practices of classroom teachers.

The findings with respect to the use of measurement principles show that teachers tend to use on a frequent basis those principles with high importance ratings. There were a few significant differences between history and mathematics teachers in the frequency of use of multiple-choice, short-answer, matching exercises, essay, problem-solving questions, observational techniques, and in scoring practices. History teachers reported a more frequent use of multiple-choice, short-answer, matching exercises, and essays questions than the mathematics teachers who responded to the survey. Mathematics teachers on the other hand reported a more frequent use of problem-solving questions and observational techniques than did the history teachers who responded to the

survey. With respect to scoring practices, history teachers reported a more frequent scoring of essay tests one question at a time for scoring consistency than did the mathematics teachers who responded to the survey. Mathematics teachers on the other hand reported a more frequent remarking of student test scripts for scoring consistency than did the history teachers who responded to the survey. However, there were no significant differences between diploma and graduate teachers in the use of measurement principles.

Significant differences were also found between history and mathematics teachers with respect to the effect of teacher aptitude in mathematics on the use of statistical analyses. A significant number of history teachers indicated that their level of comfort with mathematical concepts affects their use of statistical analyses.

Heavy workloads resulting from large numbers of examinees and wide syllabuses were listed by teachers as factors affecting the use of measurement principles in schools. These factors are seen as reducing the amount of time teachers can commit to their assessment activities. One other factor mentioned as contributing to the low usage of measurement principles is the perceived irrelevance of some of these principles (e.g., statistics, test-reliability procedures).

Teacher interviews revealed that teachers are overwhelmed by the demands associated with Kenya's 8-4-4 system of education. On the basis of the study findings, suggestions were made for improving teacher training in measurement and for further research.

CHAPTER 1

INTRODUCTION

Classroom teachers are currently encountering increasing emphasis on student assessment in Kenya. In this assessment, emphasis is laid on the development, administration and marking of tests, interpretation of test scores, evaluation of test quality, and communication of test results to students, parents, and guardians. Test results can also be used to inform the various school personnel who make educational decisions (instructional, diagnostic, placement, promotion) about students (Joint Advisory Committee, 1993).

The recent introduction of the 8-4-4 system of education (8 years elementary, 4 years secondary, 4 years university) and the rise in the interest shown by the public towards the results of student achievement in national examinations have had a major influence on assessment practices in the classroom. With the new system of education has come the need for teachers to assess their students on a frequent basis in order to monitor the achievement of the objectives laid out in the school curriculum.

Public interest in national examination results has resulted from the publication of national examination results in the press. The ranking of districts and schools by these results has enabled members of the public to see how well schools in their own district are doing in comparison to schools in other districts in the country. This awareness of national examination results and the rising cost of secondary school education have caused many parents and guardians to show an interest in the results of the assessment carried out by teachers in the classroom. Many parents and guardians have come to regard the results of student assessment as vital indicators of how well their children will do in national examinations.

The increase in public awareness of the national examination results has, in turn, put a tremendous amount of pressure on teachers to guide their students to achieve better results in the national examinations. As a direct consequence, teachers in many districts have initiated local efforts aimed at promoting student achievement in national examinations. The most notable of these efforts has been the formation of local examination committees which coordinate the development, the administration, and the marking of the district-wide "mock" (trial) examinations usually administered to grade-twelve students. Such trial examinations are designed to serve prediction purposes. However, they may also be used to diagnose student needs which subsequently become the focus of attention during revision time.

The introduction of continuous assessment of students and the need to prepare students for national examinations have created an expanding need for teachers to function as competent users of educational measurement principles. Although this need for teacher competence in educational measurement appears to be a fairly recent development in Kenya, it has been in existence elsewhere for a long time.

About forty years ago, experts recognized that measurement and evaluation principles were important for teachers (Noll, 1955). Noll suggested that "measurement and evaluation are a part of every teacher's responsibilities" (p. 88). Since these principles are necessary, some instruction in the fundamentals of measurement should be included in the preparation of every teacher. About a decade later, Mayo (1964) did a study in which he sought to determine the fundamental principles of measurement that should be taught to pre-service teachers. He surveyed "school principals, teachers, superintendents, college and university professors, and testing and research specialists thought to have the background to judge what the pre-service teacher ought

to know about measurement" (p. 79). This survey provided a body of prescribed content which has subsequently been incorporated into measurement courses. The main theme of these courses is that teachers need to gather objective data necessary for effective evaluation and decision making in schools (Criswell, 1992; Fennessy, 1982).

Other researchers (Green 1992; Smith, Silverman, & Borg, 1980) have also expressed support for the need to equip prospective teachers with measurement principles. In stressing the important role test results play in decision making affecting the lives of students, Green (1992) emphasized the need for teachers to be competent in test construction and in the interpretation of student test scores. The need for such principles is supported in an informal survey of teachers by Shore (1971) in which practising teachers stated that new teachers should be competent in appraisal methods.

Consequently, in keeping with the need for assessment principles by classroom teachers, some undergraduate teacher education programmes have typically required at least one course in testing and measurement (Goslin, 1967; Lambert, 1980; Onocha & Okpala, 1991; Ward, 1983). For example, the acknowledgment of the importance of educational measurement for teachers has led to the inclusion of evaluation as a topic in the American Association of Colleges for Teacher Education's Project for the beginning teacher (Merwin, 1989). In Nigeria, teacher education programmes include a compulsory course in educational measurement designed to equip pre-service teachers with the principles necessary for the continuous assessment of students (Onocha & Okpala, 1991). Similarly, in Kenya pre-service teachers take a mandatory course in educational measurement prior to their certification.

The general purpose for measurement courses is to prepare the pre-service classroom

teachers to develop and use appropriate teacher-made tests for their instructional purposes and to understand how to interpret and communicate student performance on these tests to students and parents (Onocha & Okpala, 1991; Stiggins, 1988). However, there are indications that these goals are far from being realized as evidenced by teacher failure to develop achievement tests that yield valid scores and to effectively compute and interpret test scores (Gullickson & Ellwein, 1985; Obanya, 1984; Stiggins, 1991). Further, research on testing in schools indicates that there are problems in the quality of teacher-made tests. For instance, after reviewing about 9,000 test items written by Cleveland, Ohio classroom teachers, Fleming and Chambers (1983) found that approximately 80% of the items dealt with facts and knowledge. Only a small number of the items required students to apply their knowledge. These researchers cite a need for teachers to write better test items, particularly, those that are less ambiguous and which require more of students than the simple recall of facts and information. It has also been revealed that where teachers use multiple-choice items for assessing students, they do not follow standard procedures for writing distractors (i.e., distractors must be plausible). This leads to the development of items which are often ridiculously easy (Mehrens & Lehmann, 1973).

Additional findings have revealed that the low quality of teacher-made tests has been criticized not only by measurement experts but also by students as well. Mehrens and Lehmann (1991) have reported that students sometimes express dissatisfaction with tests that are ambiguous and irrelevant. Student complaints such as "I didn't know what the teacher was looking for" and "I studied the major details of the course but was only examined on trivia and footnotes" (p.52), are not unusual. It does appear that students appreciate tests that are clearly written, focus on main ideas, and require more of them than the simple memorization of facts or

ideas. However, as Mehrens and Lehmann (1991) found, teacher-made tests do not often reflect the main objectives emphasized during instruction. Further evidence shows that teacher-developed tests are very short, that is, they contain a minimal number of items (Fleming & Chambers, 1983). These research findings provide some evidence that teachers probably do not adhere to the use of assessment principles as taught in educational measurement courses. Consequently, the resulting inadequacies in teacher-made tests cast doubt upon the quality of information obtained from such tests and the decisions based on them.

Given the importance of the decisions based on the results of teacher-made tests, the need for the effective use of assessment principles by teachers in the assessment of student achievement can no longer be ignored. In view of the deficiencies reported in teacher assessment practices, there is a need to study teacher assessment practices in order to understand the factors which influence the use of measurement principles in school settings. In addition, there is a need to determine teacher assessment needs in order to provide information needed for the planning of pre-service and in-service courses in measurement.

Previous studies have revealed that teachers need more training in assessment principles. However, few researchers have examined these needs. Farr and Griffin (1973) argue that "college teachers must understand the measurement needs of teachers in the field" (p.28) in order to equip them with the desired skills. Stiggins and Bridgeford (1985) have reiterated the need for a greater sensitivity to teachers' assessment needs on the part of the measurement community.

Previous studies on teacher use of measurement principles have been limited to facets of measurement (e.g. use of statistics in describing student scores) and have therefore not provided a concise picture of teacher strengths and weaknesses in the use of these principles. In particular,

few attempts have been made to study teacher assessment practices in order to determine the measurement principles teachers use, the major principles for them, and those used or employed only in a minor way.

While Borg, Worthen, and Valcarce (1986) and Onocha and Okpala (1991) have explored teacher perceptions of the importance of several measurement topics and have noted that some topics are considered important and others not, very little is known about these teacher attitudes. Does it mean, for example, that teachers use those measurement principles which they perceive to be important?

Although studies have been done on teacher use of certain measurement principles, some of these studies involved teachers some of whom had taken little or no course work in measurement. In addition, some of the studies also lacked focus on a specific instructional setting. Prior studies (Borg, Worthen, & Valcarce, 1986; Dorr-Bremme & Herman, 1984; Green & Stager, 1985; Gullickson, 1982), for example, focused primarily on surveys of both elementary and secondary school teachers. However, variations have been reported in the assessment practices of teachers at these two grade levels. Studies have shown that elementary school teachers tend to favour the use of non-test methodologies (e.g. observations) in assessing their students while secondary school teachers have been shown to favour testing (Gullickson, 1986). Wilson (1990), in his study of the assessment practices of teachers in British Columbia and Ontario found that elementary school teachers tended to adapt test questions from workbooks or texts. Secondary school teachers on the other hand constructed most of the questions used in assessing their students. The implication of these findings is that teachers at the elementary and secondary school levels may have differing perceptions of the importance of measurement principles. They may also show some variation in the use of these principles. In addition, factors

influencing the use of measurement principles may differ as a function of the grade level taught.

In addition, most of the studies so far reported have been limited to the North American context in which there is no mandatory requirement in most teacher-training institutions for teacher certification in educational measurement. Very few studies have been reported in other settings in which educational measurement is an integral part of the teacher-training curriculum.

Although there is some evidence to show lack or little use of measurement principles by teachers in their assessment practices, the factors influencing this use have not been determined. Nevertheless, there seems to be an assumption among measurement experts that teachers who are trained in measurement principles will use the principles in assessing their students (Lissitz & Schafer, 1987). However, the available evidence shows that teachers who have been trained in educational measurement do not use some of the measurement principles (Carter, 1984; Gullickson, 1982; Gullickson & Ellwein, 1985; Marso & Pigge, 1988).

The apparent disparity between training in measurement and the use of measurement principles has led to some caution being expressed about the usefulness of these principles in classroom settings. Hickman (1991) has cautioned against equating possession of knowledge in educational measurement with actual teacher practices in classroom settings. This caution is consistent with the view expressed by Reynolds and Menards (1980) who noted that, for some reason(s), teachers trained in educational measurement principles may fail to use the principles in classroom situations.

In view of these concerns, the purpose of this study was therefore threefold:

- (a) To determine the frequency with which teachers use the principles of educational measurement in the assessment of student achievement.
- (b) To identify the factors which affect the use of measurement principles in schools.

(c) To determine teacher-assessment needs.

In order to address these broad issues, the following specific questions were formulated and therefore guided the study:

1. What measurement principles are taught in teacher- training institutions?
2. What aspects of measurement training do teachers consider to be important to their classroom assessment of students?
3. What measurement principles are most frequently used and which are least frequently used in classroom settings?
4. Is the use of measurement principles more relevant to certain curricular areas?
5. Are there any significant differences between diploma and graduate teachers in their frequency of use of measurement principles?
6. What factors affect the use of measurement principles in schools?
7. What assessment needs do classroom teachers have?

Various studies over the years have revealed that teachers make very little use of measurement principles in their testing activities. These studies have been limited to the North American context. In view of these concerns, the present study was undertaken with four points in mind. The first was the need to contribute to the existing literature on teacher use of measurement principles which is lacking in complementary information from developing countries, Africa in particular.

The second was that the writer hoped that an understanding of the importance teachers attach to the use of different measurement principles, the frequency of teacher use of these principles, and the assessment needs of secondary school teachers would be useful to teacher trainers in Kenya. Thirdly, it was hoped that constraints affecting the use of measurement

principles in schools would be of interest to ministry of education officials and school administrators.

Fourth, the writer hoped that the study would provide information necessary for the planning and delivery of pre-service and in-service courses in educational measurement. Given that there are two distinct groups of teachers (Diploma and Graduate), any differences in teacher use of measurement principles that would be attributable to teacher qualification would have implications for teacher training in educational measurement.

Organization of the Thesis

Subsequent chapters of this thesis are organized in the following manner: The literature regarding the research and theoretical sources that formed the basis of this study are reviewed in chapter 2 while the description of the research methodology used is presented in chapter 3. The analyses and the findings from the survey data and interview responses are reported in chapter 4. The discussion of the research findings is presented in chapter 5 while the conclusion and the implications of the findings for practice and further research are presented in chapter 6.

CHAPTER 2

REVIEW OF LITERATURE

The literature pertinent to the present study is presented in this chapter. In the introduction of this chapter, the importance of classroom assessment in decision making by teachers and students is discussed followed by a review of studies on the importance of measurement principles as seen by experts. A review of studies on teacher use of measurement principles forms the main part of the chapter. The factors likely to affect the use of measurement principles appear in the latter part of this chapter. The synthesis of the literature review is presented at the close of the chapter.

Importance of Classroom Assessment

Student assessment is considered to be one of the most important and complex activities carried out by classroom teachers (Stiggins, Conklin, & Bridgeford, 1986). In a recent remark by a policy maker about testing in general, Cooke (1994) stressed the important role played by testing both for students and educators:

"Testing is a part of teaching. It's not all of it. But it's an integral part of the process. It's a healthy thing to do- for students but also for the system. It tells students where they've been and what they need to do to move ahead. And it helps educators adapt the system to meet the student needs".

(Dave Cooke, Ontario Minister of Education and Training, 1994)

Although researchers have not given much attention to classroom assessment in the past, there is a growing realization among educators that attention should now focus on improving this form of assessment. Stiggins (1990) notes, "...we have made a complete transition from caring totally about standardized tests to caring far more about better classroom assessments and better classroom-level decision making" (p.92).

As a further manifestation of the growing interest in classroom assessment, representatives

from the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association recently met and developed standards that outline the assessment competencies deemed necessary for good teaching (National Council on Measurement in Education, 1990). Five of the seven standards focus directly on classroom-based assessment decisions. Similar efforts aimed at improving classroom assessment have been reported in Canada. In a recent meeting of the representatives of the Canadian Education Association, Canadian School Boards Association, Canadian Association for School Administrators, Canadian Teachers' Federation, and other professional bodies with interest in education, principles that guide student assessment in Canada were developed (Joint Advisory Committee, 1993). These principles embrace the effective use of assessment procedures by teachers in the development and use of tests in the classroom.

The recognition of the importance of classroom assessment represents a shift in position by those who have long emphasized standardized testing and other large-scale testing programmes. Such a recognition of the importance of classroom assessment will probably enable the measurement community to understand assessment from the perspective of the practitioner.

The available information on classroom assessment shows that teachers frequently assess their students and that they are supportive of test use in the classroom (Mehrens & Lehmann, 1987). There is also a clear indication that teachers rely on their own assessments as the primary source of information on student achievement (Mehrens & Lehmann, 1987).

Research findings have shown that teacher-developed tests and classroom observations are the most common methods of assessing students (Dorr-Bremme & Herman, 1984). In fact, as Newman and Stallings (1982) observed, the majority of the teachers they surveyed constructed more than half the number of tests used in their classes. Reynolds and Menards (1980) found that

teachers constructed an average of 73.68% of their test questions. However, there are cases in which teachers develop all the tests used in assessing their students except for examinations designed for certification purposes.

It has also been revealed that once teachers have developed test items, they tend to reuse them for testing future students. For example, Gullickson (1982) found that 84% of the South Dakota teachers he surveyed reported reusing test items in future tests with different groups of students.

The wide spread use of teacher-made tests in the classroom has been documented by research. Anderson (1989) and Mehrens and Lehmann (1987) have estimated that a student may take as many as 400 to 1,000 teacher-constructed tests prior to high school graduation. This heavy emphasis on the use of teacher-made tests may be a reflection of teacher satisfaction with tests for which they have a direct control. Mehrens and Lehmann (1987) opined that teacher satisfaction with the use of their own tests may well be due to the increased control that teachers exercise over the content of those tests used for instructional purposes. Other researchers (Dorr-Bremme & Herman, 1984) have noted that the content of classroom tests can be designed to suit a teacher's particular instructional emphases. These tests can also be administered when they are deemed most appropriate and fit particularly well with the teachers' instructional schedules.

Although teachers at the elementary and secondary school levels have been shown to have a marked preference for the use of their own tests, there is some evidence, however, to suggest the predominance of teacher-developed tests at the secondary school level as compared to elementary school level (Dorr-Bremme & Herman, 1984). This evidence seems to lend some support to the notion that secondary school teachers have more control over classroom assessment than do teachers at the elementary school level.

Elementary school teachers are seen to have access to the use of texts or work books with readily available tests that can be adopted for use in the classroom (Dorr-Bremme & Herman, 1984). Further, it has been observed that assessment in the elementary school level may involve the use of non-test methods such as teacher observations of students (Stiggins & Bridgeford, 1985). It would follow that on the basis of these findings, elementary school teachers do not have to construct as many classroom tests as their colleagues in high school. Secondary school teachers, on the other hand, tend to favour testing and have to develop many of the tests used for assessing their students.

Gullickson (1982) found that 85% of the elementary teachers in South Dakota constructed their own tests compared to 96% of the secondary school teachers. In a similar study involving teachers in British Columbia and Ontario, Wilson (1990) found that 32% of the elementary school teachers constructed the tests used in their classes compared to 52% of the secondary school teachers.

It has also been established that, along with the preferred use of their own tests, teachers show a marked preference for the use of certain types of test items. Newman and Stallings (1982) reported that objective test items are the most popular among teachers. Only a fifth of the teachers they surveyed reported using essay items. These researchers studied testing among elementary and secondary school teachers and found that the most commonly used items as ranked from most used to least used are completion, multiple-choice, matching, true-false, short-answer, calculation, and essay types. A similar study of teacher-testing practices by Gullickson (1982) yielded similar findings. In contrast, Reynolds and Menards (1980) found that secondary school teachers favoured the use of short-answer, multiple-choice, essay, matching, and true-false items in that order.

The variations in preference for the use of certain types of test items are not limited to differences in grade level taught. Variations in the use of item types also exist across curriculum areas. Gullickson (1982) found that teachers in the social sciences showed a greater preference for the use of true-false items than did those teachers in the sciences or language arts disciplines (83% vs 69% and 55% respectively).

Studies on testing in schools have also revealed that some curricular areas are more test-oriented than others. Mathematics, for instance, is considered to be more heavily test-oriented than other curricular areas (Wahlstrom & Danley, 1976). In fact, Fennessy (1982) found that the frequency of assessment was highest in mathematics followed by reading and written expression.

In addition, testing in schools has been shown to be a time consuming activity. In particular, the preparation of classroom tests takes a substantial amount of teachers' time. For example, about 32% of the teachers surveyed by Newman and Stallings (1982) estimated that they spent between 11% and 20% of their professional time on teacher-made tests. Other studies (Gullickson, 1982; Stiggins, 1988) have shown that on the average, teachers spend as much as 40% of their professional time in the classroom dealing with assessment and measurement-related issues. Gullickson (1982) noted that some teachers spent more than nine hours on an individual test. Much of this time is spent designing, developing, administering, and scoring formal classroom tests.

Meanwhile, students also spend a considerable proportion of their school time taking teacher-made tests. Dorr-Bremme and Herman (1984) reported that a typical student in the upper elementary grades spends about ten hours on the average taking reading tests and a little more than twelve hours a year taking mathematics tests. A typical tenth-grade student in an English class spends close to twenty- six hours a year writing English tests (20% of class time). In

mathematics, such a tenth-grade student spends about twenty- four hours taking tests annually.

Research has also been done on test administration in schools. Gullickson (1982), for example, has reported that test administration in schools is a formal, constrained activity in which students expect to be graded. Virtually all teachers (99%) do not allow student interaction during the course of the test. Students are supervised during the course of their tests and are not allowed to use any support material like textbooks and notes in completing the tests. It does appear that teachers are concerned about the need to ensure that student test scores are not clouded by factors external to the test. This is particularly important since the results from classroom tests provide important information needed for decision making.

Although the specific purposes of student assessment and the intended uses of the results may differ from one school to another, assessment results provide important information for the student, teacher, parent, counsellor, and other educators (Mehrens & Lehmann, 1991). Bejar (1984) argued that classroom tests are essential if we wish to provide for optimal learning on the part of the pupil and optimal teaching on the part of the teacher. Mehrens and Lehmann (1991) have emphasized that teachers have an obligation to provide their students with the best instruction possible. Hence, teachers must have some procedure whereby they can reliably and validly evaluate how well their students have learned what has been taught. Classroom tests fulfil such a function and are particularly appropriate since they can be tailored to suit a teacher's instructional objectives (Bejar, 1984).

Student scores on classroom tests form the most frequent measures of student achievement and are subsequently interpreted as a direct reflection of teacher skill and competence in classroom instruction (Hickman, 1991). As a means of feedback, test results provide the teacher with the information necessary for the provision of a more appropriate

instructional guidance for individual students as well as for the whole class.

In emphasizing the importance of assessment results, Dorr-Bremme and Herman (1984) have outlined a number of instructional benefits that are likely to result from the effective assessment of student achievement. They have stressed that student assessment results should assist the teacher in setting, refining, and clarifying realistic goals in teaching. Assessment results may also help the teacher reach a decision on what to teach and how to teach it to students of different achievement levels. By using these results, teachers may be able to know how students are progressing and how they (teachers) can appropriately adjust their teaching to meet the needs of their students. It appears that "linking testing and instruction is a fundamental and enduring concern in educational practice" (Burstein, 1983, p. 99).

It has also been found that many of the other decisions made in schools today are based on the information obtained from teacher-made tests (Dorr-Bremme & Herman, 1984; Stiggins & Bridgeford, 1985). In a nationwide survey of over 1,000 elementary, secondary school teachers, and principals, Dorr-Bremme and Herman (1984) found that next to teacher observations and opinions, teacher-made tests were considered the most important source of information for decisions on initial student placement, changing a student from one group or curriculum to another, and grading.

Gullickson (1982) found that teachers used test results for three main purposes in the classroom. These purposes, in order of importance, include providing instructional feedback for student learning, evaluation of instruction, and grading of students.

In a related study involving 228 teachers in the United States, Stiggins and Bridgeford (1985) found that teachers relied on their own tests more heavily than on any other single source of information for making grouping, grading, evaluating, and reporting decisions affecting their

students. In particular, each teacher was asked to distribute 100 points across four types of assessments in order to convey the relative importance of each in decision making. These assessment types included teacher-made objective tests, standardized tests, structured performance assessments, and spontaneous observations. Although the study did show some distinct differences in teacher responses as a function of purpose of assessment, curriculum, and grade level taught, teachers did allocate an average of 34 points to their own objective tests and 26 points to structured performance (preplanned observation and judgement strategies). They also allocated 21 points to spontaneous observations and 19 points to published tests (curriculum-embedded and standardized tests).

An earlier study by Salmon-Cox (1981) yielded similar findings. This author reported that 42% of the 87 high school teachers she interviewed used their own tests as a primary source of information in student evaluation. Thirty-percent of the teachers used interaction, 21% relied on home work, 6% used observation, and 1% used standardized tests.

Although teachers at the elementary and secondary school levels do make a substantial use of assessment results in decision making, there are differing uses of assessment results as a function of the school level taught. Goslin (1967) found that primary school teachers tend to use assessment results for the purpose of identifying individual differences among pupils and for tailoring the curriculum to meet these differences. In contrast, secondary school teachers tend to use test results as a means of providing feedback to students. The intention is to enable students to use test data in planning for further courses or for assessing their career prospects.

Given the instructional benefits of student assessment, Frisbie (1988) has argued that teachers should ensure that test scores are reliable before they are used to make instructional decisions. He further contends that where scores cannot be relied upon as accurate reflections of

student achievement, they cannot be used to make sound instructional decisions or to communicate meaningful achievement results to students and their parents.

Decision making based on student assessment results is not something that is unique to teachers only. Students also have been found to base their decisions on the results of teacher-made tests (Stiggins, 1990). They set expectations and decide how to relate to school on the basis of test data provided to them by their teachers. Well-designed tests are useful for self-diagnosis by students, assisting them in identifying specific strengths and weaknesses (Hopkins, Stanley, & Hopkins, 1990). Stiggins (1990) notes that students stand to benefit when teachers provide them with accurate assessment data. With such data, students are more likely to make sound decisions regarding their studies. However, Stiggins (1990) also points out that students are more likely to make unsound decisions if they receive inaccurate assessment data from their teachers. Dornbush and Natriello (cited in Crooks, 1988) noted that students were less likely to work hard if they thought the evaluation of their work did not accurately reflect the level of their performance and effort.

Test results also affect students in many other ways. Test results guide their judgement of what is important to learn, affect their motivation and self-perceptions of competence, structure their approaches to, and timing of personal study (e.g., spaced practice), consolidate learning, and affect the development of enduring learning strategies and skills (Crooks, 1988).

Gullickson (1984) surveyed a stratified random sample of 391 third and tenth-grade teachers in South Dakota in order to document the possible effects of tests on students. His findings revealed that teachers are in agreement regarding the effects of teacher-made tests on students. In particular, teachers agree that "tests increase student effort, create competition, improve student interaction, affect student self-concept, and in general, improve the learning

environment" (p. 247).

Importance of Measurement Principles in Classroom Assessment as seen by Experts

Given the importance of the decisions based on student assessment results, assessment deserves very careful planning and considerable investment of time from educators. In particular, the construction, administration, marking, the interpretation of test results, and the evaluation of test items require the efficient use of assessment principles by teachers in order to ensure that sound test data are generated. Educational measurement curriculum is ostensibly designed to equip teachers with the principles necessary for these assessment activities.

A number of test construction principles are recommended in measurement courses. Various authors (Gronlund, 1981; Marshall & Hales, 1971; Sumner, 1987) have stressed that teachers should ensure that tests are constructed on the basis of well developed test blueprints or tables of specifications. The test blueprint outlines the content of the areas to be tested and enables the classroom teacher to match this content to the objectives covered during instruction. It provides greater assurance that the test will provide some balanced measure of the intended learning outcomes emphasized during instruction and that no important objective or content area will be inadvertently omitted (Mehrens & Lehmann, 1991). Gronlund (1981) argues that unless test preparation is guided by a test blueprint, there is usually a tendency on the part of the teacher to overload the test with items measuring low level cognitive skills and to neglect the more complex learning outcomes.

Measurement courses also provide specific guidelines for the construction of true-false, matching, short-answer, multiple-choice, and essay items. The overriding concern in the development of all types of items is the need to ensure that the items are free of technical flaws such as excessive wording, ambiguity, and bias. Mehrens and Lehmann (1991) insist that every

test should be reviewed critically by other teachers in order to minimize technical flaws and inadequate content sampling.

Although teachers may construct high quality tests, errors can easily be introduced into the scores by poor test administration procedures. A critical aspect of test administration is the need for teachers to specify the scoring criteria that will be used in evaluating student responses. Popham (1990) stresses that students should be told whether such factors as spelling, penmanship and punctuation will count during scoring. In a mathematics test, for instance, students should be told whether they will receive part scores for showing a correct procedure even though they may have obtained an incorrect answer. Mehrens and Lehmann (1991) argue that pupils can respond appropriately to test items when they know the "ground rules".

Other researchers (Bushway & Nash, 1977; Gronlund, 1981; Popham, 1990; Hopkins, Stanley, & Hopkins, 1990) acknowledge the seriousness of cheating by students during testing and indicate that it results in test invalidity. Testing conditions have a great influence on the extent to which cheating will occur during test administration. Responsible test administration requires that the opportunities for cheating be reduced as much as possible through conscientious student monitoring and seating arrangements. Cheating during testing may be in the form of students using their notes to complete their tests when they are not supposed to, or seeking help from other students.

It is also important to note that no matter how well tests are constructed and administered, poor marking procedures can invalidate the results from tests. One factor that may affect the validity and reliability of test scores is the inconsistency in the teacher's scoring of short-answer and essay items. The scoring of these types of items is susceptible to a considerable amount of subjectivity. This subjectivity may reflect the scorer's idiosyncrasies which are extraneous to the

original purpose of the test, but which can make a student's score on a single test to vary from one examiner to another and from scoring to scoring by the same examiner (Popham, 1990). The student's handwriting, fluency in the language of communication, and the scorer's prior knowledge of a particular student's ability may affect the level of objectivity in the scoring of essays (Briggs, 1970; Chase, 1968; Marshall & Hasel, 1971; Popham, 1990). Markham (1976) found that student test papers with better handwriting consistently received higher scores than those with poor handwriting regardless of the quality of content. Apparently, teachers must make a conscious effort to minimize the influence of handwriting in the appraisal of student responses in tests or examinations.

There are a number of marking principles which have been shown to reduce the amount of subjectivity that goes into the marking of short-answer and essay test items. It is recommended that subjective test items should be scored on the basis of a clearly defined marking scheme which indicates the relevant points the examiner should look for in a student's response (Popham, 1990). Such a scheme should enable an examiner or several examiners to score all the student responses to an item from a single frame of reference. This measure helps to reduce fluctuations in scoring.

In addition, teachers are advised to evaluate essay responses anonymously (Popham, 1990). The underlying rationale seems to be that the teacher's knowledge of a particular student's ability or name will affect the appraisal of that student's responses. In some cases, a student's response to the first essay item may introduce some bias in the way subsequent items are scored. To curb this problem, teachers are advised to score all students' answers to one question before scoring the responses to the next question (Marshall & Hales, 1971). This measure is designed to ensure that there is little variation in the teacher's marking consistency.

Teachers are also advised to make appropriate comments on students' scripts while

marking. Such comments are designed to help the students learn from the mistakes made in answering questions (Popham, 1990).

Statistical analysis of test scores forms an integral part of measurement courses. The use of descriptive statistics such as means and standard deviations are considered important in the summarizing of student test scores (Hopkins, Stanley, & Hopkins, 1990). A knowledge of the mean test score is considered important for individual students who may wish to interpret their achievements with respect to the rest of the class.

Item analysis also receives a substantial emphasis in measurement courses (Ebel & Frisbie, 1986; Mehrens & Lehmann, 1991). It is suggested that teachers should check the difficulty and discrimination indices of the test items used for assessing their students. This analysis provides information on how well the items served to discriminate among students of high and low abilities. It also helps to show the items which were very difficult or very easy for the examinees in question. Such information may be useful in test revision and in the discussion of test results with students.

Mehrens and Lehmann (1991) stress that reliability should be determined particularly for major assessments such as end of term examinations. This is because the results from such examinations are usually weighted heavily in determining the students' end of term grades. Given that these grades are likely to be used for making important decisions affecting students, high test score reliability is required.

Test reliability is an important aspect of educational measurement. Frisbie (1988) suggests that teachers need to determine the internal consistency of their tests (split- half reliability). Other methods of determining test reliability, such as the test-retest, alternate form, and alternate test-retest methods are considered as being somewhat demanding for teachers especially where

computers are not readily available. However, Marshall and Hales (1971) take the rather distinct position that the determination of the latter forms of test reliability would not constitute an efficient use of teacher effort and student time. This is because each of these forms of reliability requires the administration of two tests.

It is the expectation in assessment programmes that once student results become available, they should be used for various purposes by teachers, parents, students, and guidance and counselling personnel. The results should therefore be communicated to the parties concerned in a manner that can easily be understood.

Hopkins, Stanley, and Hopkins (1990) note that most parents want to know how well their children are doing in terms of achievement tests in relation to other students and how much industry and effort they are putting into their studies. These authors stress the need for the use of parent-teacher conferences in the dissemination of student grades but acknowledge the limitation imposed by scheduling problems and the time required for such conferences.

Given the emphasis placed on the use of measurement principles by measurement experts, how evident are these principles in teacher assessment practices?.

Teacher Use of Measurement Principles

In this section, teacher use of measurement principles is discussed followed by a discussion of factors which affect the use of these principles in schools.

Studies on teacher testing practices reveal very little evidence of teacher use of measurement principles in schools. Three in-depth studies regarding the characteristics of teacher-made tests have been carried out by Fleming and Chambers (1983). Their analysis of teacher-made tests have revealed that there are problems in the quality of these tests. Almost 80% of all questions were at the low cognitive level. Teachers preferred to use short-answer questions most

often in their tests and developed few questions to test students on behaviours linked to the application of knowledge. Similar findings have been reported by Gullickson (1982).

In addition, Fleming and Chambers (1983) found that instructions and statements of point values for test questions were frequently absent and between 15% and 20% of all tests exhibited grammatical and spelling errors. Contrary to expectation, teachers of English generally avoided essay questions. In another study, Haertel (cited in Crooks, 1988) found that classroom examinations often failed to reflect the teachers' stated instructional objectives, and frequently required little more than repetition of material presented in the textbook or class or solution of problems much like those encountered during instruction.

Gullickson and Ellwein's (1985) study of teacher testing practices provides further evidence of lack of test quality control strategies by teachers. These researchers examined whether teachers implemented prescribed guidelines in carrying out the post hoc analyses of their tests, and whether teacher practices were related to their preparation in measurement, teaching experience, grade level taught, and curricular area of instruction. Subjects were 323 elementary and secondary school teachers, 90% of whom had taken a mandatory measurement course for their certification. The subjects included teachers of grade three, teachers of science, social science, and language arts at the seventh-grade, and teachers of biology, world history, and English II at the tenth-grade.

The results showed that a few of the teachers computed summary statistics needed to evaluate test performance. Ten percent of them reported using the mean, median, and standard deviation in describing student test scores. Although 30% of the respondents reported using item difficulty and reliability, only 1.2% of them knew the procedure for computing test reliability.

In addition, there were no differences in teacher practices due to experience and grade

level taught. However, language arts teachers, contrary to expectation, reported a significantly greater use of item difficulty and reliability than either the science or social science teachers. This unexpected observation arose perhaps due to the inadequate understanding teachers have about item difficulty and test reliability (Gullickson & Ellwein, 1985).

Teacher failure to use descriptive statistics and item analysis techniques has been reported in other studies. Of the 326 Ohio teachers surveyed by Marso and Pigge (1988), most reported little or no use of test means and standard deviations. These teachers also reported little or no use of test reliability and never did item analyses of their tests. Gullickson (1982) found that many teachers did not take time to improve their tests and usually reused items without a careful item analysis. In addition, about 42% of the teachers reported the range of test scores. About 10% to 13% of the teachers used the mean, median, and standard deviation. Thirty-one percent of the teachers reported the use of item difficulty whereas 29% reported use of test reliability information.

In a related study, Stiggins and Bridgeford (1985) reported that teachers do not adhere to procedures which would likely promote valid and reliable assessment of students. These researchers reported that teachers do not clearly articulate and communicate scoring criteria to students, define acceptable levels of performance, and do not check judgements against other data (e.g. test scores). However, it was found that teacher attention to test quality control increased as grade level increased.

Carter (1984) studied teacher test development skills among 310 high school reading and language arts teachers in the US. She found that teachers had difficulty recognizing items written to measure higher order thinking skills. In addition, teachers felt quite insecure about the knowledge of good item writing practices. When presented with item faults such as item cues,

teachers indicated that their own classroom tests were characterized by such technical inadequacies. Other researchers (Gronlund, 1976; Nunally, 1978) have also noted that teachers are unfamiliar with item writing principles.

Carter (1984) found that teachers expressed the desire to write items that would provide them with the best diagnostic information about their students. However, they felt quite inadequate to do so because their training did not provide instruction in this area. One teacher put it best when in reference to this he summarized:

"I am always able to give my kids grades based on what they do in my class and on their test grades but I'm not sure how much I believe these grades give me diagnostic information...I am less able to pinpoint weak areas where individual students might need help" (Carter, 1984, p. 59).

Carter (1984) also found that teacher test items were based on the paraphrasing of lecture notes. Teachers reported that they usually spent very little time revising or editing their test items apparently because of the demands on their time.

Factors Affecting Use of Measurement Principles

A number of factors have been suggested to explain the little use of measurement principles by teachers. These factors are outlined in the paragraphs below.

Training in Measurement

Those who have assessed teachers' knowledge of assessment practices have consistently voiced concern that teacher knowledge and skills in measurement are inadequate due to lack of training (Goslin, 1967; 1987; Mayo, 1964, 1967; Roeder, 1972). This is not surprising given that some teacher training institutions in North America do not offer mandatory measurement training for teachers. Goslin (1967), in an exploratory survey involving 1,450 secondary school teachers in 75 schools and 89 elementary teachers in 16 schools in the United States found that 20% of the

public secondary school teachers had never taken a course in educational measurement either at the undergraduate or at the graduate level. About 25% of these teachers had taken one course in educational measurement. Twenty-five percent of the elementary teachers had never taken a course in tests and measurement. Roeder (1972) surveyed 860 teacher education institutions in the U.S. and reported that 270 (31%) required prospective elementary teachers to complete a measurement course. Although some of these studies are fairly old, recent studies on the state of measurement training show that the situation has not changed much. In one of the most comprehensive studies on teacher training in measurement, Lissitz and Schafer (1987) found that measurement has not been given the emphasis that its role in teaching deserves. These researchers surveyed the measurement training practices of the institutions affiliated with the American Association of Colleges for Teacher Education. They found that less than half of the teacher education programmes had a formal measurement course as a graduation requirement. In another recent study, Wise, Lukin, and Roos (1991) found that the percentage of education programmes requiring a measurement course has risen. However, they found that the majority of the educational programmes still do not require a formal course in measurement.

In other related studies (Carter, 1984; Fluit & Gifford, 1980), many teachers have argued that the topic of test construction has not been addressed sufficiently in undergraduate teacher-training programmes and that programmes are needed to help them develop, manage, and provide practice for evaluation. Similar findings have been reported by Farr and Griffin (1973).

Dorr-Bremme and Herman (1984) found that teachers received little training devoted to increasing their routine classroom assessment skills. According to the principals who were surveyed, there were few staff development activities in the area of assessment. Whenever these services were provided, they focused heavily on the analysis and explanation of test results,

directions for administering tests, and interpretation and use of the results of different types of tests. Further, 35% of the elementary teachers reported having received training in the use of test results to improve instruction. About 20% of the secondary school teachers indicated having received staff development in topics related to testing less often than elementary teachers did.

Other researchers (Coffman, 1983; Stiggins, 1990; Ward, 1982) have reported that many teachers have had no formal course work and that most of them have had no in-service training in measurement. Newman and Stallings (1982) reported that only about 25% of Florida's teachers have had some adequate training in measurement preparation. About 10% to 20% of the teachers surveyed by Hills (1977) in Florida could answer basic questions on classroom measurement principles.

Teacher lack of training in measurement seems to be supported by the findings from Green and Stager's (1985) study in which 555 elementary and secondary school teachers expressed a need to upgrade their skills in testing and measurement. About 47% of the respondents expressed an interest in in-service training with a focus on test reliability and validity, while 42% of these teachers showed an interest in learning more about the construction of classroom tests. Other researchers (Carter, 1984; Fluit & Gifford, 1980; Stiggins & Bridgeford, 1985) have found that teachers are in support of programmes designed to upgrade their skills in test construction, implying that they lack the requisite knowledge in this area. This need is not surprising given the heavy use of teacher-made tests in classrooms and the fact that these tests frequently contain item construction errors (Marso & Pigge, 1988).

Newman and Stallings (1982), however, have reported that there has been a significant emphasis on required training in measurement in recent times. However, they note that despite this new development, the problem of inadequate teacher preparation in measurement has not

been resolved. This is further complicated by the fact that teachers who have taken courses in measurement do not use some of the principles taught in these courses (Gullickson, 1984; Gullickson & Ellwein, 1985; Marso & Pigge, 1988).

Teacher failure to use measurement principles has led to the questioning of the importance of these principles in the assessment of student achievement. How important are these principles in the assessment practices of classroom teachers?

Teachers' Perceptions of the Importance of Measurement Principles

Given that some studies have shown that teachers are not knowledgeable about basic measurement concepts (Mayo, 1967; Rudman et al., 1980), it would seem logical to assume that these teachers would take the opportunity to pursue measurement training whenever this opportunity becomes available. However, the unpopularity of measurement courses among pre-service teachers (Gullickson, 1986; Stiggins & Bridgeford, 1985) and the avoidance of these courses (given a choice) by teachers raises some serious concern. Gullickson (1986) found that student registration in measurement was low (25% or less) when the course was optional and that no more than 50% of the required students took the course. This researcher contends that one reason for the avoidance of measurement courses lies with the failure of the training to address important teacher needs. This view is shared by other Farr and Griffin (1973) who have argued that measurement courses tend to overemphasize standardized tests and that little attention is devoted to classroom tests designed by teachers.

A number of studies have been done to establish the importance of measurement principles in the assessment of student achievement. Borg, Worthen, and Valcarce (1986) studied teachers' perceptions of the importance of topics and concepts taught in educational measurement courses. They surveyed a random sample of 1000 elementary and secondary teachers in Utah. Of the 479

teachers (48%) who responded to the survey, 70% indicated that relating tests to instructional objectives, analyzing tests in order to improve them, and using tests to diagnose student needs were very important principles for a teacher's routine assessment activities.

Over 50% of the teachers rated several aspects of measurement as being moderately important. These aspects include measurement use in the schools, interpretation of test norms, test validity and reliability, factors that can distort students' test scores, and measuring student aptitude and readiness. Statistics used to describe scores were rated as being least important in the assessment practices of classroom teachers.

However, significant differences were found between secondary and elementary teachers in the importance attached to a number of measurement topics. A significant proportion of the grade 7-12 teachers attached more importance to the construction and use of essay, multiple-choice, short-answer, sentence completion, and true-false test items than did teachers at the elementary school level. A larger proportion of the teachers in grades 4-6 rated these types of test items as being very important than did the teachers who taught the lower grades.

In a similar study, Onocha and Okpala (1991) investigated practising teachers' perceptions of the importance of topics taught in educational measurement courses in Nigeria's teacher training institutions. Subjects were 1500 elementary teachers who had registered for the 1988/89 in-service programme in education at the University of Ibadan. These teachers had taken a compulsory course in educational measurement as part of the requirement for their training. Sixty percent of the teachers indicated that the role of measurement and evaluation in teaching, determination of validity and reliability, relating test scores to instructional objectives, and constructing and using essay test items were important assessment principles required by classroom teachers. Also considered as being important were factors distorting students' scores,

using tests to diagnose students' needs, test administration, marking, and reporting procedures.

Statistics used to describe test scores, the interpretation of different kinds of scores, constructing specialized tests, constructing questionnaires and rating scales, use of observational techniques, and the interpretation of test norms were rated as not being important by more than 50% of the teachers. An earlier study by Mayo (1964, 1967) yielded similar findings.

Despite teacher resistance to the use of statistical analyses in dealing with student assessment, more recent specialists have been careful to argue for the merits of these analyses in the improvement of teacher test construction skills. Blessum (cited in Crooks, 1988) concluded that the use of item analysis resulted in an improvement not only in the quality and fairness of each individual examination, but also in the educational quality of successive tests. It has also been argued that the analysis makes a teacher become increasingly sensitive to such nagging problems as test ambiguity, ineffective distractors, and poor wording.

Further arguments in favour of item analysis have been advanced. It is believed that the benefits of item analysis may help to reveal which teaching strategies were most effective, allowing teachers to emphasize the strategies in future. Gronlund (1981) argued that the benefits of item analysis may transcend the improvement of the test itself. It may provide a sound basis for the efficient class discussion of test results, remedial work, and increased skill in test construction.

Availability of Time

After observing that teachers do not compute summary statistics of their test results and that they do not consider item analysis as a helpful strategy in improving test quality, Gullickson (1984) noted that many teachers may not perceive item analysis as being practical in the classroom setting. He accepts the view that teachers are faced with serious time constraints arising from their busy instructional schedules and other school commitments.

Densmore (1989) and Smith and Adams (1972) also acknowledge that teachers have a lot of work to do. Densmore notes that out of a sense of professional dedication, teachers often volunteer for additional responsibilities including after school and evening activities. As a result, most of the teachers who have taken a course in measurement may not have enough time to put into practice the principles of test construction and interpretation which they have learned.

Other researchers (Mehrens & Lehmann, 1991) have expressed similar concerns about the issue of lack of time in the assessment of student achievement. These researchers have acknowledged that the process of developing high quality test items is a demanding one and that it requires a substantial investment of time and effort. Thus, when faced with time constraints, teachers may choose to use only those measurement principles which have a direct bearing on instruction. Principles which do not have any instructional value from the teacher's point of view may be ignored. Gullickson (1986) notes that teachers may perceive test analysis as calling for more work than is justified by its benefits. He goes on to insist that this analysis can be omitted without any obvious negative effects in teacher assessment practices (i.e. a test may be administered, scored, graded, and feedback provided without such analysis).

Stiggins and Bridgeford (1985) insist that "teachers simply do not have the time to measure a student and then wait for several weeks while the computer transforms the result into a percentile or grade equivalent score" (p.69). These researchers argue that in order to make the decisions that will facilitate instruction, teachers need data immediately and have to gather it from a constantly changing environment.

Availability of Resources

According to Hambleton, Anderson, and Murray (1983), lack of resources such as computers may partly account for teacher failure to carry out item analysis. These authors take

the rather distinct position that improved teacher assessment practices may result from the introduction of microcomputers into the testing programme in schools. The underlying assumption is that computers will make it possible for teachers to determine the reliability of their tests and to carry out item analysis without having to spend much time doing so.

Teacher Competence in Mathematics

Researchers who have assessed teacher competency in the use of the knowledge of statistics (Rudman et al., 1980) have indicated that teachers do not have a good grasp of statistical concepts. This suggests that pre-service measurement courses, despite their relatively heavy emphasis on statistical computations, do not necessarily result in a level of understanding that would lead to a comfortable use of statistical concepts by teachers in the analysis of student test results (Gullickson, 1986). However, it is also known that statistical analysis involves a substantial amount of computations which may require some background in mathematics. Although Kubiszyn and Borich (1993) have suggested that some good background in mathematics may not be necessary in the understanding and use of statistical techniques, this view is subject to question. Tests yield numerical scores. Since test theory is based on statistical principles, teachers who feel weak in mathematics may have great difficulty in applying the principles in classroom settings. Nimmer (1984) argues that quantitative measures of reliability, validity, and item analysis may be frustrating to those teachers who have a limited background in mathematics.

Class Size

Shephard (1989) acknowledges that practical considerations dictate test format in ways that further constraint content. For example, given the huge number of examinees and time limits, teachers use multiple-choice formats to the exclusion of tasks that would require students to

produce or explain a correct answer. Essay tests of subject matter (where appropriate) are rare. Although multiple-choice items can elicit important conceptual distinctions from students, tests composed entirely of such items do not measure a student's ability to organize relevant information and present a logical argument (Shephard, 1989). Shephard also opines that time limits and the need to survey as much content as possible do not facilitate problems aimed at the interrelation of facets of the same topic.

Recent Thoughts About Educational Measurement

There is, however, a growing body of thought among some educational researchers which seems to question the utility of measurement principles in classroom settings. Some researchers maintain the stance that measurement as a discipline is not a critical aspect of the daily assessment carried out by classroom teachers. Presumably, it is argued that educational measurement may in fact be irrelevant (McLean, 1985). Arguments have been put forward to the effect that classroom teachers have been shown to be effective in their assessment activities without necessarily embracing the ideals of measurement (Gullickson, 1986; Wilson, 1989). For example, the majority of the teachers surveyed by Gullickson (1984) did not rate college courses very highly in preparing them to use classroom tests. However, these teachers believed they had a sufficient understanding of the principles of test construction and utilization.

In recent times, there has been a remarkable rise in the interest expressed between what is taught in educational measurement and classroom practices used to evaluate student achievement (Stiggins, 1988). The failure of teachers to use measurement principles in classroom settings has led to the questioning of the utility of these principles for teaching and learning. According to Anderson (1989), there appears to be a disparity between the science of measurement and assessment as practised in the classroom.

Similar concerns have been expressed by other researchers. In their extensive review of literature on classroom assessment, Stiggins, Conklin, and Bridgeford (1986) have noted that the major interest in educational measurement is based on data generated from standardized paper and pencil tests used for the purposes of accountability. In contrast, Gullickson (1986) found that teachers favoured the use of non-test methodologies such as observation, rating scales, and anecdotal records in the assessment of students.

In addition, Anderson (1989) has noted that the main focus of educational measurement is on issues such as test reliability and validity, standard error of measurement, principles of test design, and item scoring and analysis. However, he contends that these principles are aimed primarily at large-scale test programmes in which the test is administered, scored, analysed, and interpreted as a single isolated entity as in the case of a final examination. He continues to argue that the assessment of student achievement in the classroom is characterized by several testings or data collections throughout an instructional period using a vast array of information-generating and collection procedures. According to this researcher, teacher interest is directed towards a variety of instruments and procedures developed for instructional purposes and not toward the issues and procedures of mainstream educational measurement research and literature.

In addition, Gullickson (1984) and Stiggins and Bridgeford (1985) have noted that the world of the classroom teacher is so different from that of the professional evaluator and that the gap, instead of being narrowed, is as wide as ever. In what appears to be an extension of this view, Wilson (1989) notes that the singleness of purpose which is so helpful in aiding the technical sophistication of the professional is usually absent in the classroom. He continues to point out that the teacher evaluator in a school setting serves several audiences (students, principals, ministry officials, parents) with frequently differing uses of the information generated during student

assessment. Consequently, "the development of objective instrument is replaced with a variety of purposes that makes the teacher's decision making process far more complex" (Wilson, 1990; p.5). This view is shared by Farr and Griffin (1973).

Shephard (1989) acknowledges that classroom assessments are designed to support instruction, are adapted to local context, and are sensitive to short term change in students' knowledge. This researcher further contents that classroom tests and observations need not meet the same standards of accuracy as standardized tests. The rationale inherent in this argument is that errors made in judging individual students are less serious and are more easily redressed as teachers gather new evidence regarding the achievement of their students. Despite the fact that single teacher tests are more likely to be less reliable (in a statistical sense) than a standardized test, pooled data gathered about individual pupils in the course of a school year has much more accuracy (Shephard, 1989).

Earl and Cousins (1993), in their recent review of research on classroom assessment noted that classroom assessment is influenced a lot by the environment inhabited by teachers (school). The school administration may impose regulations and expectations which may influence many of the assessment activities carried out by teachers. School administrators specify testing guidelines some of which may not require teachers to use assessment principles as taught in measurement courses. Wilson (1989) notes that "the school shapes the structure of its reporting system through its policy statements concerning median ranges, failure rates, and symmetry of score distributions" (p.142).

Other researchers (McLean, 1985; Nimmer, 1984) have voiced concern over what they see as the inappropriateness of certain measurement procedures in classroom settings. These researchers have taken issue at most of the test building principles used by measurement

specialists (e.g., domain specification, test blue printing, item format selection, piloting, item analysis, reliability). According to these researchers, measurement principles are assumed to work in the classroom setting, but they do not. McLean (1985), who did a study of student evaluation in Canada, made a remark which seems to question the usefulness of measurement principles in classroom settings:

"The elemental stuff of the measurement courses and text books is as foreign to classroom teachers as spectral analysis of glaze mixture is to potters - and apparently as irrelevant" (p. 39).

Kubiszyn and Borich (1993) have expressed concerns which are consistent with McLean's contention. These authors have advanced the view that internal consistency methods for determining test reliability assume that the test in question measures a single basic concept (related items). However, they point out that most classroom tests involve questions designed to measure several different concepts.

Wilson (1989) in his recent study takes a rather distinct position regarding classroom assessment. He accepts outrightly that classic measurement principles are foreign to the classroom, but that teachers can be very sophisticated in assessing their students even in the absence of measurement training. This view is supported by Gullickson (1986) who has argued that teachers have been successful without a substantial knowledge of measurement and without the use of techniques considered beneficial by measurement experts. These views are at variance with Stiggins' (1991) recent claims about teacher assessment illiteracy in the classroom.

The differences in view regarding the utility of measurement principles provide a clear indication regarding the lack of consensus among experts about the value of these principles in classroom settings. However, despite the differences in view regarding the relevance of measurement principles, there is a clear indication that teachers appreciate the value of some of

these principles in the assessment of student achievement.

Summary of the Literature Review

In summary, some substantial amount of effort has been devoted to the study of the use of measurement principles by teachers. It is now clear that some teacher training institutions in North America do not provide training in the use of measurement principles. Further, where training has been provided, many teachers do not perceive themselves to have been adequately prepared for their assessment practices. This view is supported by findings which show that teachers have expressed the need for programmes designed to equip them with the requisite skills for test development.

From the studies reviewed, one can conclude that teachers attach much importance to measurement principles which relate directly to instruction. These include the principles for test construction, test administration, scoring, determination of validity and reliability, using tests to diagnose students' needs, relating tests to instructional objectives, and reporting. Among the principles considered as being least important are statistics used to describe scores, interpretation of different kinds of test scores, interpretation of test norms, and use of standardized achievement tests.

In addition, one can also conclude that teacher-made tests have serious quality problems, indicating very little or no use of measurement principles by teachers. Among the factors considered as contributing to the little use of measurement principles by teachers are lack of training in measurement, the irrelevance of certain measurement principles, lack of time, and teacher comfort with mathematical concepts (e.g., in statistical analysis). Other factors include ministry policies that may have a direct impact on assessment, policies of student assessment unique to individual schools, and availability of resources such as computers.

Given that most of the studies on the use of measurement principles have been limited to North American settings, very little is known about the use of measurement principles in other settings. In addition, very little is known about the extent to which different factors influence the use of measurement principles in schools. This study was designed to provide answers to these concerns. With the above objectives in mind, the writer preferred research methods which closely paralleled those used by other researchers in related studies (Borg, Worthen, & Valcarce, 1986; Newman & Stallings, 1982; Onocha & Okpala, 1991). The research methods for this study are discussed in the next chapter.

CHAPTER 3

RESEARCH METHODOLOGY

The research methodology used in this study is described in this chapter. The setting of the study (a brief description of Kenya and the Riftvalley Province), the target population and study sample, the instruments used, and the data collection procedures and methods used in data analyses are described.

The Setting of the Study

The setting of the study is described in this section. This description includes a brief description of Kenya and the Riftvalley Province.

Data collection for this study was conducted in the Riftvalley Province (Kenya) in the months of July-September, 1994. Kenya, one of the three countries that make up East Africa, obtained its independence from the United Kingdom in 1963. The country is divided into eight provinces- Central, Coast, Eastern, North Eastern, Nairobi, Nyanza, Riftvalley, and Western. Except for Nairobi, each of the seven provinces is divided into a number of districts. There are 46 districts in the country.

Riftvalley Province, with an estimated population of six million people, is the largest province in the country. It is also the province with the largest number of secondary schools in the country.

Secondary school enrolments in the province continue to increase each year due to the rapidly rising population. This high student enrolment, together with the recent introduction of the new system of education (8-4-4), have had serious implications which pose major challenges to school administrators and teachers in the assessment of student achievement not only in the

province but also in the whole country as well. Inadequate financial resources and the preparation to cope with these challenges, made worse by the recent global economic recession, are some of the major constraints in the school system.

The Target Population and Study Sample

The target population and study sample are briefly described in this section. This description covers the training, employment of teachers, and the sampling procedures used in this study.

Professionally trained history and mathematics teachers serving in public secondary schools in the Riftvalley Province formed the target population for this study.

Diploma and graduate teachers in two curricular areas (mathematics and history) and serving with the Teachers' Service Commission (TSC) were targeted for this study. These teachers had taken educational measurement as part of the requirement for their certification. The training period for diploma teachers usually takes two years while that of graduate teachers takes three years. However, with the recent changes in the education system, the training periods have been changed to three and four years respectively.

All trained secondary school teachers serving in public schools in Kenya are employed by the Teachers' Service Commission (TSC). The TSC is a corporate body established by the Teachers' Service Commission Act (1968), which is responsible for the recruitment, registration, deployment, discipline, transfer (inter-school, inter-district, and inter-provincial), and other matters related to teachers' service in Kenya. The TSC is an arm of the Ministry of Education.

Every year, the TSC receives lists of newly certified teachers from teacher-training institutions and assigns them to schools on the basis of individual school needs. In addition, all

teachers sign declaration forms prior to deployment binding themselves to serve in any part of the country as may be deemed necessary by the TSC. The TSC also strives to ensure that a balance of diploma and graduate teachers is maintained in all the provinces in the country.

The present study began with the acquisition of a research study permit from the relevant authorities. This permit (Appendix A) was subsequently used to facilitate the writer's access to teacher records maintained at the Riftvalley Provincial Education Office (Nakuru). These records contain information on the number of teachers in the province, teacher names and qualification, teacher curriculum area of instruction, and school addresses. The number of teachers in the province keeps changing due to teacher transfers from one province to another, deployment of newly certified teachers, and teachers leaving the profession. The records used in this study were accurate as of April, 1994.

From a population of 750 graduate-mathematics teachers, 600 diploma-mathematics teachers, 500 graduate-history teachers, and 400 diploma-history teachers, a stratified random sample of 100 teachers in each of the designed categories was drawn for the study. The sampling procedure resulted in a sample that consisted of 200 mathematics teachers and 200 history teachers, thus giving a total sample of 400 teachers.

It was the writer's hope that the stratification of teachers by qualification and curriculum area of instruction ensured representation of teachers from the two categories of interest. It was also hoped that the randomization provided for a representative sample that included respondents from a wide variety of schools - small, medium, large, long-established, and newly-established. The sampling procedure hopefully provided for the participation of respondents with the information sought in section K of the questionnaire (See Appendix B).

The Questionnaire

The questionnaire used in this study is described in this section. This description provides information on the development of the instrument, contents of the instrument, and the procedures followed in the pilot testing of the instrument.

The mail questionnaire approach was chosen to survey a sample of secondary school teachers in the Rift Valley Province (Kenya), because it was convenient and within the financial means of the researcher (Borg & Gall, 1989). In addition, Mouly (1978) suggested that the questionnaire approach enabled the researcher to preserve respondent anonymity (thus it may elicit more candid responses) and also allowed "greater uniformity in the way questions were asked and thus ensured greater compatibility in the responses" (p.189).

The research instrument was developed based on the Teacher Education Syllabus in Measurement and Evaluation in Kenya (Ministry of Education, 1993) and was greatly supplemented by the research literature. This syllabus, together with The Standards For Teacher Competence in Educational Assessment of Students (1990) in the United States, and The Principles for Fair Student Assessment in Canada (1990), were used to identify important assessment competencies for classroom teachers and the specific assessment activities associated with each of these competencies. A number of items were also adopted from survey instruments used by other researchers (Borg, Worthen, & Valcarce, 1986; Gullickson, 1982; Gullickson & Ellwein, 1985; Mayo, 1964; Newman & Stallings, 1982; Plake, Impara, & Fager, 1993). However, these items were subsequently modified to suit the writer's study and the Kenyan context.

The questionnaire (Appendix B) was developed on the premise that assessment procedures are cyclical in nature. That is, a test is developed to meet specific purposes. It is then administered, marked, and results analysed and used in the context of the intended purposes by teachers. However, it was found necessary to begin the questionnaire with items on the uses of student assessment results because it was felt that this would stimulate the respondents' interest in completing the instrument.

The questions on the survey questionnaire were preceded by instructions on how they were to be answered. Due to the complex nature of the instrument, respondents were provided with an example on how they were to complete it.

The questionnaire was composed of ten sections, namely, (A) Uses of student assessment results, (B) Principles of test construction, (C) Types of test questions and other assessment methods, (D) Principles of test administration, (E) Principles of test marking, (F) Summarizing and interpreting test scores, (G) Evaluating test quality, (H) Reporting of student assessment results, (K) Factors affecting the use of principles in tests and measurements, and (L) Personal information.

Each of the sections except L included a number of items. The number of these items varied for each of the sections A to K (see appendix B). In sections A to H, respondents were asked to indicate their frequency of use of measurement principles. They responded according to the scale: Never, Rarely, Sometimes, and Always.

Teachers were also asked to rate the importance of each of the measurement principles included on the survey instrument. In this case, they responded using a Likert-type scale which ranged from Not important at all to Very important. The respondents were also required to

indicate whether or not they had received instruction in each of the measurement principles provided on the survey instrument by choosing either Yes or No.

In section K of the questionnaire, teachers were asked to rate the extent to which each of six factors affected their use of measurement principles in test construction, test administration, test marking, summarizing and interpreting test results, evaluating test quality, and reporting of student assessment results. They responded according to the scale, Does not affect my use of the skills (1) to Greatly affects my use of the skills (4).

Another item, Others (please specify), was added to section K to enable respondents to indicate other factors not specified by the writer. This was to add richness to the data.

The section on personal comments was meant to allow respondents to make comments on any aspect of this study and also to expound on their responses to the questionnaire items. Section L of the survey instrument required each respondent to provide personal information on the curriculum area of instruction as well as the qualification level (diploma or graduate). In addition, teachers were asked to indicate whether or not they were interested in participating in the interview phase of the study and whether they would like to receive a summary of the study findings. They were also asked to provide their contact addresses and telephone numbers in order to facilitate further communication with the writer.

The survey questionnaire was distributed to 15 teachers in three Kenyan secondary schools for pilot testing in June, 1994. All the teachers had been in the field for at least one year prior to the pilot study. The objective of this study was to ascertain that respondents could follow instructions associated with the format of the questionnaire and to identify items that were not clear. Each teacher was asked to review each item on the questionnaire to check for ambiguity

and appropriateness to the Kenyan context. These individuals made valuable comments to promote maximum clarity in the instrument. Consequently, some changes were made in the questionnaire before the mail survey was carried out.

The Interview Guide

In this section, the interview guide is briefly described. This description covers the contents of the instrument and the pilot testing that was carried out.

The interview guide contained a number of open- ended questions designed to provide information on factors affecting teacher use of measurement principles. These factors included availability of time, availability of resources, Ministry of Education policies on examinations, Ministry of Education policies on continuous assessment of students, individual school policies on student assessment, and teacher's level of comfort with mathematical concepts. The instrument also contained questions designed to determine teacher assessment needs. The instrument can be found in Appendix C.

The interview guide was pilot- tested with a representative sample of teachers in two schools. Six teachers who had taken part in the pilot phase of the questionnaire voluntarily participated in the preliminary interviews. These teachers came from the same population as the main study sample. However, they did not subsequently form part of the eventual sample that provided data for this study.

The pilot study served a number of purposes. It was conducted to evaluate and improve the guide and the interview procedure. By so doing, it helped the interviewer develop some experience in using the procedure before any research data for the main study were collected. The pilot study also served to determine whether the planned procedures would produce the data

desired.

Data Collection and Analysis Procedures

The data collection procedures used in this study are outlined in the paragraphs below. First, the procedures used for the collection of data in the mail survey are described. This is followed by the description of the procedures used for data collection in the interview phase of the study. The methods used for data analysis in each case are also described.

The Mail Survey

The mail survey was designed to provide quantitative information on teacher training in measurement, importance of measurement principles, and factors affecting the use of these principles in schools.

The questionnaires were distributed by mail to secondary school teachers in Riftvalley Province. A self-addressed, stamped return envelope was enclosed with each questionnaire to be used in mailing the completed instrument to the writer. Also included was a covering letter (see Appendix D) which stated the purpose of the study, indicated the study's sponsorship, the respondent's important role in completing the questionnaire, and appealed to the respondent to participate in the interview phase of the study. The covering letter also assured each respondent of complete anonymity in case of the mail survey responses and confidentiality of the information provided in the interview. The writer also stressed the richness of data through teacher participation in the study and promised to provide a summary of the research findings to each respondent who wished to receive it. The letter was duly signed by the writer's advisor and also by the writer himself.

All the questionnaires were mailed out in early July, 1994. Of the 400 questionnaires

mailed out, 248 were returned. Two of these were rejected because they had been completed by untrained teachers. This brought the usable questionnaires to 246, thus giving a response rate of 61.5%. By qualification, 54.1% (133) of the graduate teachers responded to the questionnaire compared to 45.9% (113) of their diploma counterparts. By curriculum area of instruction, 50.8% (125) of the mathematics teachers responded to the survey compared to 49.2% (121) of the history teachers.

The responses for each item on the questionnaire were coded prior to entering the data into the computer. For the purpose of data analysis, it was found necessary to recode the responses on the effects of various factors on the use of measurement principles. The new response scale ranged from, Does not affect my use of the principles (0) to Greatly affects my use of the principles (3). Analysis of the questionnaire data was performed using the Statistical Package for the Social Sciences (SPSS Release 4.1 for IBM VM/CMS) computer programme. Descriptive statistics -means, standard deviations, frequencies, percentages, and t-tests were used.

Semi-structured Interview

The semi-structured interview was designed as a follow up study to the mail survey. Its purpose was to explore in detail the factors affecting the use of measurement principles in schools. The study also served to identify teacher assessment needs.

Teachers from one district (Kericho) participated in the interview phase of this study. Following the receipt of questionnaires from the mail survey, those teachers who had indicated their willingness to participate in the interview were identified (32). The sample was stratified by teacher qualification and curriculum area of instruction. This resulted in a sample which comprised of 16 graduate and 16 diploma teachers respectively. Sixteen teachers in each of the

curricular areas (history & mathematics), were drawn for interviews. Arrangements for interviews were made through phone and personal visits to teachers in their schools.

For the purpose of teacher interviews, the writer used the research permit (Appendix A) to obtain an introductory letter from the District Education Officer at Kericho. This letter was addressed to all secondary school heads (principals) and appealed to them to offer any necessary assistance to the writer during his visits to their schools. The writer visited individual teachers in their schools during the month of September, 1994, and carried out a thirty-minute interview with each of the thirty-two teachers identified for this phase of the study. Interview responses were recorded in the form of field notes. These field notes and teacher comments were analysed for themes which would help inform the mail survey findings. The results of the data analyses for both the mail survey and teacher interviews are presented in chapter 4.

CHAPTER 4

RESULTS

In this chapter the results of the data analyses are presented. This data analysis follows the sequence of the research questions addressed by the study. The analysis of quantitative data is presented and supplemented by qualitative data.

Teacher Training in Measurement

Question 1 of the research study focused on teacher-training in the use of measurement principles. Teachers were asked to indicate whether or not they had received instruction in the use of various measurement principles. The results of the data analysis are presented in Table 1.

The results show that most of the teachers have received training in the principles of educational measurement. These principles include the use of student assessment results for various purposes in the classroom, test construction principles, use of different types of test items in the assessment of student achievement, principles governing test administration, test marking, the use of statistics in summarizing and interpreting student test scores, principles for evaluating test quality, and the reporting of student assessment results.

Of the 246 teachers who responded to the survey, 58.9-98.4% of them indicated that they had received instruction in the use of measurement principles and concepts provided on the survey questionnaire. The highest reported percentages in instruction were in the principles for summarizing and interpreting test scores (87.4-97.2%), the use of student assessment results for various purposes in schools (79.3-98.4%), and the principles for test construction (78.9-95.9%). The results, however, indicate that fewer teachers have received instruction in the use of test marking principles. In particular, 66.7% of the teachers surveyed indicated having been taught

to assess the degree of inter-rater consistency in marking. About 72.% of the teachers indicated that they had been instructed to remark student scripts to check for consistency in scoring standards. About 60% (59.3%) of the teachers indicated that they had received instruction in the use of parent conferences for reporting student grades.

Table 1

Frequency and Percentage Distributions of Teachers Trained in the Use of Measurement

Principles. (N = 246)

<u>Uses of Student Assessment Results</u>	Frequency	Percentage
Were you ever taught to use student assessment results to:		
1. assign grades to students?	242	98.4
2. change a student from one curriculum to another?	195	79.3
3. promote students from one class to another?	205	83.3
4. counsel students in their academic work?	229	93.1
5. aid students in their career planning?	206	83.7
6. provide students with feed back on their academic progress?	238	96.7
7. diagnose student needs?	208	84.6
8. plan remedial instruction or work for academically "weak" students?	229	93.1
9. plan advanced instruction or work for academically "strong" students?	208	84.6
10. modify teaching techniques?	236	95.9

...Table 1 Contd

Table 1 Contd

<u>Test Construction</u>	Frequency	Percentage
When constructing a test, were you ever taught to:		
1. define content areas or topics to be tested?	232	94.3
2. develop a test blueprint (table of specifications)?	210	85.4
3. set test questions (e.g., multiple choice, essay, etc.) to match the cognitive level to be tested?	236	95.9
4. use item writing guidelines in the development of test questions?	194	78.9
5. develop items that reflect instructional objectives?	231	93.9
6. assess the degree to which the items reflect the content domains to be tested?	215	87.4
7. review the test questions for possible bias, ambiguity, etc.?	230	93.5
<u>Types of Test Questions and Other Assessment Methods</u>		
Were you ever taught the use of:		
1. Multiple-choice questions?	217	88.2
2. True-false questions?	207	84.1
3. Essay type questions?	212	86.1
4. Short-answer questions (e.g., two or three sentence answer)?	213	86.6
5. Matching exercises (linking or matching items from two lists)?	192	78
6. Problem-solving questions (finding a solution based on a word problem)?	197	80.1
7. Observational techniques (e.g., checklists, rating scales) used to judge student performance or products?	191	77.6

...Table 1 Contd

Table 1 Contd

<u>Principles of Test Administration</u>	Frequency	Percentage
Were you ever taught to:		
1. Inform the students of the scoring criteria (e.g., grammar, handwriting, etc.) to be used in evaluating their responses?	210	85.4
2. Structure the testing situation to discourage cheating (e.g., seating arrangements)?	217	88.2
3. Maintain a constant presence in the classroom during the course of the test?	217	88.2
<u>Principles of Test Marking</u>		
Were you ever taught to:		
1. Develop model answers for the test?	217	88.2
2. Develop a scoring key or marking scheme before scoring tests?	238	96.7
3. Score student responses one question at a time for all the students when marking essays?	184	74.8
4. Take frequent breaks when marking tests?	145	58.9
5. Remark the first few student test scripts to check for consistency in your scoring standards?	179	72.8
6. Assess the degree of inter-rater (marker) consistency if more than one person scores an examination?	164	66.9
<u>Summarizing and Interpreting Test Scores</u>		
Were you ever taught to:		
1. Rank order students according to their test scores?	232	94.3
2. Construct the frequency distribution of the student test scores?	234	95.1
3. Calculate the mean of the test scores?	239	97.2
4. Calculate the median of the test scores?	230	93.5
5. Calculate the range of the test scores?	235	95.5
6. Calculate the variance or standard deviation of the test scores?	232	94.3
7. Calculate the percentiles associated with each test score?	223	90.7
8. Compare test scores with established norms?	215	87.4

...Table 1 Contd

<u>Evaluating Test Quality</u>	Frequency	Percentage
Were you ever taught to:		
1. Calculate the difficulty of each question?	217	88.2
2. Calculate how well an item discriminates between students of high and low abilities?	213	86.6
3. Use the item difficulty and discrimination statistics to improve the test for future use?	202	82.1
4. Assess the test-retest (stability) reliability of the test?	210	85.4
5. Assess the internal consistency of the test (e.g., split-half reliability)?	191	77.6
<u>Reporting of Student Assessment Results</u>		
Were you ever taught how to:		
1. Set objective criteria for assigning grades to students?	225	91.5
2. Provide written comments regarding student strengths and weaknesses?	233	94.7
3. Provide information on student performance in such a way that is easily understood by parents?	219	89.0
4. Use parent conferences to report student grades?	146	59.3

Importance and Frequency of Teacher Use of Measurement Principles

Most educational measurement experts stress the use of several principles in the assessment of student achievement. Such principles outline possibilities. The objective of this study was to identify actual teacher practices with respect to the use of measurement principles. Teachers were asked to rate the importance and their frequency of use of several measurement principles. The results on these two aspects of the study are presented together in order to establish whether there is a close relationship between the importance and frequency of use of measurement principles. These results are presented by sections (i.e., uses of student assessment

results, test construction, types of test questions used, test administration, test marking, summarizing and interpreting test scores, evaluating test quality, and reporting of student assessment results).

Importance and Frequency of Teacher Use of Student Assessment Results

Table 2A shows the results of the data analysis regarding teachers' perceptions of the importance and their frequency of use of student assessment results.

It can be seen that the use of assessment results for the purpose of providing feedback to students on their academic progress ($M = 2.95$), assigning grades to students ($M = 2.94$), counselling students in their academic work ($M = 2.93$), and modifying teaching techniques ($M = 2.85$) received high importance ratings from teachers. The use of assessment results for the purpose of planning advanced instruction or work for academically "strong" students ($M = 2.16$) was the only item which received an importance mean rating below 2.50.

The results also show that teachers frequently use assessment results for those purposes with high importance ratings. Thus, teachers frequently use assessment results for the purpose of providing students with feedback on their academic progress ($M = 2.81$), assigning grades to students ($M = 2.66$), and counselling students in their academic work ($M = 2.40$). However, it is also clear that there is some discrepancy between the importance rating of certain purposes served by student assessment results and the frequency with teachers use the results for these purposes in their schools. It is interesting to note that though teacher use of assessment results for the purpose of modifying teaching techniques ($M = 2.85$), planning remedial instruction or work for academically weak students ($M = 2.84$), promoting students from one class to another ($M = 2.63$), helping students in their career planning ($M = 2.59$), and changing students from one

curriculum to another ($M = 2.51$) received high importance ratings from teachers, teachers do not frequently use assessment results for these purposes in their schools (mean frequency of use = 1.35, 1.85, 1.59, 1.87, and 1.48 respectively).

Table 2A

Importance and Frequency of Teacher Use of Student Assessment Results (N = 246)

Uses of Student Assessment Results	Mean Importance	SD	Mean Frequency of Use	SD
1. assigning grades to students	2.94	0.26	2.66	0.55
2. changing a student from one curriculum to another	2.51	0.82	1.48	0.91
3. promoting students from one class to another	2.63	0.78	1.59	1.26
4. counselling students in their academic work	2.93	0.28	2.40	0.66
5. aiding students in their career planning	2.59	0.80	1.87	0.94
6. providing students with feedback on their academic progress	2.95	0.23	2.81	0.47
7. diagnosing student needs	2.62	0.64	2.03	0.76
8. planning remedial instruction or work for academically "weak" students	2.84	0.44	1.85	0.91
9. planning advanced instruction or work for academically "strong" students	2.16	1.07	1.35	0.98
10. modifying teaching techniques	2.85	0.44	2.35	0.65

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always.

Importance and Frequency of Teacher Use of Test Construction Principles

Table 2B shows the results of the data analysis on teachers' perceptions regarding the importance and their frequency of use of test construction principles.

Table 2B

Importance and Frequency of Teacher Use of Test Construction Principles (N = 246)

<u>Principles of Test Construction</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Define content areas or topics to be tested	2.83	0.43	2.66	0.58
2. Develop a test blueprint (table of specifications)	2.13	1.02	1.43	1.04
3. Set test questions (e.g., multiple choice, essay, etc.) to match the cognitive level to be tested	2.80	0.50	2.36	0.78
4. Use item writing guidelines in the development of test questions	2.13	0.98	1.61	0.95
5. Develop items that reflect instructional objectives	2.79	0.52	2.46	0.69
6. Assess the degree to which the items reflect the content domains to be tested	2.69	0.58	2.22	0.76
7. Review the test questions for possible bias, ambiguity, etc.	2.83	0.44	2.23	0.89

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

Regarding the principles for test construction, the definition of the content areas to be tested ($M = 2.83$) and the review of test questions for possible bias and ambiguity ($M = 2.83$)

received the highest importance ratings from teachers. The setting of test questions to match the cognitive levels to be tested ($M = 2.80$), developing test items that reflect instructional objectives ($M = 2.79$), and developing items that reflect the content domains to be tested ($M = 2.69$) also received high importance ratings from teachers. The frequency of teacher use of each of these five principles for test construction is also high (mean frequency of use = 2.66, 2.23, 2.36, 2.46, and 2.22 respectively).

The development of test blueprints ($M = 2.13$) and the use of item writing guidelines in the development of test questions ($M = 2.13$) were rated as being somewhat less important compared to the other principles of test construction. The low importance ratings of these principles are reflected in the low frequency with which they are used by teachers during test construction (mean frequency of use = 1.43 and 1.61 respectively).

Importance and Frequency of Teacher Use of Different Types of Test Questions and Other Assessment Methods

Table 2C shows the results of the data analysis regarding teachers' perceptions of the importance and their frequency of use of different types of test questions.

It can be seen that the use of short-answer ($M = 2.13$) and essay items ($M = 2.07$) received moderate importance ratings from teachers. It can also be seen that there is some slight discrepancy between the importance ratings of these types of items and the frequency with which they are used by teachers for assessing students (mean frequency of use = 1.78 and 1.64 respectively).

However, the use of true-false questions ($M = 1.24$), matching exercises ($M = 1.48$), and

multiple-choice questions were rated as being least important by teachers ($M = 1.24$, 1.48 , and 1.83 respectively). One can also note that there is some slight discrepancy between the importance ratings of these types of items and the frequency with which they are used by teachers (mean frequency of use = 0.51 , 0.85 , and 1.19 respectively).

It is also interesting to note that observational techniques used for judging student performance or products ($M = 2.12$) were rated as being moderately important although they are not frequently used for assessing students (mean frequency of use = 1.50).

Table 2C

Importance and Frequency of Teacher Use of Different Types of Test Questions and Other Assessment Methods (N = 246)

<u>Types of Test Questions and Other Assessment Methods</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Multiple-choice questions	1.83	1.20	1.19	1.09
2. True-false questions	1.24	1.11	0.51	0.59
3. Essay-type questions	2.07	1.29	1.64	1.24
4. Short-answer questions (e.g., two or three sentence answer)	2.13	1.14	1.78	1.12
5. Matching exercises (linking or matching items from two lists)	1.48	1.08	0.85	0.81
6. Problem-solving questions (finding a solution based on a word problem)	1.94	1.28	1.64	1.27
7. Observational techniques (e.g., checklists, rating scales, etc.) used to judge student performance or products	2.12	1.07	1.50	1.07

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

Importance and Frequency of Teacher Use of Test Administration Principles

Table 2D shows the results of the data analysis on teachers' perceptions regarding the importance and their frequency of use of test administration principles.

With respect to test administration, the highest importance ratings were assigned to the structuring of the testing situation in order to discourage cheating by students ($M = 2.89$), teacher constant presence in the classroom during the course of the test ($M = 2.86$), and informing students of the scoring criteria to be used in evaluating their responses ($M = 2.56$), respectively.

It can also be seen that teachers frequently use these principles when administering tests in their classrooms ($M = 2.32, 2.49$, and 2.00 respectively). In addition, there is very little discrepancy between the importance ratings of these principles and the frequency with which they are used by teachers in the assessment of student achievement.

Table 2D

Importance and Frequency of Teacher Use of Test Administration Principles (N = 246)

<u>Principles of Test Administration</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Inform the students of the scoring criteria (e.g., grammar, handwriting, etc.) to be used in evaluating their responses	2.56	0.74	2.00	0.90
2. Structure the testing situation to discourage cheating (e.g., seating arrangements)	2.89	0.42	2.32	0.86
3. Maintain a constant presence in the classroom during the course of the test	2.86	0.42	2.49	0.70

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

Importance and Frequency of Teacher Use of Test Marking Principles

Table 2E shows the results of the data analysis on teachers' perceptions regarding the importance and their frequency of use of test marking principles.

Table 2E

Importance and Frequency of Teacher Use of Test Marking Principles (N = 246)

<u>Principles of Test Marking</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Develop model answers for the test	2.67	0.66	2.40	0.81
2. Develop a scoring key or marking scheme before scoring the tests	2.99	0.11	2.87	0.37
3. Score student responses one question at a time for all the students when marking essays	2.07	1.10	1.57	1.09
4. Take frequent breaks when marking tests	2.02	0.95	1.78	0.83
5. Remark the first few student scripts to check for consistency in your scoring standards	2.36	0.79	1.40	0.96
6. Assess the degree of inter-rater (marker) consistency if more than one person scores an examination	2.20	0.97	1.35	0.98

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

The results show that virtually all teachers rated the development of scoring keys or marking schemes ($M = 2.99$) as being very important. The development of model answers for tests ($M = 2.69$) were also rated as being highly important. The high importance ratings of these principles are reflected in the high frequency with which they are used by teachers (mean

frequency of use = 2.87 and 2.40, respectively). However, it is also clear that some discrepancy exists between the importance rating of the remarking of student test scripts for scoring consistency ($M = 2.36$) and the frequency with which teachers actually remark student test scripts ($M = 1.40$). The same can also be said of the importance rating of the assessment of inter-rater consistency ($M = 2.20$) and the frequency with which teachers determine the inter-rater consistency of their scoring when more than one person scores an examination ($M = 1.35$).

Importance and Frequency of Teacher Use of the Principles for Summarizing and Interpreting Test Scores

Table 2F shows the results of the data analysis on teachers' perceptions of the importance and their frequency of use of the principles for summarizing and interpreting test scores.

It can be seen that the rank ordering of students according to their test scores ($M = 2.69$) and the determination of the mean test score received high importance ratings from teachers ($M = 2.62$).

With respect to the use of these descriptive statistics, one can also see that teachers frequently rank order students according to their test scores ($M = 2.48$). However, there is some discrepancy between the importance rating of the determination of the mean test score and the frequency with which teachers compute the mean of student test scores ($M = 1.84$). Discrepancies between the importance rating and the frequency of teacher use of statistical principles are also reflected in the case of the determination of the range of test scores (mean importance rating = 2.35; mean frequency of use = 1.61), and comparison of test scores with established norms (mean importance rating = 2.35; mean frequency of use = 1.57).

The determination of percentiles ($M = 1.67$), variance or standard deviation of test scores

($M = 1.74$), and the median test score ($M = 1.75$) were rated as being slightly important. The low importance ratings of these statistical principles are reflected in their low frequency of use by teachers (mean frequency of use = 0.74, 0.77, and 1.06 respectively).

Table 2F

Importance and Frequency of Teacher Use of the Principles for Summarizing and Interpreting Test Scores (N = 246)

<u>Summarizing and Interpreting Test Scores</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Rank order students according to their test scores	2.69	0.57	2.48	0.68
2. Construct the frequency distribution of the student test scores	2.21	0.96	1.35	0.88
3. Calculate the mean of the test scores	2.62	0.73	1.84	0.91
4. Calculate the median of the test scores	1.75	1.18	1.06	0.92
5. Calculate the range of the test scores	2.35	0.87	1.61	0.96
6. Calculate the variance or standard deviation of the test scores	1.74	1.20	0.77	0.83
7. Calculate the percentiles associated with each test score	1.67	1.12	0.74	0.84
8. Compare the test scores with established norms	2.35	0.89	1.57	1.06

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

Importance and Frequency of Teacher Use of the Principles for Evaluating Test Quality

Table 2G shows the results of the data analysis on teachers' perceptions of the importance and their frequency of use of the principles for evaluating test quality.

Of the five principles for evaluating test quality, teachers assigned moderate importance to the determination of item difficulty ($M = 2.18$) and item discrimination indices ($M = 2.08$). The use of item difficulty and discrimination statistics for the purpose of improving tests for future use ($M = 1.94$), the determination of the internal consistency ($M = 1.88$), and test-retest reliability ($M = 1.89$) were rated as being slightly important in the assessment of student achievement. In general, the results also show very little discrepancy between teachers' perceptions of the importance of these principles and the frequency with which they use them in the assessment of student achievement (mean frequency of use = 1.51, 1.39, 1.17, 1.21, and 1.02, respectively).

Table 2G

Importance and Frequency of Teacher Use of the Principles for Evaluating Test Quality (N = 246)

<u>Evaluating Test Quality</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Calculate the difficulty of each question	2.18	1.18	1.51	1.13
2. Calculate how well an item discriminates between students of high and low ability	2.08	1.15	1.39	1.05
3. Use the item difficulty and item discrimination statistics to improve the test for future use	1.94	1.20	1.17	1.04
4. Assess the test-retest (stability) reliability of the test	1.89	1.23	1.21	1.03
5. Assess the internal consistency of the test (e.g., split-half reliability)	1.88	1.05	1.02	0.93

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

Importance and Frequency of Teacher Use of the Principles for Reporting Student

Assessment Results

Table 2H shows the results of the data analysis regarding teachers' perceptions of the importance and their frequency of use of the principles for reporting student assessment results.

Of these principles, teachers assigned the highest importance rating to the provision of written comments regarding student strengths and weaknesses ($M = 2.86$). Very high importance rating was also assigned to the communication of information on student performance in a way that can easily be understood by parents ($M = 2.81$).

Table 2H

Importance and Frequency of Teacher Use of the Principles for Reporting Student

Assessment Results

<u>Reporting of Student Assessment Results</u>	Mean Importance	SD	Mean Frequency of Use	SD
1. Set objective criteria for assigning grades to students	2.76	0.52	2.22	0.83
2. Provide written comments regarding student strengths and weaknesses	2.86	0.37	2.38	0.80
3. Provide information on student performance in such a way that is easily understood by parents	2.81	0.45	2.33	0.84
4. Use parent conferences to report student grades	2.37	0.83	1.09	0.79

Scale: Importance, 0 = Not important at all to 3 = Very important; Frequency of Use, 0 = Never to 3 = Always

The high importance ratings of these principles are clearly reflected in the high frequency

with which they are used by teachers (Mean frequency of use = 2.38 and 2.33 respectively). However, there is some discrepancy between the importance rating of the use of parent conferences for reporting student assessment results (Mean importance rating = 2.37) and the frequency with teachers use these conferences for this purpose in their schools (Mean frequency of use = 1.09).

Teacher Use of Measurement Principles by Teacher Curriculum Area

For the purpose of further analysis, the 246 teachers who gave usable responses to the original mailing were categorized by their curriculum area of instruction. When categorized this way, respondents were found to include 125 mathematics and 121 history teachers, respectively, thus revealing a sample not biased towards subjects of any one curriculum area of interest. There was an expectation at the beginning of this study that mathematics teachers would tend to use statistical analyses on a more frequent basis than their history counterparts. The underlying assumption was that these teachers had not only been introduced to statistical concepts in educational measurement, but also in their mathematics courses. There was also a further assumption that the same teachers use statistical concepts on a frequent basis in their mathematics classes.

Several t-tests were carried out to determine whether there were any significant relationships between teacher curriculum area of instruction and the frequency of teacher use of measurement principles. A probability level of .01 was used in all statistical cases of significance. The results of the analysis are presented in Appendix E. Only those cases which were significant are shown in Table 3.

Table 3

**Teacher Frequency of Use of Different Types of Test Questions and Other
Assessment Methods by Teacher Curriculum Area (Mathematics, N = 125; History, N =
121)**

<u>Types of Test Questions and Other Assessment Methods</u>	<u>Teacher curriculum area</u>	<u>Mean</u>	<u>SD</u>	<u>t</u>
1. Multiple-choice questions	Mathematics	1.31	0.55	-22.38*
	History	2.10	0.70	
2. True-false questions	Mathematics	0.30	0.48	-6.01*
	History	0.73	0.62	
3. Essay-type questions	Mathematics	0.75	1.10	-16.75*
	History	2.56	0.52	
4. Short-answer questions (e.g., two or three sentence answer)	Mathematics	1.27	0.23	-8.11*
	History	2.30	0.66	
5. Matching exercises (linking or matching items from two lists)	Mathematics	0.80	0.89	-0.90
	History	0.89	0.72	
6. Problem-solving questions (finding a solution based on a word problem)	Mathematics	2.68	0.59	23.51*
	History	0.57	0.80	
7. Observational techniques (e.g., checklists, rating scales, etc.) used to judge student performance or products	Mathematics	1.83	1.01	5.28*
	History	1.15	1.01	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always;

* $p < .01$.

Table 3 shows that there were significant differences between mathematics and history teachers in their frequency of use of different types of test items and other assessment methods. History teachers reported a more frequent use of multiple-choice, true-false, essay, and short-answer questions than did the mathematics teachers who responded to the survey. Mathematics teachers on the other hand reported a more frequent use of problem-solving questions and

observational techniques than did the history teachers who responded to the survey.

Significant differences were also found in teacher scoring practices. History teachers reported a more frequent scoring of essays, one question at a time, than did the mathematics teachers who responded to the survey. However, mathematics teachers reported a more frequent remarking of student test scripts to check for scoring consistency than did the history teachers who responded to the survey.

Teacher Use of Measurement Principles by Teacher Qualification

For the purpose of further analysis, the 246 teachers who responded to the survey were categorized by their qualification level. This categorization revealed a sample which consisted of 113 diploma and 133 graduate teachers respectively.

T-tests were carried out to determine whether there were any significant differences between diploma and graduate teachers in their frequency of use of measurement principles. The results of the analysis are presented in Appendix F (no significant cases). A significance level of .01 was used in all cases. The results show that there were no significant differences between diploma and graduate teachers in their frequency of use of measurement principles.

Effects of Various Factors on the Use of Measurement Principles in Schools

Teachers were asked to rate the extent to which each of six factors (i.e., availability of time, availability of resources, Ministry of Education policies on examinations, Ministry of Education policies on continuous assessment in schools, School policies on student assessment, and the teacher's level of comfort with mathematical concepts) affected their use of measurement principles in a number of assessment activities. The teachers responded using a four-point scale which ranged from, Does not affect my use of the principles (1) to Greatly affects my use of the

principles (4). For the purpose of data analysis, the scale was adjusted so that it ranged from , Does not affect my use of the principles (0) to Greatly affects my use of the principles (3).

Teachers also listed other factors not specified by the writer. Later, a number of interviews were conducted with a sample of teachers (32) who had responded to the mail questionnaire. The purpose of the interviews was to allow some of the respondents to expand upon and clarify their responses in an effort to enrich the questionnaire data.

For the purpose of identity and anonymity, interviewees were assigned numbers corresponding to their qualification and curriculum area of instruction. These numbers appear in parentheses at the end of direct quotes (or verbatim statements) from the interviewees. For example, DM1 refers to diploma mathematics teacher number one, whereas DH1 refers to diploma history teacher number one. Similarly, GM1 means graduate mathematics teacher number one, while GH2 means graduate history teacher number two.

Following the completion of the interviews, teacher interview responses and comments to the mail survey were content-analysed for recurring themes. Where appropriate, these themes have been used to supplement the quantitative responses. The results of the quantitative data analysis are presented in Table 4. Means and standard deviations are reported for each item with the standard deviations appearing in brackets.

The results (see Table 4) show fairly large variations in teacher responses regarding the effects of various factors on the use of measurement principles in schools.

Availability of Time:

The availability of time was rated as having a moderate effect on teacher use of measurement principles in the evaluation of test quality ($M = 2.09$) and in the summarizing and

interpretation of test scores ($M = 2.06$). However, the availability of time was rated as slightly affecting teacher use of measurement principles in test marking ($M = 1.95$), test construction ($M = 1.83$), test administration ($M = 1.48$), and in the reporting of student assessment results ($M = 1.41$).

Table 4

Means and Standard Deviations of Teachers' Perceptions of the Relative Effects of Various Factors on the Use of Measurement Principles in Schools. (N = 246)

Factors	TC*	TA	TM	S/I	ET	RR
Availability of time	1.83 (1.17)	1.48 (1.21)	1.95 (1.14)	2.06 (1.04)	2.09 (1.06)	1.41 (1.18)
Availability of resources (eg., photocopying machines, computers).	1.92 (1.19)	1.11 (1.20)	0.77 (1.10)	1.35 (1.23)	1.17 (1.23)	0.90 (1.13)
Ministry of Education policies on examinations	1.37 (0.20)	0.91 (1.15)	1.07 (1.20)	1.07 (1.17)	1.17 (1.22)	0.94 (1.18)
Ministry of Education policies on continuous assessment of students	1.23 (1.20)	0.93 (1.10)	0.92 (1.11)	1.11 (1.14)	1.05 (1.14)	1.00 (1.14)
Your own school's policies on student assessment	1.44 (1.27)	1.55 (1.25)	1.39 (1.25)	1.36 (1.22)	1.32 (1.23)	1.43 (1.26)
Your level of comfort with mathematical concepts	0.57 (1.03)	0.36 (0.80)	0.57 (1.00)	1.25 (1.28)	1.11 (1.23)	0.69 (1.07)

Scale: 0 = Does not affect my use of the principles to 3 = Greatly affects my use of the principles.

*TC = Test Construction; TA = Test Administration; TM = Test

Marking; S/I= Summarizing & Interpreting Test Scores; ET =

Evaluation of Test Quality; RR = Reporting of Results

A number of teachers, however, expressed great concern about the lack of time in which to construct and mark student tests. Three of these teachers had the following to say about the ever pressing issue of time in student assessment:

There is virtually no time for us to pay close attention to content representativeness of our tests, peer review of tests for technical adequacy, or even to carry out objective scoring of student responses. Most of the time is spent on teaching. (DM2)

The demands on the teacher's time are such that he may delay test preparation unduly. Thus he may put together a test at the last possible moment without any allowance for review and criticism. The likelihood of a good test resulting from such an effort is somewhat remote. (GH4)

Teaching, unlike other professions, is an open-ended activity. The list of the things we do and those we could do, just goes on and on. If we had enough time, tests could always be reviewed and improved, more test items could always be included on tests, and marking could always be done with greater care. (GH2)

The teachers interviewed complained about assessment in these terms. They complained that one needs to make notes, schemes of work, lesson plans, set and mark continuous assessment tests for several classes. They indicated that these aspects of their work leave teachers with very little time in which to engage in technical details in the assessment of student achievement.

Some of the teachers indicated that they stayed late in school, until 6:00 in the evening so they would not take student test scripts home with them. Others reported that they often spend their weekends trying to complete test construction and marking of student tests.

Many teachers also complained about their work, which they said has become much more complex because of the increased responsibilities teachers assume outside the classroom. They indicated that they are required to supervise student morning preps, take student roll calls daily,

oversee students during meal times in the dining halls, coach sports teams, and act as student club patrons. The list, as one teacher put it, is even longer.

Availability of Resources:

The availability of resources was rated as having a slight effect on teacher use of measurement principles in test construction ($M = 1.92$), and as having the least effect on the use of these principles in test marking ($M = 0.77$) and in the reporting of student assessment results ($M = 0.90$).

A variety of concerns regarding the availability of resources emerged during the personal interviews and from the analysis of teacher comments to the mail survey. Thirty-two percent (78) of the teachers complained about the lack of enough chairs or desks in their schools. They complained that the lack of these facilities made it extremely difficult for them to curb student cheating during test administration. The comment made by one of these teachers is a case in point:

It is unfortunate that students have to share these facilities (desks and chairs) especially when important tests are being written. We always have to live with the understanding that test results are contaminated by student cheating (GM1).

Some of the teachers also expressed disappointment about the lack of basic facilities for typing tests in their schools. This seemed to be a wide spread concern among teachers in rural schools faced with financial difficulties. "Often, we have to set tests way in advance so that they can be taken for typing in other schools which have typing facilities" (DH2). However, these teachers indicated that there are times when they cannot cover the anticipated test topics because of unforeseen circumstances in their schools. Given these circumstances, the teachers conceded that there is nothing much they can do to alter the content of the tests already typed in order to match the content of the areas covered during instruction. These teachers went on to point out

that their students are often tested on topics not fully covered during instruction. As a result of the mismatch between what is taught and what is tested, the teachers felt that student test scores may not always be valid for the purpose for which they were originally intended.

Ministry of Education Policies on Examinations

Ministry of education policies on examinations were rated as having a slight effect on teacher use of measurement principles in test construction ($M = 1.37$), and as having the least effect on the use of these principles in test administration ($M = 0.91$) and in the reporting of student assessment results ($M = 0.94$).

However, there was some concern among history teachers that the types of questions used for testing students in national examinations made it difficult for teachers to test students on higher order cognitive skills:

The tests we set are largely determined by the format and types of questions set by the Kenya National Examinations Council. Higher order cognitive skills are not usually tested and so we limit our tests to low level cognitive skills such as knowledge and comprehension (GH5).

Mathematics teachers, however, seemed satisfied with the types of questions used for assessing students in national examinations. In particular, these teachers indicated that the types of questions used required students to apply their knowledge. However, they indicated that they always have to imitate the national examination in the types and number of test items used for assessing their students.

Some teachers expressed the view that they do not use some of the measurement principles because the grades they assign to students in their classrooms are not usually taken into consideration by the ministry when grading students on completion of high school. As a result,

"teachers tend to relax their efforts in dealing with classroom tests" (GM7).

In their personal comments to the mail survey, many teachers commented that they do not use student assessment results for the purpose of student promotion because this would contradict ministry policy. One school headmaster had this to say about the policy:

Using student assessment results for the purpose of student promotion would be tantamount to contradicting Ministry of Education policy. The policy discourages repetition of classes by secondary school students even where a student does not merit proceeding to the next class because of poor test results. The ministry has succumbed to parental outcry over the high tuition cost of student repetition of classes. (GH3)

Teachers also frequently cited ministry policy as greatly limiting their use of student assessment results for the purpose of changing a student from one curriculum to another.

The introduction of subject allocation modalities for Form 3 students has limited the choice of subjects and some students have to contend with subjects they are poor in because they are included in the modality taken by the particular class. (GM1)

A number of teachers also commented that students usually opt to take curricular areas that offer better job prospects even if their achievement in those curricular areas is poor.

Ministry of Education Policies on Continuous Assessment of Students:

As the results in Table 4 show, Ministry of Education policies on continuous assessment of students were rated as having a slight effect on teacher use of measurement principles in test construction ($M = 1.23$) and in the summarizing and interpretation of student test scores ($M = 1.11$). The same policies on continuous assessment of students were rated as having the least effect on teacher use of measurement principles in test marking ($M = 0.92$) and test administration ($M = 0.93$).

However, about 24% (60) of the teachers who responded to the mail survey expressed

concern about the ministry policy on continuous assessment of students. These teachers felt that the policy, which calls for the frequent assessment of students, is not appropriate given the wide syllabuses to be covered in most curricular areas. Some teachers also felt that school administrators often took the ministry policy too far. One history teacher's comment is a case in point:

We are supposed to administer four tests per term. These are rather too many for the teacher to construct and score effectively. Hence we take testing in general as something that has to be done for student record purposes. There is no dedication among us in this area (GH7).

School Policies on Student Assessment

School policies were rated as having a slight effect on teacher use of measurement principles in test administration ($M = 1.55$), test construction ($M = 1.44$), and the reporting of student assessment results ($M = 1.43$).

However, during the interviews, teachers complained about the lack of clear school policies on student assessment. In particular, they expressed disappointment with school heads who they said often released examination time tables quite late in the school term, thus leaving them with very little time in which to prepare tests. This disappointment is reflected in a comment made by one history teacher:

It is not easy to match items to the desired cognitive levels to be tested when teachers have to beat deadlines in handing in examinations for typing. Teachers cannot be expected to design high quality tests when a week's notice is given for this purpose. A well defined policy that allows teachers enough time to prepare tests would really help (DH4).

The mail survey teacher comments and teacher interviews revealed that school policies with respect to student assessment show much variation. Some schools require teachers to assess

their students on a monthly basis, whereas others require students to be assessed twice in a school term of three months. In some schools, tests are administered during class time. However, in others, these tests are administered at night, after school in the evening, or in the morning before classes begin. In some schools, teachers have to find their own appropriate time in which to administer tests, and quite often this test administration takes place over the weekends. However, some teachers seemed unhappy with this last arrangement. One teacher complained, "We also need time to attend to personal issues" (GH3).

Some of the teachers also expressed concern about the administration of tests during class time. They pointed out that administering tests during class time greatly restricted the scope to be tested and also retarded the coverage of the syllabus. In a comment that typifies many that were made with respect to the scope to be tested, two teachers had this to say:

We have to administer tests during our normal class time. This is a school policy. But how much can you actually test in forty minutes? It really represents a tiny fraction of what has been taught in a month (DH2).

You will find that the syllabus to be covered is usually too long compared to the amount of time devoted to it. Therefore, teachers spend most of their time teaching in order to cover this syllabus. Administering tests during class time retards syllabus coverage (GM4).

Teacher's level of Comfort with Mathematical Concepts

The results in Table 4 show that teacher level of comfort with mathematical concepts was rated as having a slight effect on the use of measurement principles in the summarizing and interpretation of student test scores ($M = 1.25$) and in the evaluation of test quality ($M = 1.11$). Teacher level of comfort with mathematical concepts was rated as having the least effect on teacher use of measurement principles in test administration ($M = 0.36$), test construction ($M =$

0.57), and test marking ($M = 0.57$).

Some further analysis was done to determine whether there were any significant differences between history and mathematics teachers in the effect of their mathematical aptitude on the use of statistical analyses. The results of the analyses are as shown in Table 5.

Table 5

The Effect of Teacher Level of Comfort with Mathematical Concepts on the Use of Statistical Principles by Teacher Curriculum Area (History, $N = 121$; Mathematics, $N = 125$)

Statistical Principles	Teacher curriculum area	Mean	SD	t
Summary & Interpretation of test scores	History Mathematics	1.74 0.78	1.17 1.19	-6.40*
Evaluation of test quality	History Mathematics	1.50 0.73	1.24 1.11	-5.12*

Scale, 0 = Does not affect my use of the principles to 3 = Greatly affects my use of the principles.

* $p < .01$.

The results revealed significant differences between mathematics and history teachers regarding the effect of their level of comfort with mathematical concepts on the use of statistical principles. A significant number of history teachers indicated that their level of comfort with mathematical concepts affects their use of statistical principles.

Heavy teacher workloads, large class sizes, the irrelevance of some measurement principles, and wide syllabuses were also frequently cited by teachers as factors which influence their use of measurement principles in schools.

Teacher workloads

Over 41% (102) of the teachers who responded to the survey reported that heavy workloads greatly restricted the amount of time available for meaningful application of measurement principles in the classroom. One teacher complained at length about his heavy teaching load and the extra responsibilities he had in his school:

I am the only history teacher in the school, so I teach all the classes in this subject. Besides that, I also teach Kiswahili as there is no other teacher for this subject in the school. I hold the responsibility of the deputy headmaster, and also guidance and counselling master. I have so many responsibilities that I cannot carry out my class work effectively. But then there is nothing I can do because the TSC has not posted enough teachers to the school. Consequently, I usually set short-answer questions which are easy to mark, given the large number of classes which I teach. One unfortunate aspect of this is that many of the questions test low level cognitive skills. Further, the marking schemes used for scoring these tests are also poorly developed, thus leading to inconsistent awarding of points to student responses. (GH5)

One history teacher explained how heavy teacher workloads made it difficult for teachers to develop tests that would yield valid test scores:

Too much workload restricts both the format and the length of the test to be set and administered to students. Tests are shorter in length than we would like them to be. In addition we use multiple-choice items almost to the exclusion of essays. Our multiple-choice items are usually plagued by distractors that are often not plausible. When there is too much work, we do not have enough time to have peers review our tests for ambiguity and bias (DH3).

Some of the history teachers also indicated that they feared setting long, comprehensive essay tests because of the heavy marking that these would entail. One of these teachers put it clearly when in reference to this marking she said:

I teach nine classes, each with about fifty students. So if I administer a three essay test to the classes at the end of the month, I will have to score 1350 essays. This is an immense task given that I have to prepare lesson plans and teach at the same time (GH7).

Although history teachers found it convenient to assess their students using short-answer

test items, they expressed concern that the results of these tests do not concur with those which students usually obtain in national examinations. "Students do very well in our tests because of the narrow scope tested. However in national examinations we notice a big discrepancy in student achievement" (DH3).

Class Size

About 44% (107) of the teachers who responded to the survey expressed concern about the large class sizes they have to deal with during testing. In particular, they complained that large class sizes made it extremely difficult for them to control student cheating during test administration. Many teachers also indicated that statistical analyses of test scores are usually not done when class sizes are large. Two of these teachers expressed concerns shared by many of their colleagues teachers who responded to the mail survey:

Large student numbers (45 to 55 per class) bring in the problem of cheating because classes are overcrowded. Seating arrangements cannot easily be reorganized in such a case. Hence students steal from one another or copy answers from note books. This is reflected by the similar answers to test questions given by students. Consequently, the validity of student test scores is highly questionable (DM7).

If a teacher has so many students to teach and test, then such things as calculating means, standard deviations, item difficulties and test reliability are usually left out. We feel that they do not affect the normal routine (GH6).

Many teachers felt that meaningful testing calls for the investment of a considerable amount of time on the part of the teacher, especially in test construction and marking. They complained that the number of scripts to be marked in a test are usually quite many because of the large class sizes in schools. "So, calculating the mean, median, range, and standard deviation is time consuming and in some cases, the school administration considers it a waste of time" (DH1).

One teacher had this to say concerning the use of summary statistics in describing student

test scores:

I have found these to be impractical when one is in the field. I may be the only one (which I highly think I am not) not to use these principles but reactions from other teachers will confirm that we do not use most of the principles listed even though we are aware of them. The use of summary statistics can only work where the number of students per class is manageable. Even then, some of these principles are not necessary for us (GM7).

Irrelevance of Some Measurement Principles

Most of the teachers interviewed expressed the feeling that some of the principles taught in educational measurement are irrelevant to the assessment practices of classroom teachers.

These concerns are reflected in the comment made by one of the teachers:

What is taught in teacher-training institutions is too theoretical and some of that information is irrelevant in the practical situation. Such concepts as test reliability procedures, item analysis, test blueprinting, statistical details such as standard deviation or variance, are really not necessary in the assessment practices of the teacher (GM1).

There was some strong reservation among teachers regarding the value of statistical principles in classroom settings. Mathematics teachers, for instance, expressed the concern that the instructional value of statistical principles has not been demonstrated even though they know how to use the principles.

Teachers also acknowledged that they make very little use of statistical principles because these are not required by the Ministry of Education. "School inspectors in Kenya are only interested in whether teachers develop schemes of work, lesson plans, and student progress reports" (GM4).

However, when asked how a directive from the ministry requiring that teachers provide statistical details of student test scores would affect their practices, the same mathematics teacher (GM4) had this to say:

I will provide the details as required because my job security hinges on implementing ministry policy. However, I would still stand unconvinced about the real value of these principles in the classroom setting. These principles are only important if one intends to pursue further education (GM4).

One of the interviewees reported having not used descriptive statistics in describing student test scores in a three-year period following deployment by the TSC:

I must confess that ever since I started teaching (about three years ago), I have not used the mean test score, median, or percentiles...No one requires these details from teachers (DM5).

However, there was some consensus among the interviewees that, if required, statistical analyses of student test scores could best be handled through the establishment of a departmental unit within each school.

Some of the teachers also indicated that they have to sample as much content area as possible in assessing their students in order to prepare them adequately for the national examinations. However, they pointed out that in doing so, they do not have to sample interrelated items as often stressed in measurement courses. They indicated that national examinations are usually based on items drawn from several topics in the syllabus and that these items are not usually interrelated. In the words of one teacher:

Detailed procedures for determining the internal consistencies of tests and item analysis techniques belong to the professional test developer who is interested in studying item properties...Our use of tests is to keep instruction going. And we need test data that best provides a quick and general impression of student achievement in several topics in the syllabus (GH4).

Eighteen of the thirty-two teachers who were interviewed commented that they do not need to spend time calculating item difficulties because they usually get a general impression of these difficulty indices during the marking of student test scripts. These comments are reflected by

those made by one mathematics teacher:

I do not need to calculate the item difficulty in order to know how difficult a particular item was to the students. I usually get a general impression of this item difficulty when scoring student responses (DM4).

Wide Syllabuses

Other concerns related to student assessment also surfaced during personal interviews with teachers. These concerns related to the recent changes in the education system. Twenty-five of the thirty-two teachers interviewed complained about the wide syllabus to be covered before students sit for national examinations. These complains are reflected by the comments made by two history teachers:

The 8-4-4 system of education in Kenya is poorly planned. It involves a lot of work to be covered in a very short time. The syllabus is too wide for the teacher to cover and at the same time carry out meaningful assessment of students (GH8).

In order for us to cover the syllabus before students take form four examinations, we often have to teach during our private times, such as at nights or over the weekends. Important as it is, student assessment has assumed secondary importance, done only for purposes of ranking students at the end of the term (GH2).

Over 43% (106) of the teachers who responded to the mail survey felt that because of the wide syllabuses to be covered, there is hardly any time for them to review tests for clarity, representativeness of content domain, and to ensure that test items cover all the cognitive domains of interest. In the words of one teacher, "the whole thing (assessment) has been turned into a numbers game without regard to the quality of information conveyed by assessment scores" (GM6).

Some teachers also seemed unhappy with the recent introduction of the 8-4-4 system of

education and the factors related to its implementation:

The 8-4-4 system of education has contributed to low morale among teachers. We have not fully adjusted to the system and its inherent heavy workloads (DH4).

Teacher Assessment Needs

It was deemed necessary to identify teacher assessment needs, the ultimate purpose of which was to document and characterize the need for remediation or in-service training in measurement. Previous studies have shown that teachers are interested in improving their assessment skills.

The majority of the teachers interviewed (19) expressed concern about the rather general and theoretical nature of their measurement training, its relatively heavy focus on the use of statistics, the less time devoted to the teaching of measurement courses in general, and the non-existence of professional development workshops in the area of assessment for teachers in the field. These teacher concerns about the inadequacy of their training in educational measurement are exemplified by the following comments made by four different teachers:

Measurement instructors did not present realistic examples of the classroom use of measurement principles. The presentation of concepts tended to be in the abstract with few real life examples of how the concepts could operate in the classroom setting (GM2).

This course was not taught with the teacher in mind. Right from the first day of class, we were told that our assignments for the course would be based on the completion of three statistics exercises. I would personally have preferred to be taught how to develop a test and a marking scheme, rather than analyse data without full knowledge of the process that produced it (GM3).

We spent half the time for the course dealing with such concepts as statistics, test validity and reliability, regression analysis, and correlations of scores. Important areas such as test construction, development of marking schemes, and procedures for communicating test scores to students and parents were just mentioned in passing. Hence indications on the questionnaire showing that I was taught the concepts do not necessarily imply that I perceive the quality of instruction to have been adequate (GH8).

Unlike the case in primary schools, there are very few in-service courses or seminars for teachers at the secondary school level. These courses or seminars would help supplement our training in assessment, especially in important assessment activities such as test construction and marking (DH2).

The majority of the teachers interviewed (21) expressed concern that their training in measurement was inadequate in the areas of test construction, development of marking schemes, marking procedures, and the reporting of student assessment results. A few of the interviewees also felt that they had not been adequately prepared in the use of statistical techniques in the classroom.

Some of the diploma teachers who took their training around the time of the introduction of the 8-4-4 system of education in the nineteen-eighties felt that their training in measurement had been particularly hasty. "We went through a 'crash' programme in our training. We had to be deployed to schools as soon as possible because there was a serious teacher shortage in schools" (DH7). These teachers indicated that their instructors had not had enough time in which to take them through the procedures of test development and test scoring. They pointed out that the little time devoted to the teaching of measurement courses was spent on the teaching of statistical concepts at the expense of procedures they considered to be most important in the assessment of students.

A single lecture was devoted to test development. To be honest, no reference was made to the development of marking schemes and scoring procedures. We have had to learn these techniques from other teachers in the field (DM8).

The little time available should have been spent teaching principles that really matter in the assessment practices of the teacher. Statistical details should not have had a lot of emphasis (DM3).

Of the thirty-two teachers who participated in the interview, twenty-six of them indicated

that they would like to learn more about test development, development of marking schemes, and how to mark tests. They indicated that these assessment activities have a direct bearing on teacher routine assessment practices in the classroom. However, teachers decried the lack of professional development opportunities for teachers in the field. "With no supplementary seminars or in-service programmes in assessment techniques, one is bound to be deficient or 'rusty' in student assessment" (GM7).

Most of the interviewees (22) felt that in-service courses should be planned and delivered to teachers in the field, and that these should focus on assessment procedures suggested by teachers. The teachers also suggested that, where possible, in-service courses or seminars on the assessment of student achievement should be delivered by those teachers who develop and mark examinations administered by the Kenya National Examinations Council. According to the interviewees, teachers who have served with the Examinations Council, either as developers of examinations or examiners, have a better understanding of the complexity of assessment as carried out in the classroom than college instructors who are far removed from the classroom setting.

Ten of the interviewees expressed interest in learning more about the statistical techniques for summarizing student test scores. Surprisingly, their interest in learning more about these techniques did not stem from their desire to use them but from the concern that future ministry policy might call for the use of the same techniques.

In summary, there is a clear indication that some of the measurement principles are perceived as being irrelevant in the classroom. Teacher comments provide some indication that heavy teacher workloads resulting from the long syllabuses to be covered, large number of lessons to be taught, large class sizes, and the non-teaching workload requirements are some of the

teachers' major constraints when it comes to the use of measurement principles in classroom settings. It appears that these workload demands put pressure on the amount of time teachers have to devote to their assessment activities in the classroom. Comments on issues related to teacher workloads seem to focus on the overwhelming changes and demands associated with the introduction of the 8-4-4 system of education.

CHAPTER 5

DISCUSSION OF RESULTS

The results reported in chapter 4 are discussed in this chapter. The study findings are discussed in parallel with relevant findings from previous studies. This discussion, in order of presentation, includes teacher training in measurement, importance of measurement principles as seen by teachers, teacher use of measurement principles, and factors affecting this use in Kenya. The benefits and the limitations of the study are discussed in the latter part of this chapter.

Teacher Training in Measurement

The results clearly indicate that teachers in Kenya receive instruction in the fundamental aspects of educational measurement. In particular, teachers receive training in the procedures for test construction, test administration, marking, the use of descriptive statistics, test validity and reliability, item analysis, and the reporting of student assessment results.

The findings of this study, therefore, are at variance with those of studies elsewhere which show that many teachers have not been trained in the use of measurement principles (Coffman, 1983; Dorr-Bremme & Herman, 1984; Goslin, 1967; Lissitz & Schafer, 1987; Mayo, 1964, 1967; Newman & Stallings, 1982; Stiggins & Bridgeford, 1982; Wise, Lukin & Roos, 1991). However, the findings support other studies which show that whenever measurement training is provided, there is usually a substantial emphasis on the use of statistics (Gullickson, 1986; Gullickson & Hopkins, 1987). The teaching of test marking procedures do not receive as much emphasis.

The reportedly high percentage of teachers who have received training in the use of measurement principles comes as no surprise. Teachers in the setting of this study take a mandatory measurement course prior to their certification. The variation in the percentage of

teachers who received training in the use of various measurement principles may be a reflection of the difference in emphasis by different professors. It may also be that some of the teachers who were surveyed had forgotten about some of the measurement principles taught in college. This is to be expected when one takes into account the fact that some of the teachers graduated from college several years ago and have probably never had to use some of the measurement principles taught.

Bearing the methodological caveats in mind (small sample size of the interview subjects and non-objective teacher comments) there are two observations about training in measurement which have emerged from this study. Although other researchers (e.g., Gullickson, 1984) have reported that teachers are comfortable with their knowledge of testing and test use, this study and others (Leiter, 1976; Newman & Stallings, 1982) provide evidence of lack of teacher competence in testing techniques, particularly in test construction, marking, and the reporting of student assessment results. Although it has long been assumed that teacher incompetence in the assessment of student achievement would be remedied by training in educational measurement, this study has shown that whenever this training is provided, it does not usually address the assessment concerns of classroom teachers. This is reflected by the desire expressed by teachers for assessment programmes designed to equip them with the requisite skills in test construction, marking, and the reporting of student assessment results.

Although the assessment needs expressed by teachers in this study may be a reflection of the needs of teachers in the district where the interviews were conducted, there is strong reason to suggest that teachers in other parts of the province may have similar assessment needs. This is to be expected given that the teachers interviewed were trained in the same institutions as other

teachers in the province (also country).

It is also interesting to note that despite the small sample size of the teachers interviewed, the assessment needs of teachers in this study reflect the assessment needs of teachers in other settings (Carter, 1984; Green & Stager, 1985; Stiggins & Bridgeford, 1985). This similarity in the assessment needs of teachers in different settings provides some indication that teachers are inadequately prepared in measurement for their routine classroom assessment activities.

It appears that when measurement training is offered as a requirement for teacher certification, there is usually an assumption that teachers need to be well grounded in the quantitative aspects of measurement as practised by measurement experts. However, it seems that assessment in the classroom may not lend itself to the use of quantitative aspects of educational measurement. Stiggins (1992) notes that one of the major causes for the absence of assessment training in the teacher-training curriculum of some institutions has been the deep-seated mismatch between what teachers need to know about assessment and the content of assessment courses when they are offered. Gullickson and Hopkins (1987) have reported that measurement instructors often choose to concentrate on the application of statistical analyses when teaching measurement courses. These researchers found that whenever professors do have more time available, rather than giving more attention to test development and grading, they instead tend to emphasize statistics and standardized testing issues. Apparently, comments consistent with these earlier findings have emerged from this study.

Two issues that concern the appropriate instructional design for teacher preparation in measurement can be identified from this study. First, the amount of time devoted to the teaching of important measurement principles for the classroom teacher may not be enough for teachers to

be well grounded in the use of these principles.

Secondly, the teaching of measurement courses is probably presented in theoretical terms with few real life examples of how the procedures would work in classroom settings. Many teachers probably go through their measurement training with very little understanding of the assessment tasks and demands facing them. Although the pre-service teaching practice does allow teachers to practise what they have learned about assessment, the short duration of this exercise may not be enough for teachers to have a considerable understanding of the complexity of assessment in the classroom and whether the procedures taught in measurement can hold in practice.

The results of this study have revealed that teachers in the field have a clear understanding of the assessment demands facing them. Hence, the relevance and use of measurement principles is made much more apparent when one is faced with the realities of test development, marking, and grading. It would, therefore, be more profitable for those interested in improving training in measurement to talk with teachers regarding the intricacies of assessment in the classroom. It has been argued that teacher-training in tests and measurement needs to be reoriented (Fennessy, 1982; Gullickson, 1984; Newman & Stallings, 1982). Suggestions have been made that courses need to be teacher-oriented and more relevant to what actually goes on in the classroom and in the teacher's life. This study reaffirms this need.

Importance of Measurement Principles from the Teachers' Point of View

Another outcome of this study is that teachers attach the greatest importance to the preparation of tests (examinations), test administration and marking, and to the use of student assessment results for various instructional purposes in the classroom. These findings are in

keeping with the results of other surveys in North America and Nigeria (Borg, Worthen & Valcarce, 1986; Gullickson, 1986; Mayo, 1964, 1967; Onocha & Okpala, 1991) in which factors distorting students' test scores, use of achievement test scores for the diagnosis of student needs, determination of test validity, and reporting procedures were perceived to be important. The similarities in these findings therefore provide some clear indication that teachers, despite differences in context, value similar assessment procedures.

Like the findings reported by other researchers (Borg, Worthen, & Valcarce, 1986; Gullickson, 1986; Mayo, 1964, 1967; Onocha & Okpala, 1991), the use of statistics for summarizing student test scores and for evaluating test quality were perceived to be of less importance in the assessment practices of classroom teachers. However, unlike the findings in previous studies, the procedures for the determination of test-reliability were not perceived as being very important by teachers in this study.

Although there may not be many clear explanations for the differences in the findings above, one possible explanation may be offered. The difference in the findings could have occurred because of the differences in the wordings of the questionnaire items on test-reliability. Teachers in this study were asked to indicate how important it was for them to know how to determine test-reliability in assessing their students. In contrast, Borg et al. (1986) asked teachers to indicate how important it was for test reliability and validity to be taught in undergraduate measurement courses. Similarly, the teachers surveyed by Onocha and Okpala (1991) were asked to indicate whether it was important or not that the determination of test validity and reliability should be taught in pre-service teacher training programmes. Thus, teachers may find it very important to learn about the determination of test-reliability but may not find it very important to

compute test- reliability in dealing with student assessment.

The results of this study provide some further indication that teachers are positively disposed to measurement principles which have a direct bearing on their day-to-day assessment activities. Thus, the extent to which measurement instructors can link measurement concepts to classroom applications will probably determine to a large extent the value that teachers will place on the measurement course. It is particularly important to note that concepts which are not clearly related to classroom application (e.g. statistical analyses) tended to receive relatively lower importance ratings. Therefore, if measurement training is to meet the needs of classroom teachers, it should be oriented towards the teaching of assessment principles considered to be very important by teachers. This suggestion is bound to generate some debate among measurement experts. Questions can be asked as to whether the measurement curriculum should be designed around the perceived importance of topics by teachers who are exposed to it. Arguments can be advanced to the effect that the content of measurement courses should be decided by those who are most knowledgeable about the procedures of educational measurement (measurement experts). However, it is possible for those who are interested mainly in the principles of psychometrics not to effectively address the assessment concerns of classroom teachers.

Use of Measurement Principles in Schools

The findings with respect to the use of measurement principles reflect the results of previous studies in North America (Fleming & Chambers, 1983; Gullickson, 1982; Newman & Stallings, 1982; Reynolds & Menards, 1980). Like these prior studies, teachers in this study tend to use the short-answer, essay, problem-solving, and multiple-choice questions in assessing their students. However, unlike the findings from the studies referred to above, teachers in this study

reported a less frequent use of true-false and matching items.

A number of possible explanations may be offered for the differences above. First, the types of test items used in the assessment of student achievement may be a function of the grade level taught. In this study, the assessment practices of teachers at the secondary school level were investigated. Previous researchers (Gullickson, 1982; Newman & Stallings, 1982) studied teacher testing practices among elementary and secondary school teachers and found a frequent use of matching and true-false items. Therefore, the presence of elementary teachers among the teachers surveyed in these earlier studies could have accounted for the reportedly high frequency of use of these types of items. This could be true if matching and true-false items feature prominently in the assessment of student achievement at the elementary school level.

The influence of government examinations may also have accounted for the little use of true-false and matching items by teachers in this study. Many teachers indicated that they do not find it necessary to expose their students to true-false and matching exercises in the classroom because such items are not used for assessing students in national examinations.

The types of test items used and the scoring practices of teachers also vary as a function of the curriculum area taught. This is to be expected and the data in this study support this notion. Multiple-choice, short-answer, and essay items are used for assessing students in history. In contrast, problem-solving items are commonly used in the assessment of student achievement in mathematics.

With respect to test scoring (marking), one can conclude that the quality of assessment in history can be improved by procedures which allow teachers to score all the student responses to one essay question before scoring the responses to the next question. This procedure would help

ensure that some consistency is maintained in scoring. Procedures that involve the remarking of student test scripts in short-answer and essay items may not be possible given the time constraints faced by teachers in classroom settings.

Mathematics teachers on the other hand may improve their scoring consistency by remarking student test scripts. This may be possible given that student responses to mathematics tests do not involve much reading on the part of the examiner. However, in general, the use of a marking scheme (common marking scheme by two or more examiners) is probably the only feasible safeguard against scoring inconsistency by teachers involved in student assessment.

The findings with respect to the use of student assessment results clearly demonstrate that assessment results serve a variety of purposes in the classroom. Teachers perceive student assessment results as serving instructional purposes, both for feedback to the students and feedback to the teacher, with grading of students maintaining an important role in that feedback. These findings are similar to those reported by other researchers (Dorr-Bremme & Herman, 1984; Gullickson, 1982; Newman & Stallings, 1982; Stiggins & Bridgeford, 1985).

However, two apparent departures from the findings reported in the above studies are worth elaborating. Student assessment results in the context of this study do not play a significant role in student promotion, in the changing of students from one curriculum to another, and in the planning of specialized instruction for students.

A number of possible explanations may be offered for these differences in findings. First, the issues of student promotion and curriculum orientation to meet student needs do not lie directly with the classroom teacher, but with the Ministry of Education. In Kenya, there is automatic promotion of students at the secondary school level irrespective of student test

achievements in the preceding classes. This stems from the fact that student repetition of classes is discouraged because it would have serious financial implications for parents who would have to meet additional costs in school fees payments. In a country where tuition costs at the secondary school level have become almost out of reach for many parents, few of these parents would be willing to have their children repeat classes.

Secondly, the curriculum in most schools is fixed due to ministry guidelines. Each school in the country has to choose a certain set(s) of curricular areas from a number of curriculum alternatives based on the availability of resources required for the implementation of the chosen curriculum. Hence, by virtue of these curriculum alternatives, some students may take the curriculum for which their assessment results are least promising.

In some cases, a student's choice of a particular curriculum may be determined by the prevailing job market conditions. Thus, a student may opt to take a curriculum that offers better job prospects even if that particular student's assessment results are not promising in the curriculum chosen. To some extent, this may limit the extent to which teachers can use student assessment results for the purpose of tailoring student strengths to the appropriate curriculum.

Additional findings of this study provide some clear indication that teachers are concerned about the validity and reliability of their tests. They are concerned about practical steps they can take to ensure that test results are reliable and valid for the purposes for which they are designed. As a way of improving the quality of their assessments, many teachers take steps to ensure that test results are not tainted by student cheating. Students are supervised during the course of their tests. Such findings are to be expected in a setting where teachers take part in supervising students during the administration of national examinations. These findings are consistent with the

findings in previous surveys (Dorr-Bremme & Herman, 1984; Plake, Impara, & Fager, 1993). These studies revealed that test administration in schools is usually marked by formal testing procedures and that teachers are quite knowledgeable about the use of these procedures. The findings are also similar to those reported by Gullickson (1982, 1984) in which teachers reported actively structuring the testing situation in order to discourage cheating by students. However, contextual factors such as large numbers of examinees and lack of facilities (i.e., chairs and desks) do not allow teachers to carry out effective test administration.

The findings above clearly suggest that teachers understand and appreciate the need for sound assessment data. However, many of them consider the psychometric guidelines for dealing with assessment quality as being impractical in the classroom situation. Many teachers do not appreciate the value of determining item difficulties and the reliability coefficients associated with tests administered to students. Such quantitative measures are seen as being somewhat irrelevant and often time consuming in an environment in which teachers have to manage very full and demanding schedules.

It appears that teachers are concerned about sources of measurement error that can arise from student conduct during test administration, problems with the assessment environment, and problems with the scoring process especially when this is subjective in nature. The implication of these findings is that measurement instructors ought to point out to prospective teachers the potential pitfalls inherent in the assessment process. They also need to show them how to avoid the pitfalls in very practical and specific terms that translate into efficient actions teachers can take to maximize the quality of their assessments.

From this study, one may conclude that the elements which characterize much of the

evaluation practice, e.g. test-blue prints, item analysis, and statistical manipulation of student test scores, are not very common in classroom settings. These findings are consistent with the results of other surveys (Gullickson, 1982; Gullickson & Ellwein, 1985; Marso & Pigge, 1988; Mayo, 1967). The findings, therefore, lend some credibility to McLean's (1985) contention that the fundamentals of measurement are not very evident in classroom settings.

Teacher failure to develop and use test- blueprints suggests that teacher-developed tests may not represent an adequate sampling of the universe of content taught and may fail to include the use of items that test students on higher cognitive skills. This may partly account for the frequent use of items which test students on low level cognitive skills. Problems of this kind in teacher-made tests have been widely reported by other researchers (Fleming & Chambers, 1983).

The results pertaining to teacher use of statistical analyses are not surprising. First, descriptive statistics, item analysis, and test reliability procedures are generally not accorded much importance by teachers. It would, therefore, seem unlikely that teachers would make a profound use of the procedures in the assessment of student achievement.

Secondly, it has been argued that teachers may perceive statistical analyses as calling for more work than is justified by the benefits. Nimmer (1984) pointed out that quantitative measures associated with validity, reliability, and item analysis tend to be time consuming and appear to be of little practical value to the classroom teacher. Meanwhile, Gullickson (1986) also noted that the omission of statistical analyses of student test scores does not seem to have any obvious negative consequences in the teacher's assessment practices. Apparently, comments in support of these views have emerged from this study.

It is also interesting to note that teachers in this study indicated the use of percentiles and

test-retest reliability procedures. However, one would have expected them to report no use of these procedures especially given the low importance rating of the procedures and the time limitations in schools. It can therefore be concluded that teachers probably confused percentiles with percentage scores. In addition, they may have very little understanding about test-retest reliability. This reinforces the finding by Gullickson and Ellwein (1985) in which teachers reported computing test reliability coefficient without the understanding of the procedures involved in the computation of this index.

In general, it can be concluded that while there is a considerable agreement among measurement specialists regarding the importance of statistical analyses, there is lack of empirical evidence to show the positive instructional benefits of these analyses. As Stiggins, Conklin, and Bridgeford (1986) pointed out, it appears that until the measurement community comes to understand assessment in terms relevant to the teacher and translates its concepts into these terms, it will be difficult to change teachers' perceptions regarding the importance of statistical concepts in the classroom.

The results of this study provide some clear indication that teachers are more concerned with routine assessment practices such as test construction, marking, grading, reporting, and not with the use of fundamental measurement principles. In the classroom, teachers want answers to such questions such as: How do I construct the end of month test? How do I mark it quickly and fairly in order to get information that will guide the next phase of instruction? How best can I communicate the assessment information to the student's parent or to the school counsellor in order to help the student who is not doing well? Such questions are of utmost importance to teachers who are constantly faced with the need to obtain feedback on the effectiveness of their

instructional approaches and who have to provide information to multiple audiences that are interested in the results of student achievement.

There is, therefore, a need for those teaching measurement courses to distinguish between the measurement requirements of professional test developers and those of classroom teachers. Professional test developers may be concerned with single-test situations such as the large scale standardized test programmes and graduation examinations in schools. These single-test situations may lend themselves to systematic research but may be of little interest or use to classroom teachers.

In this study there were no significant differences between diploma and graduate teachers in their use of measurement principles. These findings are to be expected since student assessment in schools is a collaborative exercise among teachers of different qualifications. This collaboration stems from the way teaching is organized in schools. Teaching in the context of this study is organized in such a way that graduate and diploma teachers can teach different classes within the same grade. This arrangement means that diploma and graduate teachers teaching the same curriculum area will be guided by common assessment procedures. Thus, the use of measurement procedures does not appear to be related to teacher qualification but to factors within the context in which student assessment takes place. Such a finding is consistent with the view expressed by Earl and Cousins (1993) that assessment is determined somewhat by the environment in which it occurs (school).

Factors Affecting Use of Measurement Principles in Kenya

A number of observations can be made about the factors which seem to influence the application of measurement principles in classroom settings. Although the relative influence of

various factors on the use of measurement principles were not conclusive in this study, they certainly provide some indication as to what the major factors happen to be. Ministry of education policies on examinations (and otherwise) are seen as having a bearing on two aspects of classroom assessment. These aspects include the types of questions teachers use in assessing their students and the uses of assessment results in the decisions involving student promotion and curriculum orientation.

There is some indication in this study that teachers tend to imitate the national examinations in every test. The types of questions used for assessing students in the classroom and the test administration and marking procedures used may all reflect the procedures used in national examinations. Thus, it appears that if the types of questions used for assessing students in the national examinations do not include items which require students to apply their knowledge, teachers will likely not use items that call for the application of knowledge. In effect, this is a clear indication that the evaluation of student achievement by teachers in the classroom occurs within a procedural framework which is to some extent determined by outsiders. This phenomenon is not new. Wilson (1990) reported that externally mandated testing programmes have an influence on assessment practices in the classroom. In his study of the assessment practices of British Columbia teachers, he found that the anomaly of the use of matching items at the secondary school level was due to the appearance of these types of items on some government examinations. Such findings undercut the criticism that has often been levelled against teachers for using items that test students on low level cognitive skills.

Heavy teacher workloads resulting from the teaching and non-teaching work requirements reduce the amount of time teachers can commit to their assessment activities.

From the results of this study, one can conclude that teachers who are uncomfortable with the use of mathematical concepts may find themselves handicapped when it comes to the application of statistical analyses. This seems to reinforce the view held by some researchers (Paulos, 1988; Randhawa, 1992) that a good background in mathematics may be necessary in the application of statistical analyses. However, this study did not reveal any significant differences between mathematics and history teachers in the use of statistical analyses. This suggests that factors other than teacher aptitude in mathematics may be influencing the use of statistical analyses in classroom settings. It appears that the low usage of statistical analyses may lie mainly with what teachers see as the irrelevance of these procedures to their assessment activities.

Benefits of the Study

Unlike previous studies (e.g., Gullickson, 1982; Gullickson & Ellwein, 1985; Marso & Pigge, 1988), this study was more comprehensive because it was designed to investigate teacher use of measurement principles in several assessment areas, i.e., in test construction, administration, marking, summarizing of test scores, evaluation of test quality, and the reporting of student assessment results. Therefore, the study has provided some indication regarding the principles of measurement required by classroom teachers.

There is a clear indication in this study that teachers are more disposed to the use of those measurement principles which have a direct bearing on routine classroom assessment activities (i.e., test construction, administration, marking, and reporting). Quantitative aspects of measurement such as item analysis and procedures for determination of test reliability do not appear to be of prime concern to many classroom teachers.

The factors influencing the use of measurement principles in classroom settings were also

investigated. This is an area that had not been addressed before. Therefore, the study acted as a vital link between prescription and practice in the use of measurement principles.

In this study, a number of problems related to the use of measurement principles in classroom settings have been uncovered and brought into sharper focus. Addressing these challenges may pave the way for improved use of educational measurement principles in schools.

Limitations of the Study

In Kenya, many teachers teach several curricula. It is likely that some of these teachers did not follow directions carefully and responded with regard to a curriculum other than that specified.

The findings reported for the interview data need to be interpreted with some caution because the number of cases is not high. Although generalizability was not an objective of the qualitative phase of this study, the results of the interviews may apply to teachers in the district where the interviews were conducted. There is ample evidence also to suggest the generalizability of the results to other settings in the province (those faced with the same constraints in the use of measurement principles). It is important to note that many of the teachers who participated in the mail survey expressed concerns (i.e., through teacher comments) similar to those expressed by the teachers who were interviewed.

This study was conducted during a very busy time in the school calendar. The month of July during which the survey data were collected was busy with internal examinations, games, and athletic competitions. In addition, the month of September during which the interviews were conducted was also a busy one for teachers as they were preparing students for national examinations. These commitments on the part of teachers could have affected to some extent the

responses teachers gave.

In this study, teacher practices with respect to the use of measurement principles were investigated. However, it is widely known that the responses of individuals regarding their behaviours are usually susceptible to response biases especially when practices do not conform to professional expectations (Borg & Gall, 1989). The temptation to report what is accepted professionally could have been alluring for some of the respondents. Yet it is a fact of experience that words and deeds sometimes differ. Therefore, a response bias on the part of some teachers could have influenced the results.

It is also important to note that teaching is not a highly regarded profession in Kenya. However, because it has become increasingly difficult for school leavers to get jobs in other professions, some of the teachers could have taken up teaching as a career because employment is still guaranteed in this profession. Hence, some of the respondents could have included teachers for whom teaching was not a first choice career. Such teachers might have given superficial responses to some of the questionnaire items.

In addition, teachers in only one province in Kenya were surveyed. There are eight provinces in the country. A national study would have been preferable but was not possible because of lack of time and financial constraints facing the researcher. Also, teachers in two curricular areas and qualification levels were surveyed in the study. Teachers in other curricular areas and those in primary schools did not participate in this study.

As is true of all surveys, some teachers did not respond to the questionnaire (38.5%). Of those who responded (246), 133 were graduate and 113 were diploma teachers respectively. By curriculum area of instruction, 125 mathematics teachers responded to the survey compared to

121 history teachers. About 39% (154) of the teachers did not respond to the questionnaire. It is therefore not clear how the results could have turned out if the non- responding teachers had chosen to respond.

Despite these potential short comings, the study was conducted in the province in which the writer was born, educated, and had worked as a teacher for seven years. To some extent, the writer was probably known to many of the teachers who responded to the survey. This familiarity, together with the assurance of anonymity and confidentiality, should have encouraged the teachers to provide valid responses.

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The results of this study provide some clear indication that teachers in Kenya receive training in the fundamental aspects of educational measurement. However, some teachers do not consider themselves to have been adequately prepared in the areas of test development, marking, and the reporting of student assessment results.

The major criticism that has been levelled against measurement courses as taught in teacher-training institutions is the heavy emphasis on the use of statistics at the expense of assessment procedures considered to be most important in the assessment of student achievement. From this study, one may conclude that the major educational measurement principles for teachers are those in the areas of test construction, test administration, marking, and the reporting of student assessment results. However, the use of descriptive statistics, item analysis, and procedures for computation of test reliability are not considered to be of major importance in the assessment practices of classroom teachers.

Although certain measurement principles are perceived as being very important, a number of obstacles make it rather difficult for teachers to make an effective use of these principles in the assessment of student achievement. Teaching workloads, wide syllabuses, large class sizes, and non-teaching work requirements put pressure on the amount of time teachers can spend on test preparation and other related assessment activities. With class sizes of 40-55 students and student loads of 360 or more, teachers foresee little likelihood of effectively applying many of the measurement principles in the assessment of student achievement. These findings lend some support to Stiggins' (1992) recent claim that "teachers manage very full and demanding schedules"

(p.101).

There are a number of recommendations that can be derived from this study. First, educational measurement courses should probably focus more on the assessment procedures perceived to be of major importance in the assessment of student achievement. In fact, it is of utmost importance that deliberations on content of measurement courses should begin with the consideration for their relevance and impact on the assessment practices as practised by teachers in classroom settings. Measurement instructors will provide relevant assessment techniques only when they understand the classroom assessment environment and the demands it places on the teacher. Measurement instructors, therefore, ought to be aware of and appreciate the complexity and very demanding workload requirements faced by teachers who may be teaching more than 250 students a day. Such working knowledge of assessment in the classrooms serves two main purposes. It enables those providing instruction in measurement to form a clear vision of what it is that teachers need to know. It also helps them to develop some sensitivity to the teacher's life in the classroom. This may help establish credibility with teachers who can in turn act as a valuable source of practical ideas about which measurement principles are both beneficial and practical.

In order to increase the probability that teachers can effectively use relevant measurement principles, teachers must have a chance to practice the use of the principles under adequate supervision from their instructors. Those training teachers in measurement should therefore take advantage of instructional strategies that involve teachers in doing assessment. Where possible, assessment procedures should be modelled in practical terms so that teachers can appreciate them. The rather abstract approach with which measurement concepts are presented during training should be reduced by involving teachers in assessment simulations, applications of relevant

assessment principles to real-world problems, and by using assignments that make teachers critical consumers of the assessments that guide the lives of their students. These assessment simulations can best be realized through such assessment procedures as test construction, development of marking schemes, and actual marking. Establishing links with teachers in the field will probably enable teacher-trainees to have access to samples of tests administered to students and student responses to such tests. A critique of the strengths and weaknesses of these tests, coupled with the marking of samples of student responses to such tests (marking of dummies) will probably make it possible for teacher-trainees to deal with practical and beneficial assessment activities. These strategies engage teachers and enable them to see their assessment world through their training. It is only by linking measurement training with practical and beneficial assessment procedures that teachers can be provided with the principles which are most appropriate in the classroom. Hence, making educational measurement a compulsory aspect of teacher training without ensuring that the principles taught have relevance to classroom assessment will only lead to pre-service teachers going through the paces just for the purpose of certification.

Secondly, one must also recognize the fact that some of the teachers currently practising in schools graduated from programmes that did not provide adequate assessment training in some vital classroom assessment procedures. This study and others (e.g., Carter, 1984) have revealed that teachers are interested in improving their assessment skills. It may, therefore, be necessary for the Ministry of Education to organize in-service courses or seminars for these teachers in order to equip them with the desired assessment principles.

Those organizing in-service courses should consider the possibility of involving teachers with sound assessment experience in the difficult task of teaching assessment principles to

teachers who have assessment needs. In particular, those teachers who set, and those who mark national examinations are cognizant of sound assessment ideas and do constitute a vital training resource which should be tapped. There is evidence that teachers regard their colleagues as one of the most important sources of ideas for student assessment (Stiggins & Bridgeford, 1982).

In-service courses in student assessment should allow for participation by teachers. It would in fact be a mistake if participants came expecting to sit back and be instructed. Participants should have at least an active involvement in student evaluation activities, modelled in practical ways that enable teachers to see how the assessment procedures taught could work in their classroom settings. Most of the time should be spent sharing concerns and experiences instead of participants listening to someone outline how assessment should be done. It is also important that teachers discuss practical steps they can take to ensure that the assessments in their classrooms yield high quality data (validity and reliability). These steps may hold promise for teachers who may not have the time or find it relevant to carry out quantitative measures of test reliability.

In order to help facilitate the effective use of relevant assessment principles, it may be necessary for the ministry (TSC) to deploy more teachers to schools so as to cope with the large class sizes and the resulting heavy workloads which have become typical of most schools in the country in recent times. Addressing teachers' workload demands may help release more time for teachers to enhance their commitment to the use of assessment principles.

The results of this study, however, cannot be considered totally definitive. Further research is needed to establish the extent to which regional variation exists with respect to teacher use of measurement principles and whether factors affecting this use can be identified reliably in

other settings. In view of the fact that the relative influence of various factors on the use of measurement principles was not very clear in this study, it is recommended that future studies involve some modification in the wording of some of the items in the study instrument (e.g., use of lack of time instead of availability of time). Such additional studies should also be used to determine the extent to which factors such as class size and the teachers' workload affect the use of measurement principles. These factors were apparent in the qualitative phase of this study.

Regarding teacher assessment needs, more research is called for to determine the needs of teachers on a larger scale. Such a study should provide information which is needed in the development and implementation of strategies to meet the designated needs.

As for teacher-training in measurement, the major concern with the results reported in this study is not so much whether one has had the training or not but rather it is whether teachers perceive the training to have been relevant and adequate. It is in fact a massive oversimplification to treat teacher-training in educational measurement principles as a dichotomous variable (taught or not taught). Researchers in future could examine the relevance and adequacy of measurement courses, the emphasis laid on various aspects of measurement, and the amount of time devoted to the teaching of these courses.

It also seems natural to assume that teachers would apply measurement principles in their assessment activities if they were more knowledgeable about educational measurement. However, as the results of this study have shown, this may not necessarily be true. Hence, the persuasion of teachers to use measurement principles will inevitably require that the measurement community determines which principles are both beneficial and practical in classroom settings. These are important but challenging tasks. However, they seem necessary because of the growing public call

for better assessment of students and improved student achievement in national examinations.

REFERENCES

- Anderson, J. O. (1989). Evaluation of Student Achievement. Teacher Practices and Educational Measurement. Alberta Journal of Educational Research, 35, 123-133.
- Bejar, I. I. (1984). Educational diagnostic assessment. Journal of Educational Measurement, 21, 175-189.
- Borg, W. R., Worthen, B. R., & Valcarce, R. W. (1986). Teachers' perceptions of the importance of educational measurement. Journal of Experimental Education, 55, 9-14.
- Borg, W. R., & Gall, M. D. (1989). Educational Research. An Introduction (5th Edition). White Plains, New York: Longman.
- Briggs, D. (1970). Influence of handwriting on assessment. Educational Research, 13, 50-55.
- Burstein, L. (1983). A word about this issue. Journal of Educational Measurement, 20, 99-101.
- Bushway, A., & Nash, W. R. (1977). School cheating behaviour. Review of Educational Research, 47, 623-632.
- Carter, K. (1984). Do teachers understand principles for writing tests?. Journal of Teacher Education, 35, 57-60.
- Chase, C. I. (1968). The impact of some obvious variables on essay test scores. Journal of Educational Measurement, 5, 315-318.
- Coffman, W. E. (1983). Testing in the schools. A historical perspective. Paper presented at the Centre for the Study of Evaluation Annual Invitational Conference,

University of Los Angeles.

Cooke, D. (in press, 1994). Assessment In Schools

Cox, S. (1981). Teachers and standardized achievement tests: What's really happening?.

Phi Delta Kappan, 62, 631-634.

Criswell, J. R. (1992). Preparing teachers for the '90S: Redirecting the content of measurement courses. Education, 113(3), 465-472.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Densmore, K. (1989). Professionalism, proletarianism and teachers' work. In T. Popkewitz (Ed.), Critical studies in Teacher Education (pp. 148-149). Lewes: Falmer Press.

Dorr-Bremme, D. W., & Herman, J. (1984). Teachers and testing: Implications from a National Study. Los Angeles: California University, Centre for the Study of Evaluation. (ERIC Document Reproduction Service No. ED. 244 987).

Earl, L., & Cousins, J. B. (1993). Enhancing classroom assessment. In K. Leithwood, A. Hargreaves & D. G rrin-Lajoie (Eds.), Years of Transition: Times for Change (pp. 2-31). Toronto, ON: MGS Publications Services.

Farr, R., & Griffin, M. (1973). Measurement gaps in teacher education. Journal of Research and Development in Education, 7, 19-28.

Fennessy, D. (1982). Primary teachers' assessment practices. Some implications for teacher training. Paper presented at the annual conference of the South Pacific Association for Teacher Education, Victoria (Australia). (ERIC Document Reproduction Service

No. ED. 229 346)

- Fleming, M., & Chambers, B. (1983). Teacher-made tests - windows on the classroom. In W. E. Hathaway (Ed.), Testing in the Schools: New directions for testing and Measurement, 19 (pp. 29-38). San Francisco: Jossey-Bass.
- Fluit, J., & Gifford, C. (1980). Who's teaching teachers how to teach test-wiseness?. Contemporary Education, 51, 152-154.
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. Educational Measurement: Issues and Practice, 7(1), 25-35.
- Frisbie, D. A., & Ebel, R. L. (1986). Essentials of educational measurement. New Jersey: Prentice-Hall, Inc.
- Goslin, D. A. (1967). Teachers and Testing. New York: Russel Sage Foundation.
- Green, K. E., & Stager, S. F. (1985). Teachers' attitudes toward testing. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL. (ERIC Document Reproduction Service No. ED. 253 588).
- Green, K. E. (1992). Differing opinions on testing between preservice and inservice teachers. Journal of Educational Research, 86(1), 37-42.
- Gronlund, N. E. (1973). Preparing Criterion-referenced tests for classroom instruction. New York: MacMillan.
- Gronlund, N. E. (1976). Measurement and Evaluation in teaching. New York: MacMillan Publishing.
- Gronlund, N. E. (1981). Measurement and Evaluation in teaching. New York: MacMillan Publishing.

Gullickson, A. R. (1982). The practice of testing in elementary and secondary schools.

Paper presented at the Rural Education Conference at Kansas University,

Manhattan. (ERIC Document Reproduction Service No. ED. 229 391).

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests: Journal of Educational Research, 77, 244-248.

Gullickson, A. R., & Ellwein, M. C. (1985). The goodness-of-fit between prescription and practice: Post hoc analysis of teacher-made tests. Educational Measurement: Issues and Practice, 7(10), 15-18.

Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23, 347-354.

Gullickson, A. R., & Hopkins, K. D. (1987). The context of educational measurement instruction for preservice teachers: professor perspectives. Educational Measurement: Issues and Practice, 6(3), 12-16.

Hambleton, R. K., Anderson, G. E., & Murray, L. N. Applications of Microcomputers to Classroom testing. In Hathaway, W.E. (1983). New Directions for testing and Measurement. Testing in the schools: Washington. Jossey-Bass Inc., Publishers and Jossey-Bass, No. 19.

Hickman, J. A. (1991). Does formal training in tests and measurements affect performance on teacher certification exams?. Education, 112(1), 72-75.

Hills, J. R. (1977). Coordinators of accountability view teachers' measurement competence. Florida Journal of Educational Research, 19, 39-44.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). Educational and Psychological Measurement and Evaluation. New Jersey: Prentice-Hall.

- Kenya Government (1968). The Teachers' Service Act: Chapter 212 of the Laws of Kenya (revised edition). Nairobi: Government Printer.
- Kubiszyn, T., & Borich, G., (1993). Educational testing and measurement: Classroom application and practice. New York: Harper Collins College Publishers.
- Lambert, R. F. (1980). Teacher attitudes on testing: A multiple perspective. College Board Review, 118, 13-14, 29-30.
- Leiter, K. C. W. (1976). Teachers' use of background knowledge to interpret test scores. Sociology of Education, 49, 59-65.
- Lissitz, R. W., & Schafer, W. D. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38, 57-63.
- Markham, L. R. (1976). Influences of handwriting quality on teacher evaluation of written work. American Educational Research Journal, 13(4), 277-283.
- Marshall, J. C., & Hales, L. W. (1971). Classroom test construction. London: Addison-Wesley Publishing company.
- Marso, R. N., & Pigge, F. L. (1988). Ohio secondary teachers' testing needs and proficiencies: Assessments by teachers, supervisors, and principals. American Secondary Education, 17, 2-9.
- Mayo, S. T. (1964). What experts think teachers ought to know about educational measurement. Journal of Educational Measurement, 1, 79-86.
- Mayo, S. T. (1967). Preservice preparation of teachers in educational measurement. (ERIC Document Reproduction Service No. ED. 021 784)
- McLean, L. D. (1985). The craft of student evaluation in Canada. Toronto: Canadian Education Association.

Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and Evaluation in Education and Psychology. Toronto: Holt, Rhinehart and Winston, Inc.

Mehrens, W. A., & Lehmann, I. J. (1987). Using teacher-made measurement devices. NAASP Bulletin, 71(496), 36-44.

Mehrens, W. A., & Lehmann, I. J. (1973). Measurement and Evaluation in Education and Psychology. Toronto: Holt, Rhinehart and Winston, Inc.

Ministry of Education (Kenya) (1993). Graduate Diploma in Education: Teacher Education Syllabus. Nairobi: Kenya Institute of Education.

Mouly, G. J. (1978). Educational Research. The art and science of investigation. Boston: Allyn & Bacon.

National Council on Measurement in Education (1990). The Standards for Teacher Competence in Educational Assessment of Students. Washington, DC.

Newman, D. C., & Stallings, W. M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at the American Educational Research Association annual conference, New York. (ERIC Document Reproduction Service No. ED. 220 491).

Nimmer, D. N. (1984). Measures of validity, reliability, and item analysis for classroom tests. Clearing House, 58(3), 138-140.

Noll, V. H. (1955). Requirements in educational measurement for prospective teachers: School and Society, 82, 88-90.

Nunally, J. (1978). Psychometric theory. New York: McGraw-Hill.

Obanya, P. A. I. (1984). 'Implementing a continuous assessment programme in the secondary school'. In: ADESINA (Ed.), Secondary Education in Nigeria (pp. 56-

- 71). Ife, Nigeria: University of Ife Press.
- Onocha, C. O., & Okpala, P. M., (1991). Teachers' perceptions of the relevance of topics taught in educational measurement courses. Educational Research, **33**, 228-232.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. Educational Measurement: Issues and Practice, **12**(4), 10-12.
- Paulo, J. A. (1988). Innumeracy: Mathematical illiteracy and its consequences. New York: Vintage Books.
- Popham, W. J. (1990). Modern Educational Measurement. A practitioner's perspective. Toronto: Prentice-Hall Canada Inc.
- Principles for Fair Student Assessment Practices for Education in Canada (1993). Edmonton, Alberta: Joint Advisory Committee.
- Randhawa, B. S. (1992). Construction of curriculum relevant tests by teachers and experts. In D. J. Bateson, Classroom testing in Canada (pp.38-50). Proceedings of the second Canadian Conference on classroom testing. The Centre for Applied Studies in Evaluation, University of British Columbia.
- Reynolds, W. M. & Menards, K. A. (1980). An examination of teachers' practices. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston.
- Roeder, H. H. (1972). Are today's teachers prepared to use tests? Peabody Journal of Education, **59**, 239-240.
- Rudman, H. C., Kelly, J. L., Wanous, D. S., Mehrens, W. A., Clark, C. M., & Porter, A. C. (1980). Integrating assessment with instruction: A review (1922-1980). East Lansing: Michigan State University, College of Education, Institute for Research on

Teaching (ERIC Document Reproduction Service No. ED. 189 136)

Shephard, L. A. (1989). Why we need better assessments. Educational Leadership, 46, 5-8.

Shore, J. L. (1971). Preservice experiences and teaching responsibilities for specialists in reading. Paper presented at the meeting of International Reading Association, Atlanta City. N.J. (ERIC Document Reproduction Service No. ED. 052 910).

Smith, F. M., Adams, S. (1972). Educational measurement for the classroom teacher. New York: Herper & Row.

Smith, B. O., Silverman, S. H., & Borg, J. M. (1980). A design for a school of pedagogy. Washington, DC: (US Department of Education No. E-80-4200).

Stiggins, R. J. (1988). Revitalizing classroom assessment. The highest instructional priority. Phi Delta Kappan, 65, 363-368.

Stiggins, R. J. (1990). Toward a relevant classroom assessment research agenda. Alberta Journal of Educational Research, 36, 92-97.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. Educational Measurement: Issues and Practice, 10(1), 7-12.

Stiggins, R. J. (1992). Making assessment training relevant for teachers. In D. J. Bateson, Classroom testing in Canada (pp. 97-103). Proceedings of the second Canadian Conference on classroom testing. The Centre for Applied Studies in Evaluation, University of British Columbia.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N.J. (1986). Classroom assessment. A key to effective education. Educational Measurement: Issues and Practice, 5(2), 5-17.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

- Stiggins, R. J. (1985). Improving assessment where it means the most: in the classroom. Educational Leadership, 42, 69-74.
- Stiggins, R. J., & Bridgeford, N. J., (1982). Final research report on the nature, role and quality of classroom performance assessment. Portland, OR: Northwest Regional Educational Laboratory.
- Sumner, R. (1987). The role of testing in schools. Berkshire: NFER-NELSON Publishing Company Ltd.
- Wahlstrom, M. W., & Danley, R. R. (1976). Assessment of student achievement in Ontario. Toronto: Ontario Ministry of Education.
- Ward, J. G. (1983). Teachers and testing: A survey of knowledge and attitudes. In L. M. Rudner (Ed.), Testing in Our Schools. Proceedings of the NIE Invitational Conference on Test Use. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED. 237 579).
- Ward, J. G. (1982). An overview of the AFT's teaching and testing. In S.B. Anderson & L.B. Coburn (Eds.), Academic testing and the consumer (pp.47-52). San Francisco: Jossey-Bass.
- Wilson, R. J. (1989). Evaluating student achievement in an Ontario high school. Alberta Journal of Educational Research, 35, 134-144.
- Wilson, R. J. (1990). Classroom processes in evaluating student achievement. Alberta Journal of Educational Research, 36, 4-17.
- Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education, 42(1), 37-42

APPENDIX A

The Research Permit

APPENDIX A

The Research Permit

PAGE 2

THIS IS TO CERTIFY THAT:

Prof./Dr./Mr./MissANDREW.....
 XXXXXX K CHIRCHIR
 of (Address) ...P.O.BOX.261.....
LONDANI.....
 has been permitted to conduct research in.....
SECONDARY SCHOOLS..... Location,
KERLCHO & BOMET..... District,
RIFT VALLEY..... Province,
 on the topic ..THE RELATIONSHIP.....
 ..BETWEEN TEACHER TRAINING IN.....
 ..MEASUREMENT AND CLASSROOM.....
 ..ASSESSMENT IN KENYAS SECONDARY.....
 ..SCHOOLS.....
 for a period endingSEPTEMBER..., 19...94

PAGE 3

Research permit No.QP/12/001/24C110...

Date of issue ..27-5-94.....

Fee received ..Kshs.100/-.....

Applicant's
Signature

for

 R MUSE WANASAKAAMI
 Permanent Secretary,
 Office of the President

APPENDIX B

The Survey Questionnaire

APPENDIX B

The Survey Questionnaire

A Survey of

Teacher Perceptions of the

Importance and Use

of Educational Measurement

Instructions:

This questionnaire is addressed to History and Mathematics teachers. In filling the questionnaire, kindly limit yourself to the principles as they apply only to either History or Mathematics and not any other subject that you may be teaching.

HOW TO COMPLETE PARTS A-H OF THE QUESTIONNAIRE

The following information applies to PARTS A-H of the questionnaire. You are asked to answer three questions regarding a number of items relating to a specific topic. The stem of each question is typewritten in **boldface** letters. To read the complete question, simply add each of the numbered items found in the first column to the end of each stem. Then, circle the option corresponding to your choice of answer.

Example:

Below is the first issue to be considered in PART A: USES OF STUDENT ASSESSMENT RESULTS. This section is reproduced below:

	How often do you use student assessment results to:	How important is it for you to know how to use student assessment results to:	Were you ever taught to use student assessment results to:
1. assign grades to students?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

The three questions to be answered for the first issue are:

1. **How often do you use student assessment results to assign grades to students?**
2. **How important is it for you to know how to use student assessment results to assign grades to students?**
3. **Were you ever taught to use student assessment results to assign grades to students?**

If your answers to questions 1, 2 and 3 were "Always", "Very Important" and "Yes", respectively, you would then circle the option corresponding to your choice. The complete response to this first item is provided below.

	How often do you use student assessment results to:	How important is it for you to know how to use student assessment results to:	Were you ever taught to use student assessment results to:
1. assign grades to students?	Never Rarely Often <u>Always</u>	Not at all Important Slightly Important Moderately Important <u>Very Important</u>	<u>Yes</u> No

Complete the responses for each of the remaining numbered items in the same manner.

PART A USES OF STUDENT ASSESSMENT RESULTS

	How often do you use student assessment results to:				How important is it for you to know how to use student assessment results to:				Were you ever taught to use student assessment results to:	
1. assign grades to students?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
2. change a student from one curriculum to another?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
3. promote students from one class to another?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
4. counsel students in their academic work?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
5. aid students in their career planning?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
6. provide students with feedback on their academic progress?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
7. diagnose student needs?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
8. plan remedial instruction or work for academically "weak" students?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
9. plan advanced instruction or work for academically "strong" students?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
10. modify teaching techniques?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No

PART B: PRINCIPLES OF TEST CONSTRUCTION

	When constructing a test, how often do you:	When constructing a test, how important is it for you to:	When constructing a test, were you ever taught how to:
1. define content areas or topics to be tested?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
2. develop a test blueprint (table of specifications)?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
3. set test questions (e.g., multiple- choice, essay, etc.) to match the cognitive level (e.g., recall, application, synthesis, etc.) to be tested?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
4. use item-writing guidelines in the development of questions?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
5. develop items that reflect the instructional objectives?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
6. assess the degree to which the items reflect the content domains to be tested?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
7. review the test questions for possible bias, ambiguity, etc.?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

PART C TYPES OF TEST QUESTIONS AND OTHER ASSESSMENT METHODS

	When assessing student achievement, how often do you use:				When assessing student achievement, how important is it for you to use:				Were you ever taught the use of:	
1. multiple-choice questions?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
2. true-false questions?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
3. essay-type questions?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
4. short-answer questions (e.g., two or three sentence answer)?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
5. matching exercises (linking or matching items from two lists)?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
6. problem-solving questions (finding a solution based on a word problem)?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
7. observational techniques (e.g., checklists, rating scales, etc.) to judge student performance or products?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No

PART D: PRINCIPLES OF TEST ADMINISTRATION

	When administering a test, how often do you:	When administering a test, how important is it to:	Were you ever taught to:
1. inform the students of the scoring criteria (e.g., grammar, handwriting, etc.) to be used in evaluating their responses?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
2. structure the testing situation to discourage cheating (e.g., seating arrangements, etc.)?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
3. maintain a constant presence in the classroom during the course of the test?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

PART II: PRINCIPLES OF TEST MARKING

	When marking tests in your classes, how often do you:	When marking tests in your classes, how important is it to:	Were you ever taught to:
1. develop model answers for the test?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
2. develop a scoring key or marking scheme before scoring the tests?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
3. score student responses one question at a time for all the students when marking essays?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
4. take frequent breaks when marking tests?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
5. re-mark the first few student test scripts to check for consistency in your scoring standards?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
6. assess the degree of inter-rater (marker) consistency if more than one person scores an examination?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

PART I: SUMMARIZING AND INTERPRETING TEST SCORES

	When summarizing and interpreting test scores, how often do you:				When summarizing and interpreting test scores, how important is it to:				Were you ever taught to:	
1. rank-order students according to their test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
2. construct the frequency distribution of the student test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
3. calculate the <u>mean</u> of the test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
4. calculate the <u>median</u> of the test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
5. calculate the <u>range</u> of the test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
6. calculate the <u>variance</u> or <u>standard deviation</u> of the test scores?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
7. calculate the <u>percentiles</u> associated with each test score?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No
8. compare the test scores with established norms?	Never	Rarely	Often	Always	Not at all Important	Slightly Important	Moderately Important	Very Important	Yes	No

PART C: EVALUATING THE QUALITY OF TESTS

	When evaluating the quality of the test that you have developed, how often do you:	When evaluating the quality of the test that you have developed, how important is it to:	Were you ever taught to:
1. calculate the difficulty of each question?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
2. calculate how well an item discriminates between students of high and low abilities?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
3. use the item difficulty and item discrimination statistics to improve the test for future use?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
4. assess the test-retest (stability) reliability of the test?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
5. assess the internal consistency of the test (e.g., split-half reliability)?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

PART II: REPORTING OF STUDENT ASSESSMENT RESULTS

	When reporting the results of student assessments, how often do you:	When reporting the results of student assessments, how important is it to:	Were you ever taught <u>how</u> to:
1. set objective criteria for assigning grades to students?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
2. provide written comments regarding student strengths and weaknesses?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
3. provide information on student performance in such a way that is easily understood by parents?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No
4. use parent conferences to report student grades?	Never Rarely Often Always	Not at all Important Slightly Important Moderately Important Very Important	Yes No

PART K: FACTORS AFFECTING USE OF PRINCIPLES IN TESTS AND MEASUREMENTS

This part of the questionnaire seeks to identify some of the factors which affect teacher use of test and measurement principles in the classroom. Kindly indicate the extent to which each of the factors shown below affect your use of the principles in test construction, test administration, test marking, summarizing and interpreting test scores, evaluating the quality of tests and reporting of student assessment results as identified in the previous sections of the questionnaire. Circle the number corresponding to your choice of the following categories:

1	2	3	4
Does not	Slightly	Moderately	Greatly
affect my use	affects my use	affects my use	affects my use
of the principles	of the principles	of the principles	of the principles

Factors	Test Construction	Test Administration	Test Marking	Summarizing and Interpreting Test scores	Evaluating the Quality of Tests	Reporting of Student Assessment Results
Availability of time	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Availability of resources (e.g., photocopying machines, computers, etc.)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Ministry of Education policies on examinations	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Ministry of Education policies on continuous assessment of students	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Your own school's policies on student assessment	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Your level of comfort with mathematical concepts	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Other factors (please specify):	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

Your Comments

At this time, I would like to provide you with the opportunity to express your ideas, concerns or comments regarding any aspect of this study or the use of tests and measurements in the classroom setting, in general. If you wish to clarify or elaborate on any of your responses, please take the opportunity to do so. Please use the space provided below for this purpose.

PART L: PERSONAL INFORMATION

The following are some questions about you that will be used for statistical purposes only. Your answers will be anonymous. It is important that you complete this section to facilitate the interpretation of the results.

1. What is the highest level that you have completed? (Circle the appropriate letter)

A. Diploma or SI

B. Approved graduate

C. B.Ed.

D. Post-graduate degree in education

E. Other (Please specify) _____

2. What subject or subjects have you been trained to teach? (Please specify)

(a) _____

(b) _____

(c) _____

At this time I would like you to indicate whether you would be willing to participate in an interview in which you could further elaborate on the responses you provided on the questionnaire. Your agreement would be very valuable because the information you provide will allow for a more thorough understanding of the issues relating to both the use of tests and measurements in the classroom and teacher training in measurement. You can be assured that your responses in the interview will remain strictly confidential.

I would like to participate in a short (approximately 30 minute) interview: Yes _____ No _____

I would like to receive a summary of the findings of this study: Yes _____ No _____

If you answered yes to either of the above two questions, please indicate where I may contact you.

Name: _____

Address: _____

Phone: _____

APPENDIX C

Teacher Interview Guide

APPENDIX C

Teacher Interview Guide

1(a). Of the following factors, which would you say affect your use of measurement principles in the assessment of your students?

- Availability of time, availability of resources such as computers and electronic hand calculators, Ministry of Education policies on examinations, Ministry of Education policies on continuous assessment of students, Your own school's policies on student assessment, Your level of comfort with mathematical concepts, number of students in your class, and your workload in school.

1(b). How do these factors affect your use of measurement principles in

(i) Test construction?

(ii) Test administration?

(iii) Test marking?

(iv) Summarizing and interpreting test scores?

(v) The evaluation of test quality?

(vi) The reporting of student assessment results?

2(a) How adequate or inadequate would you say your training in measurement was to your assessment practices in the classroom?

(b) In what assessment areas would you say this training was inadequate?

(c) If an inservice course were to be organized and you were requested to suggest assessment areas to be addressed, which areas would you suggest and why?

(d) Which of these assessment areas would you say typically represent(s) your assessment needs today?

APPENDIX D

The Covering Letter

APPENDIX D

The Covering Letter

Dear

I am secondary school teacher who is completing an M.A. in Education at the University of Ottawa (Canada). I am conducting a survey of trained secondary school teachers in order to gather data on their opinions regarding issues related to the uses of skills in tests and measurements in classroom assessment. It is hoped that the results of this study might be used to provide recommendations to teacher training institutions with a view to making the tests and measurement curriculum more responsive to our needs as classroom teachers. As well, it is hoped that barriers to the use of these skills in the classroom might be identified and recommendations be made to facilitate the use of these skills. The study has been approved by the Faculty of Education and is being sponsored by CIDA through the Kenya-Canada General Training Programme.

You are one of a small number of teachers in the country who are being asked to provide their opinions in these matters. In order for the results to reflect the views of all the teachers, it is important that each questionnaire be completed and returned. All responses will be anonymous unless specified otherwise by you.

I wish to assure you that your participation in this study is strictly voluntary. You may, without prejudice to you, withdraw your consent and discontinue participation in the study at any time. You may also, at any time, request that your answers to the survey or interview questions be withdrawn from the study and returned to you. However, I would hope that you do participate because your opinions and experiences are valuable. Responding to the questionnaire should take approximately 20-30 minutes of your time. I do kindly request you to complete the questionnaire and return it to me using the self-addressed enclosed envelope by 27/7, 1994.

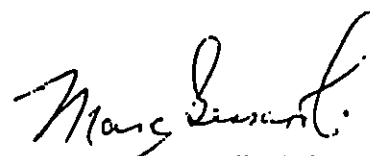
I will be interested to discuss further with you, in an interview, some of the issues which affect the use of measurement skills in the classroom settings. This interview, which will last approximately 30 minutes, is crucial for the usefulness of the study because it is the information based on your experience in the classroom that will determine the credibility of the findings. Again, I would like to assure you that all information you provide will remain strictly confidential. At no time will your responses be attributed to you or your school. If you would like to offer your views in this interview, please kindly indicate your desire to do so in the section at the end of the questionnaire.

I sincerely hope that you will be willing to participate in this study. If you are interested in receiving a summary of the results, I would be pleased to send you a copy.

Thank you in advance for your valuable input.

Sincerely,


Andrew Chirchir


Marc E. Gessaroli, Ph.D.
Thesis Supervisor

APPENDIX E

Teacher Frequency of Use of Measurement Principles by Teacher Curriculum Area

APPENDIX E

Teacher Frequency of Use of Measurement Principles by Teacher Curriculum Area

Table 6

Teacher Frequency of Use of Student Assessment Results by Teacher Curriculum Area (Mathematics, N = 125; History, N = 121)

Uses of Student Assessment Results	Teacher Curriculum Area	Mean	SD	t
1. Assigning grades to students	Mathematics	2.60	0.61	-1.84
	History	2.73	0.47	
2. Changing a student from one curriculum to another	Mathematics	1.46	0.90	-0.49
	History	1.51	0.91	
3. Promoting students from one class to another	Mathematics	1.52	1.26	-0.93
	History	1.67	1.25	
4. Counselling students in their academic work	Mathematics	2.40	0.66	-0.06
	History	2.40	0.67	
5. Aiding students in their career planning	Mathematics	1.95	0.88	1.46
	History	1.78	1.00	
6. Providing students with feedback on their academic progress	Mathematics	2.85	0.38	1.33
	History	2.77	0.54	
7. Diagnosing student needs	Mathematics	2.03	0.75	0.07
	History	2.02	0.78	
8. Planning remedial instruction or work for academically "weak" students	Mathematics	1.82	0.93	-0.45
	History	1.88	0.89	
9. Planning advanced instruction or work for academically "strong" students	Mathematics	1.41	1.04	0.95
	History	1.29	0.91	
10. modifying teaching techniques	Mathematics	2.67	0.65	0.16
	History	2.63	0.64	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 7

Teacher Frequency of Use of the Principles for Test Construction by Teacher Curriculum Area

(Mathematics, N = 125; History, N = 121)

Principles of Test Construction	Teacher Curriculum Area	Mean	SD	t
1. define content areas or topics to be tested	Mathematics	2.70	0.61	0.91
	History	2.63	0.55	
2. develop a test blueprint (table of specifications)	Mathematics	1.56	1.02	1.92
	History	2.31	1.06	
3. set test questions (e.g., multiple choice, essay, etc.) to match the cognitive level (e.g., recall, application, synthesis, etc.) to be tested	Mathematics	2.30	0.82	-1.10
	History	2.41	0.74	
4. use item writing guidelines in the development of test questions	Mathematics	1.57	0.88	-0.77
	History	1.66	1.01	
5. develop items that reflect instructional objectives	Mathematics	2.49	0.69	0.75
	History	2.42	0.69	
6. assess the degree to which the items reflect the content domains to be tested	Mathematics	2.21	0.82	-0.32
	History	2.24	0.71	
7. review the test questions for possible bias, ambiguity, etc.	Mathematics	2.26	0.88	0.51
	History	2.20	0.91	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 8.

Teacher Frequency of Use of Different Types of Test Questions and Other Assessment Methods by Teacher Curriculum Area (Mathematics, N = 125; History, N = 121)

Use of Different Types of Tests and Other Assessment Methods	Teacher Curriculum Area	Mean	SD	t
1. Multiple-choice questions	Mathematics	1.31	0.55	-22.38*
	History	2.10	0.70	
2. True-false questions	Mathematics	0.30	0.48	-6.01*
	History	0.73	0.62	
3. Essay-type questions	Mathematics	0.75	1.10	-16.75*
	History	2.56	0.52	
4. Short-answer questions (e.g., two or three sentence answer)	Mathematics	1.27	0.23	-8.11*
	History	2.30	0.66	
5. Matching exercises (linking or matching items from two lists)	Mathematics	0.80	0.89	-0.90
	History	0.89	0.72	
6. Problem- solving questions (finding a solution based on a word problem)	Mathematics	2.68	0.59	23.51*
	History	0.57	0.80	
7. Observational techniques (e.g., checklists, rating scales, etc.) used to judge student performance or products	Mathematics	1.83	1.01	5.28*
	History	1.15	1.01	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 9

Teacher Frequency of Use of the Principles for Test Administration by Teacher Curriculum Area**(Mathematics, N = 125; History, N = 121)**

Principles of Test Administration	Teacher Curriculum Area	Mean	SD	t
1. Inform the students of the scoring criteria (e.g., grammar, handwriting, etc.) to be used in evaluating their responses	Mathematics	2.03	0.92	0.56
	History	1.97	0.88	
2. Structure the testing situation to discourage cheating (e.g., seating arrangements)	Mathematics	2.40	0.83	1.55
	History	2.23	0.87	
3. Maintain a constant presence in the classroom during the course of the test	Mathematics	2.50	0.69	0.37
	History	2.47	0.72	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always*** $p < .01$.**

Table 10

Teacher Frequency of Use of Test Marking Principles by Teacher Curriculum Area (Mathematics, N = 125; History, N = 121)

Principles of Test Marking	Teacher Curriculum Area	Mean	SD	t
1. Develop model answers for the test	Mathematics History	2.45 2.35	0.79 0.83	0.98
2. Develop a scoring key or marking scheme before scoring the test	Mathematics History	2.88 2.86	0.35 0.39	0.43
3. Score student responses, one question at a time when marking essays	Mathematics History	1.31 1.84	1.11 1.00	-3.94*
4. Take frequent breaks when marking tests	Mathematics History	1.73 1.83	0.85 0.81	-1.01
5. Remark the first few student scripts to check for consistency in your scoring standards	Mathematics History	1.60 1.20	0.93 0.95	3.35*
6. Assess the degree of inter-rater (marker) consistency if more than one person scores an examination	Mathematics History	1.47 1.22	1.01 0.94	2.00

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 11

Teacher Frequency of Use of the Principles for Summarizing and Interpreting Student Test Scores by Teacher Curriculum Area (Mathematics, N = 125; History, N = 121)

Summarizing and Interpreting Test Scores	Teacher Curriculum Area	Mean	SD	t
1. Rank order students according to their test scores	Mathematics	2.43	0.68	-0.18
	History	2.49	0.69	
2. Construct the frequency distribution of the student test scores	Mathematics	1.46	0.87	2.02
	History	1.24	0.88	
3. Calculate the mean of the test scores	Mathematics	1.95	0.80	1.94
	History	1.73	1.00	
4. Calculate the median of the test scores	Mathematics	1.07	0.93	0.26
	History	1.04	0.93	
5. Calculate the range of the test scores	Mathematics	1.76	0.92	2.45
	History	1.46	0.98	
6. Calculate the variance or the standard deviation of the test scores	Mathematics	0.82	0.82	1.00
	History	0.72	0.83	
7. Calculate the percentiles associated with each test score	Mathematics	0.74	0.82	0.00
	History	0.74	0.87	
8. Compare the test scores with established norms	Mathematics	1.59	1.05	0.34
	History	1.55	1.08	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 12

Teacher Frequency of Use of the Principles for Evaluating Test Quality by Teacher Curriculum Area

(Mathematics, N = 125; History, N = 121)

Principles for Evaluating Test Quality	Teacher Curriculum Area	Mean	SD	t
1. Calculate the difficulty of each question	Mathematics	1.63	1.16	1.75
	History	1.38	1.10	
2. Calculate how well an item discriminates between students of high and low abilities	Mathematics	1.50	1.10	1.80
	History	1.26	0.97	
3. Use the item difficulty and item discrimination statistics to improve the test for future use	Mathematics	1.30	1.09	1.93
	History	1.04	0.98	
4. Assess the test-retest (stability) reliability of the test	Mathematics	1.34	1.06	2.12
	History	1.07	0.99	
5. Assess the internal consistency of the test (e.g., split-half reliability) scores	Mathematics	1.10	0.94	1.38
	History	0.93	0.91	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$.

Table 13

Teacher Frequency of Use of the Principles for Reporting Student Assessment Results by Teacher**Curriculum Area (Mathematics: N = 125; History: N = 121)**

Reporting of Student Assessment Results	Teacher Curriculum Area	Mean	SD	t
1. Set objective criteria for assigning grades to students	Mathematics	2.27	0.84	0.92
	History	2.17	0.83	
2. Provide written comments regarding student strengths and weaknesses	Mathematics	2.34	0.77	-0.68
	History	2.41	0.82	
3. Provide information on student performance in such a way that is easily understood by parents	Mathematics	2.32	0.86	-0.25
	History	2.35	0.83	
4. Use parent conferences to report student grades	Mathematics	1.18	0.77	1.75
	History	1.00	0.81	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always*** p < .01**

APPENDIX F

Teacher Frequency of Use of Measurement Principles by Teacher Qualification

APPENDIX F

Teacher Frequency of Use of Measurement Principles by Teacher Qualification

Table 14

Teacher Frequency of Use of Student Assessment Results by Teacher Qualification. (Diploma, N = 113; Graduate, N = 133)

Uses of Student Assessment Results	Teacher Qualification	Mean	SD	t
1. assigning grades to students	Graduate Diploma	2.71 2.61	0.50 0.59	-1.38
2. changing a student from one curriculum to another	Graduate Diploma	1.48 1.49	0.89 0.93	0.76 0.86
3. promoting students from one class to another	Graduate Diploma	1.49 1.72	1.24 1.27	1.42
4. counselling students in their academic work	Graduate Diploma	2.42 2.38	0.65 0.67	-0.48
5. aiding students in their career planning	Graduate Diploma	1.90 1.82	0.97 0.92	-0.66
6. providing students with feedback on their academic progress	Graduate Diploma	2.85 2.76	0.42 0.52	-1.48
7. diagnosing student needs	Graduate Diploma	2.04 2.02	0.78 0.74	-0.20
8. planning remedial instruction or work for academically "weak" students	Graduate Diploma	1.87 1.82	0.89 0.93	-0.42
9. planning advanced instruction or work for academically "strong" students	Graduate Diploma	1.36 1.34	0.98 0.98	-0.20
10. modifying teaching techniques	Graduate Diploma	2.35 2.36	0.64 0.66	0.20

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always ; * $p < .01$

Table 15

Teacher Frequency of Use of Test Construction Principles by Teacher Qualification. (Diploma, N = 113;
Graduate, N = 133)

Principles of Test Construction	Teacher Qualification	Mean	SD	t
1. define content areas or topics to be tested	Graduate Diploma	2.65 2.67	0.57 0.60	0.25
2. develop a test blueprint (table of specifications)	Graduate Diploma	1.38 1.50	0.98 1.11	0.84
3. set test questions (e.g., multiple- choice, essay, etc.) to match the cognitive level (e.g., recall, application, synthesis, etc.) to be tested	Graduate Diploma	2.41 2.29	0.70 0.86	-1.22
4. use item writing guidelines in the development of test questions	Graduate Diploma	1.67 1.55	0.98 0.91	-0.99
5. develop items that reflect instructional objectives	Graduate Diploma	2.51 2.39	0.61 0.77	-1.38
6. assess the degree to which the items reflect the content domains to be tested	Graduate Diploma	2.21 2.24	0.76 0.77	0.29
7. review the test questions for possible bias, ambiguity, etc.	Graduate Diploma	2.19 2.27	0.91 0.88	0.76

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 16

Teacher Frequency of Use of Different Types of Test Questions and Other Assessment Methods by Teacher Qualification (Diploma, N = 113; Graduate, N = 133)

Types of Test Questions and Other Assessment Methods.	Teacher Qualification	Mean	SD	t
1. Multiple -choice questions	Graduate	1.28	1.03	-1.36
	Diploma	1.09	1.15	
2. True-false questions	Graduate	0.57	0.58	-1.71
	Diploma	0.44	0.60	
3. Essay-type questions	Graduate	1.65	1.19	-0.16
	Diploma	1.63	1.30	
4. Short- answer questions (e.g., two or three sentence answer)	Graduate	1.82	1.08	-0.65
	Diploma	1.73	1.16	
5. Matching exercises (linking or matching items from two lists)	Graduate	0.78	0.69	1.34
	Diploma	0.92	0.93	
6. Problem -solving questions (finding a solution based on a word problem)	Graduate	1.53	1.29	1.46
	Diploma	1.77	1.23	
7. Observational techniques (e.g., checklists, rating scales, etc.) used to judge student performance or products	Graduate	1.39	1.09	1.68
	Diploma	1.62	1.04	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 17

Teacher Frequency of Use of Test Administration Principles by Teacher Qualification. (Diploma, N = 113; Graduate, N = 133)

Principles of Test Administration	Teacher Qualification	Mean	SD	t
1. Inform the students of the scoring criteria (e.g., grammar, handwriting, etc.) to be used in evaluating their responses	Graduate Diploma	1.98 2.03	0.87 0.95	0.42
2. Structure the testing situation to discourage cheating (e.g., seating arrangements)	Graduate Diploma	2.23 2.42	0.85 0.85	1.83
3. Maintain a constant presence in the classroom during the course of the test	Graduate Diploma	2.43 2.56	0.73 0.67	1.43

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 18

Teacher Frequency of Use of Test Marking Principles by Teacher Qualification. (Diploma, N = 113;

Graduate, N = 133)

Principles of Test Marking	Teacher Qualification	Mean	SD	t
1. Develop model answers for the test	Graduate Diploma	2.46 2.33	0.77 0.85	-1.27
2. Develop a scoring key or marking scheme before scoring the test	Graduate Diploma	2.85 2.89	0.38 0.36	0.93
3. Score student responses, one question at a time when marking essays	Graduate Diploma	1.53 1.62	1.01 1.18	0.61
4. Take frequent breaks when marking tests	Graduate Diploma	1.82 1.73	0.81 0.86	-0.80
5. Remark the first few student scripts to check for consistency in your scoring standards	Graduate Diploma	1.36 1.45	0.96 0.96	0.74
6. Assess the degree of inter-rater (marker) consistency if more than one person scores an examination	Graduate Diploma	1.29 1.42	0.95 1.02	1.11

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 19

Teacher Frequency of Use of the Principles for Summarizing and Interpreting Student Test Scores by Teacher Qualification (Diploma, N = 113; Graduate, N = 133)

Summarizing and Interpreting Test Scores	Teacher Qualification	Mean	SD	t
1. Rank order students according to their test scores	Graduate	2.52	0.66	-0.98
	Diploma	2.43	0.71	
2. Construct the frequency distribution of the student test scores	Graduate	1.31	0.90	0.88
	Diploma	1.41	0.85	
3. Calculate the mean of the test scores	Graduate	1.85	0.91	-0.15
	Diploma	1.83	0.93	
4. Calculate the median of the test scores	Graduate	1.05	0.91	0.22
	Diploma	1.07	0.94	
5. Calculate the range of the test scores	Graduate	1.69	0.96	-1.38
	Diploma	1.52	0.95	
6. Calculate the variance or the standard deviation of the test scores	Graduate	0.79	0.84	-0.35
	Diploma	0.75	0.81	
7. Calculate the percentiles associated with each test score	Graduate	0.70	0.84	0.74
	Diploma	0.79	0.84	
8. Compare the test scores with established norms	Graduate	1.65	1.06	-1.24
	Diploma	1.48	1.06	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 20

Teacher Frequency of Use of the Principles for Evaluating Test Quality by Teacher Qualification.**(Diploma, N = 113; Graduate, N = 133)**

Principles for Evaluating Test Quality	Teacher Qualification	Mean	SD	t
1. Calculate the difficulty of each question	Graduate	1.39	0.09	1.76
	Diploma	1.65	0.11	
2. Calculate how well an item discriminates between students of high and low abilities	Graduate	1.35	1.00	0.53
	Diploma	1.42	1.09	
3. Use the item difficulty and item discrimination statistics to improve the test for future use	Graduate	1.12	1.01	0.82
	Diploma	1.23	1.08	
4. Assess the test-retest (stability) reliability of the test	Graduate	1.11	0.93	1.68
	Diploma	1.33	1.14	
5. Assess the internal consistency of the test (e.g., split-half reliability) scores	Graduate	0.88	0.87	2.54
	Diploma	1.18	0.97	

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$

Table 21

Teacher Frequency of Use of the Principles for Reporting Student Assessment Results by Teacher**Qualification (Diploma: N = 113; Graduate, N = 133)**

Principles for Reporting Student Assessment Results	Teacher Qualification	Mean	SD	t
1. Set objective criteria for assigning grades to students	Graduate Diploma	2.22 2.23	0.82 0.85	-0.04
2. Provide written comments regarding student strengths and weaknesses	Graduate Diploma	2.43 2.32	0.76 0.84	-1.08
3. Provide information on student performance in such a way that is easily understood by parents	Graduate Diploma	2.36 2.30	0.83 0.87	-0.55
4. Use parent conferences to report student grades	Graduate Diploma	1.14 1.04	0.78 0.81	-0.98

Scale: 0 = Never, 1 = Rarely, 2 = Often, 3 = Always

* $p < .01$