UNIVERSITY OF CAPE COAST

ASSESSMENT PRACTICES OF TEACHER TRAINING COLLEGE TUTORS

IN GHANA

YAW MENSAH ANHWERE

DECEMBER 2009

UNIVERSITY OF CAPE COAST

ASSESSMENT PRACTICES OF TEACHER TRAINING COLLEGE TUTORS

IN GHANA

BY

YAW MENSAH ANHWERE

Thesis submitted to the Department of Educational Foundations of the Faculty of

Education, University of Cape Coast, in partial fulfillment of the requirements for

the Award of Master of Philosophy Degree in Educational Measurement and

Evaluation

DECEMBER 2009

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this University or elsewhere.

Candidate's Signature:…………………………...... Date:…………………………...

Name: Yaw Mensah Anhwere

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature:……………………… Date:………………

Name: Prof. Young Kafui. Abel. Etsey

Co-Supervisor's Signature:……………………………… Date: …………………

Name: Prof. Francis Kodzo. Amedahe

ABSTRACT

The purpose of the study was to investigate whether tutors in the teacher training colleges follow the basic laid down principles of testing practices, especially, test construction, administration and scoring of classroom or teacher-made tests in the teacher training colleges. The study adopted the descriptive design and covered 20 public teacher training colleges in Ghana comprising 230 male and 80 female teacher training college tutors.

The main instrument used for the study was the questionnaire. The instrument was developed and pre-tested by the researcher. The data was analyzed using means and standard deviations, frequency and percentages, and independent samples-t test. For all tests, the level of significance was at 0.05.

The study showed that, teacher training college tutors did not follow the basic principles of testing in the construction of teacher-made or classroom tests, and that they perceived the management of assessment practices in the colleges as extra load to their teaching activities. It is recommended that regular in-service training in testing practices be organized for teacher training college tutors by Teacher Education Division of the Ghana Education Service and the Institute of Education of the University of Cape Coast in collaboration with the experts in Educational Measurement and Evaluation at the University of Cape Coast.

ACKNOWLEDGEMENTS

DEDICATION

This work is dedicated to my parents, Mr. Kofi Anhwere and Madam
Adwoa Twumwaa, and my dear wife, Mrs. Rebecca Adwoa Anhwere.

TABLE OF CONTENTS

Page

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Background to the Study

Tests, according to Anamuah-Mensah and Quagrain (1998), have been regarded as one of the most tangible clues and the most crucial yardstick in determining the attainment of the objectives of any learning experience. Testing practices are, therefore, indispensable tools in the educational enterprise. Educators as well as others concerned with the progress of the leaner need information on the learner for decision making. There is the need for the teacher to describe the nature and extent of the learner's learning in terms of how far the aims and objectives of teaching have been achieved and what is left to be covered. This need calls for assessment of the learner (Tamakloe, Atta & Amedahe, 1996). According to Asamoah-Gyimah (2002), "classroom or teacher-made tests are frequently used as a major evaluating device of students' progress in schools'' (p. 2). Hardly can one envisage or conceptualize an educational system where the student is not put under a classroom or teacher-made tests.

According to Stiggins (1991), "teachers spend much of their instructional time ('a third to a half') in assessment related activities. While this would suggest the need for teachers to be knowledgeable with assessment practices, that is not

the case," (p. 85). Ward (1980) reported that only about half of the teachers in a national survey had received pre- service instruction in tests and measurement. The field of teacher education has been moving forward in re-conceptualizing the goals of teacher education based on new visions of effective teaching, (Reynolds, 1989). As more and more emphasis is being placed on student performance and teacher accountability, measurement and assessment are becoming increasingly important to all educators. It is critical therefore that tutors possess not only comprehensive knowledge of subject matter but also the ability to assess the learning of the subject matter. A concerted effort to use valid measurement and assessment practices would strengthen the effectiveness of teacher education and provide valuable information for programmes in the teacher training colleges. Amedahe (1989) stated that:

> Teachers from stage one/class one, even in the kindergarten to the university in the Ghanaian education system engage in some sort of assessment practices in order to determine whether learning has taken place or not, or sometimes for selection to the next ladder of education. Besides, teachers construct tests to find out problem areas of students in specific areas of topics treated. These and other reasons have necessitated the need for tutors in the teacher training colleges to engage in assessment practices in the colleges (p.1).

According to Andrews and Barnes (1990), the most important service that can be rendered to classroom teachers in a measurement course is teaching the

technique of constructing, administering and scoring classroom or teacher –made tests based on the basic principles in measurement and evaluation. A large amount of the testing done by classroom teachers including tutors in the teacher training colleges must necessarily be the measurement of the immediate results of daily lessons and larger units of work which are not suited to measurement by standardized tests.

The rationale of teacher assessment is linked with the constructive model of learning. In this model, it is important to understand what the student knows and how he/she articulates it in order to develop the knowledge and understanding, (Gipps, 1992b). According to Impara, Divine, Bruce, Liverman and Gay (1991), "teachers have many important strengths in their knowledge of interpreting standardized test scores; these teachers also have several areas of weaknesses in interpreting such information'' (p. 116).

An improved classroom procedure may result from the frequent use of tests for information because such tests will make possible the elimination of factual drill from the class period. Four ways in which classroom or teacher-made tests may be used to improve instruction are diagnosing learning difficulties, individualizing instruction, raising the students' standards, and providing more efficient measures of achievement. Hence, the need for tutors in the teacher training colleges to acquire methods of test construction, test administration and test scoring based on the basic principles in measurement and evaluation. Moreover, it also behoves on tutors in the teacher training colleges to use the appropriate tests in the decision making process of their students.

In Ghana, both the trained and the untrained teacher in the classroom, from the basic level to the university level, construct, administer and score classroom achievement tests irrespective of whether they have had training in measurement and evaluation or not .Decisions mostly taken on the students have far reaching consequences that affect the student. Policy makers have ignored the training and equipping teachers with the skill in test construction, test administration and test scoring for the fact that results have always been achieved out of tests, even in the private institutions. Amedahe (1989) stated that:

> Although teachers may succeed in their teaching to some
> extent without following to the letter the prescribed principles
> in testing their students, more could be achieved if scientific
> principles and practices deemed useful in measurement are
> followed. This is rather crucial in the Ghanaian educational
> system where we lack, to a large extent, the availability of
> standardized achievement and intelligence tests found in
> developed countries like the United States of America (U. S.
> A.) and Britain. What could be termed standardized
> achievement tests (these cannot strictly meet the definition of
> standardized tests) are the tests conducted by the West African
> Examinations Council (WAEC), which are taken nationally
> and internationally in the West African sub-region at the
> terminal points of the educational system. Examples are the

Middle School, the Junior Secondary School and the present

Secondary School Leaving Certificate Examinations (pp. 3 – 4)

Not all teachers in the secondary schools in Ghana have undergone professional training in testing techniques (Amedahe, 1989). It is, therefore, unascertainable whether these teachers and for this study, tutors in the teacher training colleges in Ghana with little or no skills in measurement and evaluation are coping with the task of testing with specific regard to construction, administration and scoring of classroom or teacher-made tests in the colleges, in addition to the management of assessment practices in the colleges.

Statement of the Problem

The practice of assessment in the teacher training colleges is based on the premise that tutors in teacher raining colleges have had a course or training in ":testing". This premise assumes that tutors in teacher training colleges construct, administer and score classroom or teacher-made tests based on the basic principles on testing in measurement and evaluation. This premise is not, in most cases, necessarily true. In addition to this, studies by Amedahe (1989) and Quagrain (1992) revealed that most Ghanaian teachers had limited skills for constructing the objective and essay type tests, which are the most frequently used instruments in our schools. This is because most initial teacher training programmes do not make adequate provision for a course in testing. Amedahe (2000) stated that "teacher –based tests may be made of a number of factors, notably among them are, training in assessment techniques, class size and a

5

particular school's policy in assessment standards with implications on validity and reliability of the assessment results" (p. 112-113).

This study sought to find out the extent to which tutors in the teacher training colleges are managing assessment practices based on the basic principles of testing. The study sought to examine the testing practices of tutors in the teacher training colleges in Ghana in terms of the development and construction, administering and scoring of classroom or teacher-made tests. Stated in a question form, the issue of the research was: To what extent do tutors in the teacher training colleges in Ghana follow the basic scientific principles in the construction, administration and scoring of their tests?

Purpose of the Study

The study sought to investigate the following specific sub-problems of testing in the teacher training colleges in Ghana. Specifically, the study sought to investigate how assessment practices are managed in the teacher training colleges under three main activities involved in the management of assessment practices in the teacher training colleges. These are:

1. The construction of classroom achievement tests by tutors.

2. Test administration and scoring in the teacher raining colleges.

3. Issues related to the management of assessment practices in the teacher training colleges.

The study also looked at major problems tutors in the teacher training colleges encounter in the management of assessment practices, and their perception towards assessment practices in the colleges.

Research Questions

Research questions that were addressed in association with the purposes listed above were:

1.   What basic principles of achievement test construction do tutors in the teacher training colleges follow in constructing test items?

2.   What differences in the constructing of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and those who have taught above three years?

3.   What basic principles of achievement test administration do tutors in the teacher training colleges follow in the administration of test items?

4.   What basic principles of test scoring do tutors in the teacher training colleges follow in the scoring of essay-type test items?

5.   What differences in the administration and scoring of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and those who have taught above three years?

6.   What differences exist in the level of knowledge in assessment practices between tutors who took a course in assessment and those who did not take a course in assessment practices during pre-service training?

7.   What differences exist in the level of knowledge in assessment practices between those who have had any in-service training in assessment practices and those who have not had any in-service training in testing practices?

 8.   Which testing format(s) do tutors in the teacher training colleges use more often in testing their students?

9.   How do tutors in the teacher training colleges put students' continuous assessment scores to use?

Significance of the Study

Gosling (1967), (as cited in Amedahe, 1989) stated "that the role of teachers in testing is too important to be left to chance" (p. 23). Since tutors in the teacher training colleges of Ghana are homogeneous, based on the qualification in teaching a training college, the conclusions and recommendations made could be quite relevant and a guide to all teacher-training colleges.

Moreover, the study could serve as an important reference source for tutors in the teacher training colleges, college administrators, Teacher Education Division (TED) of the Ghana Education Service (GES) and the Institute of Education, (IOE) of the University of Cape Coast, (UCC) in their effort to improve the management of testing in the teacher training colleges with the adequate information about what is actually involved in assessment practices in the teacher training colleges. It is hoped that the study would complement studies already undertaken in this subject matter. Besides, the study could contribute to

the improvement of testing practices, specifically, construction, administration and scoring of teacher-made tests in the teacher training colleges.

Delimitations

Issues concerning assessment practices are so numerous that it would not be feasible for any one study to identify all. There are 42 teacher-training colleges in Ghana with 38 being public while four are private. The study was confined to 20 public teacher-training colleges in the country.

Finally, the study was delimited to issues of test construction, administration and scoring based upon the basic principles of testing. Other areas of study included problems faced by tutors in the management of testing practices in the colleges.

Limitations

The target population for the study was 3,450 tutors in the 38 public teacher-training colleges in Ghana. A sample of 400 (11.6%) tutors was therefore small with regard to the target population. Moreover, the study is descriptive in nature. It is essentially, an initial step to investigate testing practices, specifically construction, administration and scoring of teacher-made tests in teacher training colleges in Ghana. A study, of such nature should cover most tutors in the teacher training colleges unlike 400 tutors who participated in the study. The uncooperative attitude of some of the participants who refused to answer the questionnaire contributed to the response of 310 tutors out of the 400 tutors who

were previously sampled for the study. This limitation of the un- co-operative attitude of some of the participants towards research contributed to obtaining a response rate of 78%.

However, one cannot be very certain as to whether responses of the subjects under study are true or false. One cannot judge the honesty and truthfulness of such responses made by the respondents.

Organization of the Study

The study has been organized into five chapters. The first chapter discusses the Introduction, which highlights the background to the study, the research problem, and the purpose for the study. The research questions have been stated, with the significance and delimitation of the study.

Chapter Two reviews the literature related to the study. The review involves empirical studies and conceptual framework. The third chapter describes the methodology used for the study. This involves the research design, population and sampling procedure, the research instrument, the pre-testing procedure, the procedure for data collection and the data analysis. In Chapter Four, the results are discussed while the final chapter summarizes the study and provides conclusions. Recommendations are given in the last section of the chapter based upon the findings of the study

CHAPTER TWO

REVIEW OF RELATED LITERATURE

This chapter deals with literature reviewed in the following areas relating

to the management of assessment were covered:

1. Historical development of testing;

2. Principles of constructing classroom or teacher-made tests;

3. Principles of test administering classroom or teacher-made tests;

4. Principles in the grading of essay tests and;

    5. General issues related to the management of assessment practices.

Historical Development of Testing

Rudman, Mehrens and Wanous (1983) noted that few studies of teacher

testing practices existed. Amedahe (1989) indicated, that "research work on

classroom testing in Ghana was probably non-existent" (p. 15). According to

Amedahe (1989), "teaching and learning can be likened to two sides of a coin.

Without testing, the teacher would not be in a position to know whether the

objectives set out to be achieved at the outset have been attained" (p. 1).

Tests constitute an integral part of the assessment process in education.

The historical development of the field of measurement in education is closely

related to similar developments in psychology. The Chinese set the pace for what

is known today as "assessment under examination conditions". The practices of assessment had been established by the Chinese long before 1115 B.C. (DuBois, 1966; DuBois, 1970; Ebel, 1972; Lindon & Lindon, 1968; Kuo, 1915).

According to DuBois (1970), the first systematic programme of testing was initiated in China as far back as 2200 B. C. At that time, China had no hereditary ruling class, and initial appointments as well as continuance in employment were based on examinations. The tests covered the examinee's knowledge of civil law, military affairs, agriculture, revenue and geography. Civil servants were tested every three years. The Indian civil service adopted the Chinese method of selecting civil servants. The United States of America also adopted the Chinese method for selecting civil servants in 1883.

Tests were developed for many occupations in the United States. As a means of validating the tests, attempts were made to show that there was a statistical relationship between scores on the test and later performance on the job. Formal testing in schools did not commence until in the 12th century by the Arabs using the Chinese method. Assessment by means of written tests was first used by Jesuits at St. Ignatio (DuBois, 1970).

The development of academic tests was pioneered in Britain, particularly at the University of London. Under its initial stage, testing and awarding degrees were recognized as a legitimate basis for decision making (Ebel, 1972).

The birth of psychology as a science in testing took place in the laboratory of Wilhelm Wundt in Leipzig, Germany. However, Galton (1967) developed psychological testing by studying individual differences. Many of the methods in

psychological testing can be traced to Alfred Binet who focused on the assessment of human abilities. Thorndike and Alfred Binet contributed immensely to the development of the current principles in testing. Thorndike (1903) published the first book in educational measurement. In this book, he indicated that whatever exists at all exists in some amount.

Types of Classroom Teacher-Made Tests

Assessment made by tutors of students' attainment, knowledge and understanding is called variously as teacher-made or classroom made test and school -based assessment (Amedahe, 1989). The rationale of teacher-made tests is linked with the constructivist model of learning. In this model, it is important to understand what the student knows and how he/she articulates it in order to develop his/her knowledge of understanding. In this model, it is learning with understanding which counts and to this end, information about existing ideas and skills is essential. Work in psychology and learning tells us similarly that for effective learning, the task must be matched to the student's current level of understanding (Gipps, 1992a), and either pitched at the level to provide practice or slightly higher in order to extend and develop the student's skills. If the new task is much too easy, the students can become bored, and if much too difficult, the student can become de-motivated (Gipps, 1992b).

Teacher or classroom- made test is essentially both an informal and formal activity. Tutors may pose questions, observe activities, and evaluate students'

work in a planned and systematic or ad hoc way (McCallum, McAlister, Brown & Gipps, 1993). Classroom tests are in most cases teacher –made tests.

Teachers have an obligation to provide their students with the best instruction possible. This implies that they must have some procedure(s) whereby they can reliably and validly evaluate how effectively their students have learnt what has been taught them (Mehrens & Lehmann, 1991). The classroom or teacher-made test is one such tool. Moreover, they are more likely to reflect today's curriculum. This is especially true in subject-matter areas as science and social studies.

Classroom or teacher-made tests can also be tailored to fit a tutor's particular instructional objectives in providing optimal learning to the student (Bejar, 1984). Without classroom or teacher-made tests, those objectives that are unique to a particular topic in the college might never be evaluated. Emphasis is always on the desirability and importance of the tutors being able to construct their test items based on the basic principles in measurement and evaluation.

A survey by Stiggins and Bridgeford (1985) on the uses of various types of tests [teacher-made objectives; standardized objectives; and structured (planned and systematically designed to include pre-specified purposes, exercise, observations and scoring procedures) and spontaneous (arises naturally in the classroom upon which the tutor makes a judgment of the student's level of development] reported that the tests are (a) for assigning grades and evaluating the effectiveness of an instructional treatment, (b) for diagnosis, (c) for remedial teaching, (d) to motivate students to learn to improve in their work, (e) to provide

the basis for guidance in selection and placement in the world of work and (f) for certification.

In a study of teachers in the United States of America, it was observed that in the pre-service education programmes, teachers were only trained to teach but not to assess (test) their students (Gullickson, 1986; Gullickson & Ellwein, 1985; and Marso & Pigge, 1989). Some of the major deficiencies commonly associated with teacher or classroom - made tests are; (a) ambiguous questions, (b) excessive wording, (c) lack of appropriate emphasis and (d) use of inappropriate item formats.

Classroom or teacher-made tests can take various forms. There are several ways in which items have been classified by format – supply and selection types; free answer and structured answer; essay and objective (Ebel & Frisbie, 1986; Gronlund & Linn, 1990; Thorndike & Hagen, 1977). Item types are grouped into two major categories – essay and objectives. Some educators argue that essay tests are more susceptible in scoring than the objective tests. However, classroom teachers exclusively use both since one cannot be used exclusively to measure all learning outcomes. According to Gronlund (1985) and Nitko (1983), essay forms are of two main types; extended response type and the restricted response types. With regard to the objective type tests, the multiple choice, short- answer/fill-in-the blanks, matching and true or false types are the most commonly used by tutors in the teacher training colleges in Ghana (Bartels, 2003).

Essay-Type Tests

According to Amedahe & Etsey (2003), an essay test item is a test that gives freedom to respondents to compose their own responses using their own words. The essay test items consist of relatively few items, but each demands an extended response.

There are two types of essay tests items. These are the restricted response type and the extended response type. The restricted response type limits the respondents to a specified length of the response while the extended response type does not limit respondents in the form and the scope of the answer.

Discussing some of the differences between the essay and the objective tests, Ebel & Frisbie (1986) pointed out the following:

1. Essay tests require an individual to organize and express his/her answers in his/her own words. In the essay or "free response" item, the student is not restricted to a list of responses from which he/she is to select the answer. Objective tests, on the other hand, require that the student either supplies a brief answer or choose the correct answer from among several alternatives.

2. An essay test consists of fewer questions but calls for lengthy answers. Efficiency and reliability are therefore likely to be superior in objective tests.

3. Different skills and processes are involved in taking the tests (Ward, Kline & Fluagher, 1986).

4. The quality of essay test is dependent largely on the skill of the person grading the answer and that of an objective test, on the skill of the test constructor.

5. Essay tests are relatively easy to prepare but more difficult to grade accurately since they are graded by humans (who may be subjective) rather than by impersonal machines.

6. Essay tests afford both the students and the tutors the opportunity to be individualistic. Objective tests afford this freedom of expression (item writing) only to the test maker.

Objective-Type Tests

An objective test requires a respondent to provide a briefly response which is usually not more than a sentence long. The objective test-items normally consist of a large number of items and the responses are scored objectively, to the extent that competent observers can agree on how responses should be scored (Amedahe & Etsey, 2003).

There are two major types of objective tests. These are the selection type and the supply type. The selection type consists of the multiple-choice type, true or false type and matching type. The supply type has variations as completion, fill-in-the blanks and short answer. Objective type test items are most useful when class sizes are very large and when there is limited time to submit the results of the test (Amedahe & Etsey, 2003). Objective tests are more susceptible to

guessing; and the distribution is determined almost completely by the test (Mehrens & Lehmann, 1991).

A true or false test item consists of a statement marked true or false. A respondent is expected to demonstrate his command of the material by indicating whether the given statement is true or false.

One of the limitations in constructing this type of objective test is the probability of getting right answer by guessing is 0.5. It can be used to assess only a limited number of educational objectives, and can be used to evaluate definitions, facts, meaning of expressions, recognition, and interpretation of charts / graphs. One advantage of the true-false format is that it is suitable for classroom short–time evaluation.

One format of the objective test formats that mostly eliminates guessing is the completion type. The short answer is the type of objective test item that is known as the supply, completion, and fill-in-the blank. It consists of a statement or question and the respondent is required to complete it with a short answer usually not more than one line. It is useful for testing knowledge of facts or recall of specific facts (e.g. "knowledge objective" in Bloom's taxonomy of educational objectives). One demerit of the format is that it can be such that there is more than one answer, and that situation makes the scoring of the test subjective.

A matching-type test is a special form of the multiple-choice format. It is not very flexible, though it is useful for testing variables that are compatible. The matching type of objective consists of two columns. The respondent is expected to associate an item in column A with a choice in column B on the basis of a well-

defined relationship. Column A contains the premises and column B the responses or options.

A multiple-choice test item is a type of objective test in which the respondent is given a stem and then is to select from among three or more alternatives (options or responses) the one that best completes the stem. The incorrect options are called foils or distracters.

There are two types of multiple-choice tests. These are the single "best response" type and the "multiple responses" type. The single "best response" type consists of a stem followed by three or more responses and the respondent is to select only one option to complete the stem. The "multiple responses" type consists of a stem followed by several true or false statements or words. The respondent is to select which statement(s) could complete the stem.

The multiple-choice format4can be used to evaluate educational objectives in the cognitive domain of learning (Bloom, 1956). This is widely used in the schools, including the teacher training colleges, and in national or public examinations. It is susceptible to guessing, but the probability of guessing correct responses decreases as the number of options increases. To minimize the probability of guessing, it is recommended that generally, items be made to have about five options.

Thorndike and Hagen (1969) identified six functions served by teacher-made tests. These are (1) motivation; (2) diagnosis and instruction; (3) differentiation and certification; (4) guidance and counseling; (5) curriculum development; and (6) selection and placement.

Construction of Classroom or Teacher-Made Tests

Tests constitute an integral part of the assessment process in education. There is a general consensus among testing experts, (for example, Cunningham, 1986; Ebel, 1972; Mehrens and Lehman, 1991 and Nitko, 1983,) concerning what makes an item a good test item. It needs to be emphasized, however, that this knowledge is necessary but not sufficient to ensure good test construction. The test constructor needs to have commitment to construct a good classroom or teacher–made test because it takes a great deal of time and effort to do so. The specific instructions for constructing test items are divided into two categories; supply and selection items. The supply items include essay and short answer formats while the selection formats include matching, true or false, and multiple-choice item formats (Cunningham, 1986).

Good tests do not just happen. They require adequate and extensive planning so that the instructional objectives, the teaching strategy to be employed, the textural material, and the evaluative procedures are all related. The steps or stages in any test construction have been delineated in most works. Mehrens and Lehmann (1991) identified the following stages and steps as being relevant to the construction of the classroom or teacher-made tests:

1. Specify the course or unit content

2. List the major course or unit objectives

3. Define each objective in terms of students behaviour

4. Discard unrealistic objectives

5. Prepare a table of specifications

6. Decide on the type of item format to be used.

7. Prepare test items that match the instructional objectives

8. Evaluate the degree to which the students have learned the objectives.

9. Revise the objectives and/or teaching material and/or test on the basis of the results.

Contributing to the stages and steps involved in the construction of a classroom or teacher-made test, Chase (1979) indicated five distinct stages. These stages are:

1. Designating the purpose of testing;

2. Choosing between norm-referenced and criterion –referenced approaches based on competencies;

3. Developing a table of specification. Here, two major principles are always involved when the classroom teacher decides to use the table of specification; i.e., preparing a framework of what the essential objectives of the learning instruction are to be assessed and the appropriate test content that reflects the various topics treated to be assessed in making appropriate decisions.

It is very important to prepare the table of specifications before the beginning of instruction since the table would the assist the tutor in organizing his/her teaching – all the resources needed to plan on using in teaching a course (Mehrens, 1984a; Mehrens & Kaminski, 1989). The use of a table of specification will ensure that (a) only those objectives actually involved in the instructional process are assessed, (b) each objective would receive a proportional emphasis on

the test in relation to the emphasis placed on the objective by the tutor, and (c) no important objective or content area would be inadvertently omitted.

The table of specifications aids immensely in the preparation of test items, in the production of a valid and well-balanced test, in the clarification of objectives to both the tutor and the students, and in assisting the tutor to select the most appropriate teaching strategy.

4.   Constructing appropriate test items formats. These are mostly of two main types, i.e., essay or objective test format. Among the objective test item formats are completion test items, multiple-choice test items, true or false test items, and matching test items

5.   Preparing the test for administration

The following factors are also to be considered in the writing of teacher-made test items. The test constructor ought to:

1. carefully define the instructional objectives,

2. prepare a table of specifications, keep it before him/her, and continually refer to it as test items are written,

3. formulate well-defined questions,

4. avoid excess verbiage,

5. have the test items based on information that the students know,

6. try to avoid race and sex bias,

7. write each test item on a separate sheet,

8. prepare more items than actually needed by creating item banks,

9. write and key the test items as soon as possible after the material has been taught,

10. prepare the test items well in advance to permit review and editing;

11. avoid specific determiners,

12. be careful when rewording a faulty test items,

13. insert some novelty into the test,

14. avoid textbook or stereotyped language and not to concentrate on past and examinations questions (Amedahe and Etsey, 2003).

Thorndike and Hagen (1977) have listed some of the common faults of the objective tests. They are:

1. The content tends towards the trivial. Most of the time the test constructor limit themselves to the specifics and the factual.

2. The type of test exercises that the test constructor uses is poorly adapted to appraising the mental processes or content that the test is supposed to measure. Most of the time, the test constructor includes a number of different item types not because each is especially effective for assessing certain objectives, but just to achieve a diversity of format.

3. Most of the questions test constructors write for tests are ambiguous. The wordings usually are either not clear, or the task to be performed is not specified. Students had to guess or use extrasensory perception to determine what the test constructor is asking and wants for an answer.

4. The characteristics of the tests are inappropriate for the purpose for which the test constructor wants to use the test results. As a matter of fact, seldom the test constructor explicitly states the purpose for which he/she wants to use the results.

5. Usually, most test constructors suddenly realize that it is time to give a test and sit down to write items with the idea of producing a certain number of items and never does plan a test to truly reflect their teaching emphasis.

Contributing to the discussion on the basic principles of test construction, Adamolekun (1985) noted that in the writing of any classroom or teacher-made tests for whatever purpose, the classroom teacher should consider the following:

1. Identify the purpose of the test i.e. what the teacher wants to achieve by the test.

2. Select the test item type that will best measure the learning outcome.

3. Obtain a representative sample of student behaviour which the teacher would want to evaluate (e. g. in the cognitive domain; does the teacher want to test knowledge of facts, comprehension, application, analysis, synthesis or evaluation?)

4. Construct test items of the proper level of difficulty.

5. Try to eliminate factors that are extraneous.

6. Make a test which will contribute to improved teaching/ learning practices i.e. what will enable teachers to teach better and students to learn better.

For objective tests, whatever the format, the same general principles guide the tutor in the construction of the test. However, additional guidelines are:

1. Keep reading difficulty level to the appropriate level of the students.

2. Guard against one item providing the clue to another item.

3. Avoid interdependent items i.e. items which are deliberately based on another item.

Administration of Teacher –Made Tests

According to the Standards for Educational and Psychological testing National Council on Measurement in Education (NCME, 1999), the test administrator must establish conditions so that each examinee can do his/her best. The physical conditions should be as comfortable as possible, and the examinees should be as relaxed as possible, even though the evidence is inconclusive regarding the effects of physical and environmental conditions on test performance. Whereas distractions during testing have little effect on the scores of high schools and college students, young children may be affected (Anastasi, 1981; Gronlund & Linn, 1990; Thorndike & Hagen, 1977; Trentham, 1975) Lindquist (1951) further pointed that some types of improper administration of tests cause bias in the results for entire groups, and render inter-group or intra-

group comparisons of little value. Test administration is therefore, equally important as the preparation of the test.

It should be noted that individuals perform better at any endeavour, including test taking, if they approach the experience with positive attitude, and that students cannot perform better when they are excessively tense and nervous. It is known that test anxiety affects optimum performance (Clawson, Firment & Trower, 1981; Culler & Hollohan, 1980; Sarason, Hill, & Zimbrado, 1964). Things that create anxiety are (1) warning students to do their best "because the test is important"; (2) telling students that they must work fast in order to finish on time; (3) threatening dire consequences if they fail; and (4) threatening students with tests if they do not behave. When administering the classroom or teacher-made test, the test constructor should make sure that the students understand the directions of the test. In addition, the teacher should keep the students informed of time remaining.

Careful proctoring should take place so that cheating is eliminated. Many a time test administrators are seen in the examination rooms, but are found to be reading novels, marking scripts and doing other things that contribute to cheating during the administering of the test (Lindquist, 1951).

Frary (1981) and Roberts (1987) pointed out that the best way to detect cheating is to observe students during the examination and not by being preoccupied at one's desk. Moreover, cheating during the administration of a test would occur when a test is poorly planned in which test items do not cover what

26

was taught or focuses on straight memorization of trivial rather than upon reasoning skills.

Discussing the basic principles related to the administration of classroom or teacher-made tests, Tyler (1950) pointed out that certain rules have been useful in actual practice in the administration of any testing programme and for that matter, could be used for administering any classroom or teacher-made test. These are:

1. Select the tests carefully, preferably in co-operation with a college assessment committee.

2. Order the tests well in advance of the date on which they are to be used.

3. Plan in detail for the administration of the tests. Choose examiners and proctors with great care.

4. Avoid overemphasis on the tests.

5. Seat the students in alternate chairs, if possible.

6. Make announcements slowly and clearly in a voice that is loud enough to be heard throughout the room.

7. Have the blanks on the front of the booklets filled out. Be sure to announce the date, how names are to be written, and other items that may need clarification.

8. Time the examination with extreme care, using a watch or a wall clock

9. Move about the room occasionally to see that students are working on the right part of the test, but do not stand gazing over a student's shoulder and do not constantly move nervously from student to student.

10. Stop the test immediately when the time is up and collect the answer booklets.

11. As soon as a certain test has been given, have all the students taking the test turn in their booklets promptly. Alphabetize and check the test papers against the class list.

12. Except in cases of illness, see that all absentees make up the test. This necessity will cause the tutor much trouble and worry, but it is unavoidable, for complete data are essential if the results of the test are to be used successfully in either teaching or guidance.

13. Endeavour to have the tests promptly scored.

14. Have the scores of each student entered on an individual cumulative – record card, and make the card available to both the students. The card may be shown to parents and guardians when necessary (pp. 61 – 63).

In addition to other studies, Amedahe and Etsey (2003) indicated that in ensuring maximum performance of students in the administration of teacher-made tests, teachers should follow certain guidelines. These are as follows:

1. Prepare students for the test. Provide information such as

    a. when the test will be given (date and time),

    b. under what conditions the test will be given, for example number of items, open book or closed book, place of test,

    c. the content of the areas the test will cover, for example, study questions or list of learning targets,

    d. emphasis or weighting of content areas,

e. the kinds of items of the test (objective ─types or essay ─type tests),

f. how the test will be scored and graded, and

g. the importance of the results of the test.

2. Students must be made aware of the rules and regulations covering the conduct of the test. Penalties for malpractice such as cheating should be clearly spelt out.

3. Re-usable tests must be kept secure. Efforts must be made to keep students from taking out copies of the test.

4. Avoid giving tests immediately before or after a long vacation, holidays or other important events where all students are actively involved physically/emotionally.

5. Avoid giving tests when students would normally be doing something pleasant, for example having lunch.

6. The sitting arrangement must allow enough space so that students will not copy each other's work.

7. Adequate ventilation and lighting is expected in the examination room.

8. Provision must be made for extra answer sheets and writing materials.

9. Students should start the test promptly and stop on time.

10. Invigilators are expected to stand at a point where they could view all students. They should once a while move among the pupils to check on malpractices. Such movements should not disturb the students. The teacher/invigilator must be vigilant. Reading novels, newspapers, grading papers are not allowed in the examination room.

11. Threatening behavoiurs should be avoided by the invigilators, for example, speeches like 'if you do not write fast, you will fail'. Students should be made to feel at ease.

12. The testing environment should be free from distractions. Noise should be kept to a very low level if it cannot be eliminated or removed. Interruptions within and outside the classroom should be reduced. It is helpful to hang a "Do not DISTURB – EXAMINATION IN PROGRESS" sign at the door.

Scoring Essay Tests

Because essay tests are less objective, there is a far greater risk that the marks that students receive on these tests, and the decisions made based on the marks, will not be valid or reliable. Therefore, it is very important that technical care be taken in developing, marking, and using essay tests. There are several ways that essay tests can be improved to contribute to better learning and teaching and to be more valid and reliable, including writing detailed specifications, writing prompts that clearly define the student's task, preparing test marking guidelines, and training markers and marking in groups (Capper, 2007).

In essay tests, the measurement of a student's ability does not end with his/her answer, but depends to a large extent on the person who reads his/her answer, as well as on the grading method used. The effectiveness of an essay test depends to a large degree on how well it is graded. Unreliable grading has been

one of the major and most valid criticisms leveled against the use of grading. (Charmey, 1984).

In grading essay tests, the test constructor must (1) use appropriate methods to minimize biases, (2) pay attention only to the significance of the answer, (3) be careful not to let personal idiosyncrasies affect the grading, and (4) apply uniform standards to all papers (Freedman, 1981). Undoubtedly, the uniformity of grading standards is a crucial aspect of grading essay tests. For without uniformity, there is no fair way of comparing students. Moreover, without uniformity, one cannot be certain that the scores represent a valid measure of the students' achievement (Spandel, 1981).

Two commonly used methods that have been developed for grading essay tests are the analytical method and the global (holistic) method (Spandel, 1981; Stiggins, 1991; Frisbie & Griswold, 1989). In the analytical method, (also known as "point-score" method), the student's score is broken down into specific points contained in the answer. Component parts such as (1) effectiveness of expression, (2) logical organization, and (3) "support of statements" are specified and assigned points or values. The major advantages of the analytical method of grading essay tests are as follows:

1. It can yield very reliable scores when used by a conscientious tutor (Ward, 1980).

2. The very process of preparing the detailed answer may frequently bring to the test constructor's attention such errors as faulty wording, extreme difficulty and complexity of the question, and unrealistic time limits.

Hence, if the model answer had been prepared before the test was administered, the question could have been reworded or the time extended.

3. The fine subdivision of the model answer can make it easier to discuss the grade given to the student (Cunnigham, 1986).

Two major limitations of analytical method of scoring are that: (1) it is very laborious and time consuming, and (2) in attempting to identify the elements, undue attention may be given to superficial aspects of the answer (Diederich, 1967).

In holistic (or sometimes known as rating method or global method), the ideal answer is not subdivided into specific points or component parts. It simply serves as a standard. The scorer makes a single overall judgment of the quality of the writing sample. The rater reads the student's work, forms a general impression, and using some standard, assigns a rating to the response. No single factor is given undue weight. Rather, all factors are taken into account in forming the judgment of the response (Stanes, 1992). The crux of this method is to select papers that vary in quality to serve as anchor points to be awarded to the student's response. (Vacc, 1998).

According to Charmey (1984), the validity of holistic scoring has not been convincingly demonstrated. In addition, the process that test constructors follow when marking quick evaluations of a writing sample's quality (i.e. which features of the writing sample contribute most to the holistic score) is unclear (Freedman & Calfee, 1983). The most important limitation of holistic scoring is that it gives no meaningful diagnostic information beyond the comparative ranking it

represents. Even if it is assumed that the score is reliable, the test constructor cannot tell much that he/she might want to know about the student. For example, a low score might represent an inability to control sentence structure, a major spelling incapacity, a total misreading of the question, or misguided attempt to be whimsical or creative. A high score might mean a correct but boring response or a genuinely creative piece of prose. All we have is a single score, where we might wish to have *a* profile (Vacc, 1998).

The second limitation on the value of holistic score emerges from its connection to its particular test group. It cannot represent an absolute value in itself. A second-stage operation is needed to give meaning to the ranking. This means that every time a holistic scoring is completed, those responsible for reporting scores need to make a fresh decision about where cutting levels should be (Kirsch & Guthrie, 1980).

The third important limitation of holistic scores is also common to all test scores: reliabilities are customarily overestimated and inescapable inaccuracy of scores tends to be ignored. All tests yield approximations of ability levels, and even the most highly developed multiple-choice tests (which have almost perfect scoring reliability) would report a wide band of possible error (Tiedt, 1983).

In view of the objectivity in scoring the essay tests, test developers have developed a number of principles to be followed in scoring essay tests to ensure to a greater degree, consistency of the scores. Most test specialists have agreed on these principles as useful in scoring essay tests (Ebel, 1972; Lindquist, 1951; Mehrens & Kaminski, 1989; Spandel, 1981; Tyler, 1950). Based on the literature,

the principles for scoring essay tests can be summarized as cited from the works above. They are as follows:

1. Check scoring rubric (or marking scheme) against actual responses.

2. Be consistent in grading. Graders are human and therefore fallible, and graders should not be influenced by the first few papers they read. (Hales & Tokar, 1975).

3. Randomly shuffle the papers before grading.

4. Grade only one question at a time for all papers to reduce "the hallo effect" (the quality of the response to one question influence the reader's evaluation of quality of the response to subsequent questions).

5. Try to grade all responses to a particular question without interruption.

6. Grade the responses without knowing identity of the students.

7. The mechanics of expression should be judged from the content. [Research has shown that when teachers are told to disregard spellings, punctuations, and grammatical errors in some papers which do not require them, they still assign lower grades to papers containing such errors in such papers (Chase, 1979; and Scanell & Marshall, 1966). Where they are relevant, they should be scored for their own sake].

8. If possible, have two independent readings of the test and use the average as the final score.

9. Provide comments and correct errors.

10. Set realistic objectives (Chase, 1979).

According to Amedahe and Etsey (2003), the following principles should be followed by teachers when scoring essay test items:

1. Prepare a form of scoring guide, either an analytic scoring rubric or holistic scoring rubric.

2. Grade the responses item by item and not script by script. The teacher should also score all responses to each item before going to the next item. This reduces carryover effect. The carryover effect occurs when the mark for a question is influenced by the performance on the previous question.

3. Keep scores of previously graded items out of sight when evaluating the rest of the items.

4. Score the essay test when you are physically sound, mentally alert and in an environment with very little or no distraction.

5. Constantly follow the scoring guide as you score. This reduces the rater drift which is the tendency to either not paying attention to the scoring guide over time or interpreting it differently as time passes. Also avoid being influenced by first few papers read. [These could make the teacher either too harsh or too lenient].

Assessment Practices

Many researchers have carried out studies in assessment practices in the United States of America. According to Schafer and Lissitz (1978), teachers were reported to have not been well prepared to perform their roles as classroom assessors. They indicated that their knowledge base in the fundamentals of testing, for example, terminology, test construction principles, and test use, was very limited (Gullickson & Ellwein, 1985; Infantino, 1976; Takeuchi, 1977). One explanation for this limited knowledge base is that, as of 1991, only four States in the United States had a legal requirement for a course in assessment for teachers (Hills, 1991).

To compound the problem of the inadequacy of knowledge base in testing practices by teachers, basic teacher preparation courses included insufficient assessment content. As early as 1955, a national survey in the United States found that teacher pre-service programmes contained insufficient content in tests and measurement (Noll, 1955). Moreover, Schafer and Lissitz (1987) noted that little change has occured since 1995 in the preparation programmes. This means that few teachers actually took formal courses in tests and measurement, and their exposure to such content in other courses was insufficient, thus they were likely to be deficient in their knowledge and skills in the management of assessment practices. Ward (1980), in a national survey in the United States, indicated that only one-half of the teachers had received pre-service instruction in tests and measurement.

Studies in the United States of America since 1957 have found that the trend of inadequate knowledge of classroom assessment principles and skills has not changed. Boothroyd, McMorris and Pruzek (1992) studied 41 seventh and eighth grade science and mathematics teachers. These teachers represented 25 public and private districts from many geographic regions in New York. They examined a sample of approximately 350 multiple-choice and completion items submitted by the teachers. They found that teachers' knowledge of measurement was not adequate. They attributed this deficiency to inadequate training in measurement given at the pre-service teacher education level. Plake, Impara and Frager (1993) also reported from a United States of America National Council on Measurement in Education (NCME)─sponsored national survey of elementary, middle and secondary level teachers and administrators from 98 school systems that almost 30% of the teachers reported that they had no training at all in tests and measurement. They further observed that teachers who completed a course or in-service training programme in measurement competency test did better than those without such background.

In Ghana, Amedahe (1989) in a study of the assessment practices of secondary school teachers in 18 secondary schools in the Central Region found that teachers lacked the skills and principles of test construction. The limited skills in testing practices and in the management of assessment practices were due to their inability to receive training in assessment practices.

In another study of student assessment procedures in junior secondary schools in 11 districts in the country, it was also found out that teachers did not

have adequate training in the management of assessment practices (Curriculum Research & Development Division [CRDD], 1999). It was reported that 55% of the teachers interviewed felt they were not confident in the testing and measurement practices because they did not have any training in testing and measurement. Moreover, Etsey (2003) in a study of teacher trainees in 24 teacher training colleges, recommended that Teacher Education Division of the Ghana Education Service mandates the teacher education curriculum planners in the country to make courses in the management of assessment practices a priority in the first-two years in the training of teacher trainees in the teacher training colleges in the country. Aidoo-Taylor (1993) pointed out that:

> Testing practices require a great deal of record –
> keeping and frequent measurement of student
> performance. It means more work for teachers.
> Teachers would have to construct tests, and other
> forms of testing instruments, mostly on their own. It
> demands more dedication and professionalism from
> teachers, and the adjustment may be painful for
> some teachers (p. 35).

According to Akyeampong (1997) (as cited in Ghana Muster Research Report, 2000), the system of assessment practices in the teacher training colleges has virtually remained the same throughout teacher training reforms. No statements of standards have ever been developed to guide the teaching, learning and assessment practices in the teacher training colleges, until the introduction of

the current Diploma (Basic Education) programme. As a matter of policy, the Institute of Education of the University of Cape Coast, expects each tutor in the teacher training college to give at least one class assignments/exercises and class test using different test formats, depending on the number of credits of the course of study in a semester. Principals, Vice-Principals, Assessment Officers and Heads of Departments in the colleges are required to assist in management of assessment practices with regards to the effective management of college assessment committee, handling of continuous assessment scores of the students and putting the teacher ─made tests scores into good use.

Summary of Related Literature Review

There has not been much research work in testing practices in Ghana. Studies from literature indicated that tutors had not been well prepared to perform their roles classroom assessors. The knowledge base in the fundamentals of testing, for example terminology, principles and test use, was very limited to test constructors.

The literature has also revealed that the level of knowledge in assessment practices was higher for those who took a course in assessment than those who did not take a course in assessment at the pre-service stage of training.

CHAPTER THREE

METHODOLOGY

The chapter discusses how the study was conducted. It is divided into five sections. The first section deals with research design, and the second; dealt with the population and sample/sampling procedure. The third section covers the research instrument (including pre-testing that was used) while the fourth section deals with data collection procedure. The last section covers how data collected was analyzed.

Research Design

With this survey study, descriptive research design was used to collect data so that inferences could be made about some characteristics, attitudes or behaviour of the population. Osuala (2001) noted that: "descriptive surveys are versatile and practical, especially to the researcher in that they identify present needs" (p. 35). Descriptive research involves collecting data in order to test hypothesis or answer questions concerning the current status of the subjects of the study. It determines and reports the way things are (Gay, 1992).

Frankel and Wallen (1993) stated that "obtaining answers from a large group of people to a set of carefully designed and administered question, lies at the heart of survey research" (p. 17). According to Polit and Hungler (1995),

descriptive study aims at describing, observing and documenting aspects of a situation as it naturally occurs rather than explaining them. A descriptive study provides a more accurate picture of events and seeks to explain people's perception and behaviour on the basis of data gathered at a point. It is appropriate when a researcher attempts to describe some aspect of a population by selecting unbiased samples who are asked to complete questionnaires, (Frankel & Wallen, 1993). Frankel and Wallen (1993) continued by noting that the big advantage of the descriptive study has the potential to provide a lot of information obtained from quite a large sample of individuals.

Seifert and Hoffnung (1994) and Frankel and Wallen (1993) noted further that there is the difficulty of ensuring that the questions to be answered using the descriptive design are clear and not misleading because survey results can vary significantly depending on the exact wording of questions. It may also produce untrustworthy results because they delve into private matters people may not be completely truthful about. They further maintained that questionnaires require respondents who can articulate their thoughts well and sometimes even put such thoughts in writing. Getting a sufficient number of questionnaires completed and returned so that meaningful analysis is made is another weakness of the descriptive study.

The descriptive design was chosen for the study because judging from the main study, it was the most appropriate design which could lead the researcher to draw meaningful conclusions from the study. Moreover, the descriptive design was used because according to Frankel and Wallen (2000), the big advantage of

the design is the potential to provide a lot of information obtained from quite a large sample of individuals. Hence, other research designs such as correlation method or evaluation method were not applicable to the study.

Population

The target population for the study was made up of 3,450 tutors of the 38 public teacher training colleges in the country. The accessible population was made up of 1,820 tutors in selected 20 public teacher training colleges in the country.

Sample and Sampling Procedure

There are 38 public teacher training colleges in Ghana which have been grouped into six zones (See Appendix A). There are 3,450 tutors in the 38 teacher training colleges. The first zone has five colleges, the second zone has seven colleges, the third zone has 12 colleges, and the fourth zone has seven colleges. The remaining two zones; five and six have three and four colleges respectively..

Four hundred tutors were selected from the 20 public teacher training colleges for the study. Twenty teacher training colleges were selected for the study. (See Appendix B). Ninety selected tutors who had to attend a teacher Training College in-service programme that had been organized by the Teacher Education Division of Ghana Education Service could not participate in the study. The researcher had to work with, three hundred and ten tutors in the 20 teacher

training colleges for the study. These were 230 males and 80 females respectively. The selection was done through a two-stage simple random sampling procedure.

The first stage involved randomly selecting at least two teacher training colleges from each zone. The researcher in his attempt to obtain a representative sample for the study decided to select at least two teacher training colleges from each zone. This sampling procedure was deemed appropriate for the study because each zone had at least two teacher training colleges. The researcher shook the receptacle and randomly picked up to determine the college with the code selected. That was done with replacement, in order to maintain the same probability for the colleges in each zone.

The same procedure was adopted at the second stage for the selection of the tutors with each teacher training college selected with at least ten tutors irrespective of the subject taught in the colleges. By the foregoing procedure, representative sample for both colleges and tutors was obtained. The selected colleges were serially coded.

In each college, the names of the tutors were then written on pieces of paper and were put in an urn. Simple random sampling technique was then used to select the 10 tutors in each teacher training college. The slips of paper were picked one after the other without looking into the pool. Once a name of a tutor was selected, it was recorded as a sample with the urn shaken while the chosen piece of paper was put back into the urn and reshuffled. The process was repeated in each college till 400 tutors were obtained for the study. Table 1 shows distribution of sample by college, by zone and by gender.

Table 1

Frequency Distribution by College, Zone and Gender

College	Zone	Male	Female
1	1	10	0
2	1	8	7
3	1	10	10
4	2	11	3
5	2	8	12
6	2	13	7
7	2	15	3
8	3	9	1
9	3	9	1
10	3	12	1
11	3	17	4
12	3	17	4
13	3	17	5
14	3	11	3
15	4	9	1
16	4	9	2
17	5	16	4
18	5	11	4
19	6	8	2
20	6	16	4

The Research Instrument

The research instrument used for the study was the questionnaire (see Appendix C). A four-section questionnaire was developed and was made up of 81 items. The items were mainly closed-ended ones with only two being open-ended ones. A four —section questionnaire was developed.

Section A, which contained five items requested information on the background of the respondents. The items in this section included gender, number of years taught in the college, highest academic qualification, subjects taught by the tutors and status in the college.

Section B of the questionnaire dealt with factors involving construction of achievement tests based on the basic principles in testing. This section contained knowledge of respondents on course in testing, attendance of in-service training, and the frequency of assessing students using the objective, essay or both formats. Item 8 in this section sought for information on level of assessment practices in test construction while the last part in this section sought for information on the level of knowledge on concepts and terms related to educational testing and practices. In writing the items in this section, the researcher considered the basic principles and the stages in test construction.

Section C covered the administration and scoring of classroom tests in the colleges. Information mainly sought for in this section was the conditions under which teacher training college tutors administer and score their tests in the teacher training colleges. Issues discussed in this section included methods used in scoring essay tests, factors that influence the tutors when marking essay tests, and

practices often used by the tutors when administering and scoring tests in the colleges.

The last section (i.e. Section D) was concerned with some general issues connected with the management of assessment practices in the teacher training colleges. Amongst them were, how often tutors plan the schedule of assessing students in the colleges, number of tests administered to students, and how students' assessment scores are put to use. Moreover, item 19 covered how effective the College Assessment Committees were in their work. The last part in this section covered issues on major problems encountered by tutors in the teacher training colleges in the management of assessment practices.

The questionnaire described above was developed by the researcher after reviewing the related literature on the principles on the teacher-made tests construction, administration and scoring. To test the validity of the items in the questionnaire, the researcher conducted a pre-test in two teacher training colleges; Akrokerri and OLA Teacher Training Colleges, Akrokerri and Cape Coast respectively. Based upon the pre-testing, the questionnaire was revised and edited before the final write up. Ten items out of the 91 items were revised to make them more specific and effective in eliciting the needed responses. It was after the pre-testing that the main pre-tested questionnaire was administered to the sample of the study.

In the development of the questionnaire, literature on assessment practices in the knowledge of basic principles tutors are to follow in the construction, administration of classroom/teacher-made tests and in the scoring of essay-type

tests The researcher's supervisors, in the Department of Educational Foundations of the University of Cape Coast (UCC), experts in the area of measurement and evaluation read through the questionnaire to check the validity – the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on the scores generated from the instrument. Specifically, they read through the questionnaire to check the construct validity of the instrument

On the basis of their comments, the statements were reviewed and a second list comprising multiple-choice items as well as likert-scale statements were produced and developed into the questionnaire. The questionnaire was administered to 30 tutors in two public teacher training colleges, Akrokerri Teacher Training College, Akrokerri and OLA Teacher Training College, Cape Coast respectively as a pilot study. The responses to the items were analysed and the final instrument made. The instrument had a Cronbach's alpha of 0.782 as the estimate of its reliability (See Appendix D), and this was considered appropriate since any co-efficient alpha above 0.70 is considered appropriate (Pavet, Diener, Colvin & Sandvick, 1991).

Pre-Testing

The questionnaire was pre-tested in two teacher training colleges namely, Akrokerri Teacher Training, Akrokerri (AKI) and Presbyterian Womens' Teacher Training Colleges, Aburi (PWTC) in zone 3 and zone 2 respectively on March 8

and 12, 2007 respectively by the researcher. The sample for the pre-testing was 35 tutors, comprising 20 tutors at AKI and 15 tutors at PWTC respectively.

Akrokerri Teacher Training College was chosen to cater for colleges with both sexes while Presbyterian Women's Teacher Training College was chosen to determine testing practices as pertained in the female teacher training colleges. Moreover, the teacher training colleges chosen for the pre-testing have similar attributes to that of the other teacher training colleges in Ghana. Besides, the tutors of the pre-testing colleges, which formed the accessible population of the study, had similar qualifications and experiences.

The questionnaire was personally administered to the 35 tutors in the two colleges. The results were analysed to determine the content validity of the instrument. Finally, items of the original questionnaire were reduced from 98 to 81 for the main study.

The pre-testing was necessary because it enhanced the content validity and reliability of the instrument, and to improve questions, format and scales after careful analysis of the items based on the comments passed by respondents concerning the weaknesses, clarity and ambiguity on all aspects of the questionnaire. For example, it was found out after the pre-testing, that items 2 and 10 of the questionnaire were not consistent with the objective of the study, and were revised.

In addition, the pre-testing provided the opportunity in assessing the appropriateness and practicality of the data collection instrument. Besides, it tested the adequacy of the procedures that were used for the study. Moreover, the

pre-testing was important because it enhanced the content validity and reliability of the instrument, and also improved items that were ambiguously constructed. On the whole, the pre-testing helped to fine-tune the instrument.

Data Collection Procedure

The questionnaire was administered to 310 tutors in the 20 sampled teacher-training colleges for the study by the researcher with the assistance of the Vice-Principals in the colleges. Permission was sought from the Principals in the colleges, who allowed the researcher to seek the assistance of the Vice Principals in administering the questionnaire, after an introductory letter had been delivered to the Principals for co-operation and assistance.

In each college, the researcher explained the purpose of the study to the Principal, Vice Principal and the sampled tutors, and assured them of anonymity and confidentiality of their participation in the study. Questionnaires were then administered to the tutors with the assistance of the Vice-Principals.

There was time for questions, during which respondents had the opportunity of asking questions that were not clear to them before responding to the questionnaire. This was because it helps to erase respondents' biases and prejudices (Trochim, 2000). This ensured good contact with the tutors to further explain the purpose of the study so that the researcher won the commitment of the tutors towards responding to items on the questionnaire and submitting them in good time.

The administration of the questionnaire to the tutors in Zones 1 and 2 was done in early April, 2007, and that of Zones 3 and 4 was in later part of April, 2007. The administration of questionnaire to participants in Zones 5 and 6 was in mid-May, 2007 when the researcher attended a two-week workshop in Tamale and Wa respectively. The questionnaire was administered to the tutors at one sitting by the researcher with the assistance of the Vice Principals. Tutors spent 15 minutes in responding to items on the questionnaire. Data collection was done on April 10 and 28, 2007 by the researcher in the colleges in Zones 1, 2, 3 and 4 respectively, while data collection of the completed questionnaire in Zones 5 and 6 was in May 8 and 12, 2007 respectively.

For the sake of anonymity and confidentiality of the results, participants were instructed not to write either their names or name of their college on the questionnaire. Four hundred tutors were given the questionnaire to respond to. However, 90 of them had to withdraw from the study because they had to attend a workshop organized by the Teacher Education Division of the Ghana Education Service for selected tutors in the teacher training colleges at Ajumako in the Central Region of Ghana. The researcher collected the completed questionnaire from the tutors after they had responded to them, while in some situations like in colleges in zones five and six, the Vice-Principals assisted in the collection of the completed questionnaire from the tutors. Three hundred and ten participants returned their questionnaires. They were 230 males and 80 females giving a 78% response rate.

Data Analysis

The data collected in this study was checked, edited, coded and statistically analyzed with both descriptive and inferential statistics based on the research questions and the literature reviewed for the study. Apart from the last two items of the questionnaire, all the items were of the close-ended types with options provided for the respondents to choose from. Items 1, 6 and 19a were of dichotomous type while items 2, 3, 4, 5, 16, 17, and 18 were of multiple-choice types. Items 8, 9, 12, 13, and 15 were measured on a five-point scale indicating "never" (scored 1) to always (scored 5). Item 12 was also on a five –point scale and weighted 1 to 5, "not at all "(scored 1) to "very large extent" (scored 5).

Item 13 was measured and coded as follows: "never" (scored 1) to "always" (scored 5). Item 15 was coded as "once" (scored 1} to "less than 4" (scored 5). A four-response Likert format was used for items 10 and 19b indicating "I never learnt about it" (scored 1) to "I remember exactly what it means" (scored 4) and "not effective" (scored 1) to "very effective" (scored 4) Item 20 was measured indicating a four response Likert scale of statements in agreement and disagreement with "strongly disagree" (scored 1) to "strongly agree" (scored 4). For all inferences to be made, the level of significance was at 0.05.

Research Question One

What basic principles of achievement test construction do tutors in the teacher training colleges use in constructing test items?

The responses of the items to this research question were analyzed using means and standard deviations of the response on principles of achievement test construction by tutors in the teacher training colleges.

Research Question Two

What differences in the constructing of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and those who have taught above three years?

To find out whether there was any statistically significant difference between tutors who have taught between 1-3 years and those who have taught above three years, responses to items were analyzed by conducting an independent samples t-test for equality of means at 0.5 level of significance.

Research Question Three

What basic principles of achievement test administration do tutors in the teacher training college follow in the administration of test items?

The responses of the items to this research question were analyzed using means and standard deviations of the principles of achievement test administration by tutors in the teacher training colleges.

Research Question Four

What basic principles of test scoring do tutors in the teacher training colleges follow in the scoring of essay-type test items?

The responses of the items to this research question were analyzed using means and standard deviations of the responses on the principles of test scoring of essay-test type items by tutors in the teacher training colleges.

Research Question Five

What differences in the administration of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and above three years?

An independent samples t-test for equality of means was conducted to compare the group mean scores for teacher training college tutors who have taught between 1 − 3 years and above three years in the teacher training colleges with respect to the administration and scoring of test items.

Research Question Six

What differences exist in the level of knowledge in assessment practices between tutors who took a course in assessment and those who did not take a course in assessment practices during the pre-service training?

To find out whether there was any statistically significant difference between tutors who have taught between 1-3 years and those who have taught above three years, responses to items to the research question were analyzed by conducting an independent samples t-test for equality of means at 0.5 level of significance.

Research Question Seven

What differences exist in the level of knowledge in assessment practices between tutors who have had any in-service training in assessment practices and those who have not had any in-service training in testing practices?

This question was answered by analyzing the data using an independent samples t- test of test for equality means at 0.5 level of significance.

Research Question Eight

Which testing format(s) do tutors in the teacher training colleges use more often in testing the students?

This research question was answered by analyzing the data using a frequency and percentage distribution table.

Research Question Nine

How do tutors in the teacher training colleges put students' continuous assessment scores to use?

The data concerning this research question was analyzed using frequency and percentage distribution. A frequency distribution table showed how tutors put the continuous assessment scores to use in teacher training colleges.

CHAPTER FOUR

RESULTS AND DISCUSSION

In this chapter, the data collected from the respondents are analyzed using frequencies, percentages, means and standard deviations, independent samples t-test and the z-test for a proportion (binomial distribution) and the results presented with discussions.

Characteristics of the Sample

Gender Distribution of Respondents

Out of a multi-stage sampled size of 310 tutors, 230 (74%) tutors were male while 80 (26%) tutors were female.

Number of Years of Teaching in College of Respondents

Question 3 of the questionnaire sought to find out the number of years the sampled tutors for the study had taught in the colleges. Table 2 shows the number of years the tutors had taught in the various colleges.

Table 2

Number of Years in the College

Year (s)	Frequency	%
1 – 3	105	34
4 – 7	87	28
8 – 11	69	22
Above 12	49	16
Total	**310**	**100**

It can be observed from Table 2 that majority 105 (34%) of the tutors had spent between 1- 3 years in their various colleges while only 49 (16%) had taught in their colleges for more than 12 years.

Highest Academic Education Level of Respondents

Education is an important factor influencing people's perception, behaviour and acceptance of practices that may be of benefit to them. It was therefore considered necessary to find out the highest academic qualification attained by tutors in the teacher training colleges.

Item 3 of the questionnaire was on finding the highest academic qualification of the participants in the study. The results are in Table 4 below.

Table 3

Highest Academic Qualification of Respondents

Qualification	Frequency	%
Bachelors	289	87
Bachelor with PGDE	18	6
Master of Education	11	4
Master of Philosophy	10	3
Total	**310**	**100**

From Table 3, 289 (87%) tutors have a Bachelor's degree as their highest academic qualification while 21 (7%) tutors have a masters' degree.

Status of Respondents

The distribution of the status of the selected tutors is shown in Table 4.

Table 4

Status of Tutors in College

Status	Frequency	%
Vice – Principal	20	7
Head of Department	50	16
Assessment Officer	20	7
Tutor	220	70
Total	**310**	**100**

From Table 4, most of the respondents, 220 (70 %,) were tutors while each college had an assessment officer and a Vice -Principal.

Research Question One

What basic principles of achievement test construction do tutors in the teacher training colleges follow in constructing test items?

This question sought to find out three major issues involved in the construction of classroom or teacher-made tests in the teacher training colleges. These are test formats used in the colleges, actual test construction practices and current level of knowledge with certain concepts in test construction as expected by tutors. Tables 5 to 7 show the mean responses and the standard deviations of the sampled tutors on the three major issues in test construction.

Item 8 of the questionnaire was used to determine how frequent respondents used the major test formats in test construction. The responses were on a five-point Likert format with categories from "never" (scored 1) to "very often" (scored 5). However, three cut-off points ranging from "never" (scored within a mean of 2.0 – 2.5), "often" (scored within a mean score of 2.5 ─3.0) to "very often" (scored above a mean of above 3.0), were used in discussing the results as shown in Table 5.

Table 5

Means and Standard Deviations on the Use of Test Formats

Test format	N	Mean	Standard deviation
Essay	310	3.68	1.19
Short answer/fill in the blank	310	3.60	1.15
Multiple choice	310	2.98	1.29
True or false	310	2.45	1.18
Matching	310	2.28	1.16

From Table 5 responses indicate that tutors used the two major types of test items; (i.e. essay and objective test formats) as assessment methods in the colleges. The result shows that with ($\underline{M} = 3.68$, $\underline{SD} = 1.19$, and $\underline{M} = 3.60$, $\underline{SD} = 1.15$), the essay and the fill-in –the blank/short answer types of test formats could be said to be very often used by the tutors while the other two having ($\underline{M}$ (2.45, $\underline{SD} = 1.18$) for the true or false test format being often used while the matching test format having ($\underline{M} = 2.28$, $\underline{SD} = 1.16$) never or less frequently used by the tutors in the colleges. This result supports previous study that classroom or teacher-made tests are mainly essay, short answer and multiple-choice types (Adamolekun, 1985; Amedahe,1989; Bartels, 2003; Gronlund, 1985). Item 9 of the questionnaire sought to find out the frequency of the use of principles in the construction of classroom or teacher-made tests. The item contained statements to be responded by using a five-point Likert format with categories from "never"

(scored 1) to "always" (scored 5). However, four cut-off points ranging from "not often" (scored within a mean of 1.0 – 2.0), "often" (scored within a mean score of 2.0 -3.0) to "very often" (scored within a mean of 3.0 – 4.0) and "always" (scored above a mean of 4.0) have been used in discussing the results as shown in Table 6.

Table 6

Means and Standard Deviations on Actual Practices of Test Construction

Test Practice	N	Mean	Standard Deviation
I write directions/instructions for the test clearly.	310	4.27	1.08
I evaluate the test as whole before I submit it for typing.	310	3.96	1.15
I read over the test items after they have been set aside for a few days	310	3.91	1.06
I prepare a marking scheme before students take the test.	310	3.91	1.13
I match instructional objectives with test items.	310	3.89	1.12
I consider the purpose of the test before developing test items.	310	3.88	1.17
I write test items based on information that students know.	310	3.55	1.16
I prepare more test items than needed before I review and select in the examination.	310	3.50	1.18
I refer to the principles of test construction when I write test items.	310	3.41	1.29
I allow students to select a specific number of test items from a given set items in an essay test.	310	3.27	1.38
I write individual test items at least two weeks before date of testing.	310	3.25	1.06
I arrange test items in order of increasing difficulties.	310	3.22	1.25
1 review test items by letting another tutor in the subject area read over them.	310	3.01	1.32

From Table 6, based on the cut-off points, responses indicate that tutors reported that the practice "I write directions/instructions for the test clearly" (M =4.27, SD = 1.08) was always done. Practices that were very often adopted by tutors in test constructions include the following: "I evaluate the test as whole, before I submit it for typing" (M = 3.96, SD = 1.15); "I read over the test items after they have been set aside for a few days" (M =3.91, SD = 1.15); "I prepare marking scheme before students take the test" (M =3.91, SD = 1.13); "I match instructional objectives with test items" (M =3.89, SD =1.12); "I consider the purpose of the test before developing test items" (M =3.88, SD =1.17). These are the major testing practices that were often done by the tutors. Practices as " I copy questions from past teacher training college examinations questions set by the Institute of Education of the University of Cape Coast" (M = 2.23, SD = .94); "I copy test questions from textbooks" (M =2.08, SD = .95); and "The institution conducts in-service training in test construction for tutors" (M = 2.03, SD = 1.17) as testing practices were not often performed by tutors in the colleges.

Research Question Two

What differences in the constructing of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and above 3 years?

This research question was used to find out whether there was any significant difference between tutors who have taught between 1-3 years and above three years in the teacher training colleges when constructing test items.

Using the number of years as the independent variable and the practices of test construction as dependent variable, an independent samples t-test of equality of means was conducted to determine whether there existed any significant difference between those who have been teaching between 1 – 3 years and those above three years in the teacher training colleges. The result is shown in Table 8.

Table 7

Test for Equality of Means on Test Construction of Tutors Teaching between 1-3 Years and above Three Years

Group	N	Mean	SD	t	df[a]	p-value
1 – 3 Years	105	88. 53	11.81			
				-1.49	308	0.14
Above 3 Years	205	90.76	12.78			

Degrees of freedom reduced because Levene's test shows violation of homogeneity of variances assumption.

The result indicated that based on the sample, the independent samples t-test for equality of means is not statistically significant (t(308) = -1.486, $p >$ 0.05). This shows that there was not any significant difference between tutors who taught between 1 -3 years and above three years.

Research Question Three

What basic principles of achievement test administration do tutors in the teacher training colleges follow in the administration of test items?

Items 1 to 7 sought to find out what basic what principles of test administration tutors in the teacher training colleges follow in the administration of essay-type test items?

This section of the research question was answered using the actual practices involved in test administration This section was scored using a – five point Likert scale with categories "never" (scored one) to "always" (scored five). A mean score, ranging between 1.0 and 2.0, scored" never"; that of 2.1 – 3.4 (scored, "often"), and a mean score of above 4.0 (scored, very often).

Table 8 shows the results of the practices, based on the physical or psychological conditions when administering classroom or teacher-made tests in the colleges.

Table 8

Means and Standard Deviations of Actual Test Administration Practices in the Colleges

Test practice	N	Mean	Standard deviation
After examination questions are typed, I proofread for mistakes before students take the test.	310	4.58	0.85
I ensure good sitting arrangement to prevent students from copying from each other.	310	4.39	0.92
During examinations, I prepare for and expect emergencies.	310	3.18	1.23
I inform students in advance about contents/topics that the test/examinations cover.	310	3.01	1.39
I give hints to students when they ask about individual test items during examinations.	310	1.47	1.39
I read novels, newspapers or mark questions when I invigilate tests/examinations.	310	1.29	0.86
During examinations, I tell students that if they do not write fast, they will fail.	310	1.15	0.96

The physical conditions that tutors indicated were very often practiced were: "I ensure good sitting arrangements to prevent students from copying from each other" (M =4.39, SD = 0.92), and "After examination questions are typed, I proofread for mistakes before students take the test" (=M =4.58, SD = 0.85). and "During examinations, I prepare for and expect emergencies" (M = 3.18, SD = 1.22). The results indicated by tutors, supports the statement that:

> The seating arrangement in the testing room should be carefully planned both to make the conditions as comfortable as possible for the students and to reduce opportunities for copying. Where the size of the room permits such an arrangement, the use of alternate seats and even alternate rows is preferable (Lindquist, 1951, p. 351).

The findings suggest that tutors ensured good physical conditions in administering their classroom or teacher-made tests. Ensuring good physical conditions is very crucial and effective supervision should not be ignored. As pointed by Ebel and Frisbie (1986), "the general goal of good test administration is to present and maintain the conditions that would permit all examinees to demonstrate their true level of achievement" (p. 210). Judging from the result, one may conclude that tutors very often ensured good physical conditions when administering classroom or teacher –made tests in the colleges.

It might be due to the type of buildings in the colleges, which might have contributed to the good physical conditions in the colleges. However, the efforts of tutors in organizing the classroom should not be overlooked.

The following four statements were used to find out whether tutors adopt the psychological conditions which specialists in measurement considered not useful when administering classroom or teacher-made tests in the colleges. Four of such conditions were ((1) threatening students when writing examinations/tests, (3) aiding students in examination rooms and (4) reading novels and performing other activities in the examination rooms when administering an examinations/ tests.

From Table 8, the analysis of the responses of tutors showed that the practice that was often carried out was, "I inform students in advance about content/topics that the test/examinations cover" (M =3.01, SD = 1.39). The other three psychological conditions that were less practiced during test administration in the colleges were, (1) "During examinations, I tell students that if they do not write fast, they will fail" (M =1.51, SD = 0.96), (2) "I give hints to students when they ask about individual test items during examinations" (M = 1.47, SD = 0.84) and (3) "I read novels, newspapers or mark questions when I invigilate tests/examinations" (M = 1.29, SD = 0.86).

Judging from the results, it might be concluded that psychological conditions were quite favourable for good test taking in the colleges.

Research Question Four

What basic principles of test scoring do tutors in the teacher training colleges follow in the scoring of essay-type test items?

Items 8 to 19 of Question 12 of the questionnaire asked tutors to indicate actual practices observed by them when scoring essay-type classroom or teacher–made tests in the colleges. The responses were assessed on a three–Likert scale categories using an average mean as, (1) "never" (scored between mean of 1.2 and 2.0), (2) "often" (scored between mean range of 2.1 and 3.0) and (3) "very often" (scored above a mean value of 3.5). The results are presented in Table 9

Table 9

Means and Standard Deviations on Actual Practices of Test Construction

Test Practice	N	Mean	Standard Deviation
I score essay test questions by question.	310	4.05	1.11
I promptly submit test results to the head of department.	310	3.78	1.20
I promptly record test results.	310	3.78	1.26
I provide comments on essay tests for students to facilitate learning.	310	3.64	1.31
I reshuffle essay scripts before I score the scripts.	310	3.63	1.33
I periodically rescore previously scored paper to check consistency in scoring.	310	3.57	1.24
I keep previously marked scored paper out of sight when scoring the rest of the papers	310	3.26	1.43
In an essay test, I score all questions for a student before I score another student's paper.	310	3.08	1.57
I prepare marking scheme for essay tests after students take the test.	310	-	-
I give a separate mark for mechanics of writing such as correct grammar, flow of expression, etc when scoring essays.	310	2.88	1.39
The first few essays I score influence the rest of the scores I give.	310	1.78	1.06
I score answer scripts with the names of the students known to me.	310	1.71	1.20

From Table 9 above, practices used in test scoring that showed that they were very often used were as follows; "I score essay test questions by question" (M = 4.05, SD =1.11), "I promptly submit test results to the head of department" (M =3.78, SD = 1.20), " I promptly record test results" (M =3.78, SD = 1.26), "I provide comments on essay tests for students to facilitate learning" (M =3.64, SD =1,31), " I reshuffle essay scripts before I score the next question" (M =3.63, SD = 1.32), and "I periodically rescore previously scored paper to check consistency in scoring" (M = 3.57, SD = 1.24). Practices that were often used in scoring essay– type tests, were ; "I keep previously scored papers out of sight when scoring the rest of the papers" (M =3.26, SD =1.43), "In essay tests, I score all questions for a student before I score another student's paper" (M = 3.08, SD = 1.57), "I prepare marking schemes for essay tests after students have taken the test" (M = 2.91, SD =1.59) and "I give separate marks for the mechanics of writing such as correct grammar, flow of expression, etc. when scoring essay test items" (M = 2.88, SD = 1.38).

Item 11 of the questionnaire sought to find out what factors influenced tutors in the teacher training colleges when scoring the essay-type tests. The result is presented in Table 10

Table10

Means and Standard Deviations on Factors Affecting Scoring of Essay Test

Items in the Colleges

Factor	N	Mean	Standard deviation
Grammar and language expression	310	3.79	1.15.
Student's previous knowledge	310	3.42	1.37
Handwriting	310	2.81	1.24.
Number of scripts	310	2.66	1.38
Length of student's essay	310	2.49	1.37
Gender of student	310	1.54	1.04

Research has shown that when teachers are told to disregard spellings, punctuations, and grammatical errors in some papers which do not require them, they still assign lower grades to papers containing such errors in such papers (Chase, 1979; Scannell & Marshall, 1966).

The tutors indicated that they very often ($\underline{M}$ =3.79, $\underline{SD}$ =1.15), considered the grammar and language expression of students when scoring essay type tests in the colleges. The result of the tutors also indicated that the " number of scripts"; .($\underline{M}$ =2.66, $\underline{SD}$ =1.38) often influenced then when scoring essay-type tests in the colleges. The tutors reported that "gender of the student" ($\underline{M}$ =1.54, $\underline{SD}$ = 1.04) never influenced them when scoring essay-type tests in the colleges.

Research Question Five

What differences in the administration of test items exist between tutors in the teacher training colleges who have taught between 1-3 years and those who have taught above 3 years?

This research question also sought to find out whether there was any significant difference between tutors who have taught for 1-3 years and above three years in the colleges in relation to the basic principles in the administration and scoring of test items in the teacher training colleges.

Table 12 shows the results of the independent samples t-test of the equality of means of the responses of the tutors and the extent of differences between the two groups of tutors with regards to the administration and scoring of test items in the teacher training colleges.

Table 11

Test for Equality of Means of Test Administration between Teacher Training Tutors in the Colleges Based on Experience

Group	N	Mean	SD	t	df[a]	p value
1 – 3 Years	105	79.05	9.49			
				1.24	308	0.22
Above 3 Years	205	77.48	11.04			

[a]Degrees of freedom reduced because Levene's test shows violation of homogeneity of variance assumption.

The independent samples t-test for equality of means shows no statistically significant difference, $t(308) = 1.24$, p>0.05. This implies that there was no statistically significant difference between tutors who had taught between 1-3 years and above three years in the administration of test items in the colleges.

Research Question Six

What differences exist in the level of knowledge in test construction practices between tutors who had received a course in assessment practices and those who did not receive any course in assessment during the pre-service training?

To ascertain whether tutors in the teacher training colleges took a course in assessment (measurement and evaluation), majority 249 (80%) tutors indicated that they took a course in assessment. However, 61 (20%) tutors said that they did not receive any course in assessment during their pre-service training.

To test for significant difference between the two groups, (tutors who had received any course in test (measurement and evaluation) and those who did not receive take any course in testing, an independent samples t-test for equality of means was conducted to test the statistical significance difference between tutors took a course in assessment practices and those who did not take a course in assessment with respect the level of knowledge in test construction.. The result is presented in Table 12.

Table 12

Test for Equality of Means of Respondents Taking Courses in Testing and

No Course in Testing

Category	N	Mean	SD	t	df[a]	p
Took a course in assessment	249	2.56	0.24			
				2.40	308	0.03
Did take course in assessment	61	2.48	0.32			
Total	310					

[a]Degrees of freedom reduced because Levene's test shows violation of homogeneity of variance assumption.

Based on the sample, the overall result was statistically significant, t (308) = 2.40 p<0.05. The results of the study have shown that those who took a course in assessment during pre-service training seem to have had higher level of test construction practices than those who did not take a course in assessment at the pre-service stage. This result supports Ward (1980) and Stiggins (1991) that there is the need for all teachers to be taken through a course in testing (measurement and evaluation) at the pre-service stage of their training.

Research Question Seven

What differences exist in the level of knowledge in assessment practices between tutors who have had in-service training in testing and those who had no in-service training in assessment practices?

Item six of the questionnaire sought to find out whether tutors had attended any in service training programme in assessment. 158 (51%) tutors indicated that they had attended ever attended in-service training programme in assessment (measurement and evaluation) while 152 (49%) tutors reported that they had never attended an in-service training in testing.

To determine whether there was any statistically significant difference between those who had an in-service training in assessment and those who did not have in-service training in assessment, an independent samples t-test for test of equality of means was conducted. The result is presented in Table 13

Table 13

Test for Equality of Means for Respondent Tutors in the Colleges who have attended In–service Training Programmes on Assessment Practices

Response	N	Mean	SD	t	df[a]	p
Yes	158	179.27	18.4			
				2.64	308	0.01
No	152	173.34	21.34			
Total	310					

[A]degrees of Freedom Reduced Because Levene's Test Shows Violation of Homogeneity of Variance Assumption.

The independent samples t-test for equality of means shows a statistically significant result, $t(308) = 2.64$, $p < 0.05$. This implies that those who had attended in-service training in assessment had a higher level of knowledge than

those who did not attend in-service training on assessment. This perhaps is due to the fact that those who attended in-service training in assessment were exposed to better management of assessment practices in the colleges.

Research Question Eight

Which testing format (s) do tutors in the teacher training colleges use more often in testing the students?

Teacher- made tests are usually classified into two; namely objective tests and essay type tests (Lindquist, 1951). Item 8 of the questionnaire sought to find out how frequent tutors used different test formats in testing students in the teacher training colleges. The distribution of the respondents is shown in Table 14.

Table 14

Frequency Distribution on Use of Test Format(s)

Test format	Never	Not often	Often	Total (%)
Essay	9 (2.9%)	54 (17.4%)	247 (79.7%)	310 (100%)
Multiple choice	45 (14.5%)	79 (25.5%)	186 (60.0%)	310 (100%)
Short answer/supply	15 (4.8%)	38 (12.3%)	257 (82.9%)	310 (100%)
Matching	92 (29.7%)	110 (35.5%)	108 (34.8%)	310 (100%)
True or false	73 (23.5%)	109 (35.2%)	108 (41.3%)	310 (100%)

Overall, 257 (82.9%) tutors indicated that they often used short answer/supply test format in the colleges. In addition, 247 (79.7%) tutors said that they often used essay test format in testing students. Moreover, 186 (60%) tutors indicated that they often used the multiple-choice test format in testing the students in the colleges. The results of the study showed that 9 (2.9%) never used the essay-type test in testing the students.

From Table 14, it could be said that tutors equally used both objective and essay test formats in testing students in the colleges. Regarding objective test formats, it was found out from the responses that, multiple choice and short answer/supply test formats were often used by the tutors. The least objective test formats used by tutors were matching and true or false, which had 29.7% and 23.9% respectively.

In all, the finding that tutors used objective tests and essay tests supported the view in the literature that teacher-made tests are mainly of objective and essay types (Ackerman & Smith, 1988, Ebel & Frisbie, 1986, Lindquist, 1951, and Murchan, 1989). Thus, the findings have implications for instruction in measurement for appropriate teacher training colleges.

Research Question Nine

How do tutors in the teacher training colleges put student's continuous assessment scores to use?

To ascertain how students' continuous assessment scores are put to use after scoring, respondents were asked to indicate how they put the students' scores

to use. The purposes were: improving teaching and learning, recording, selection and placement, remedial teaching, certification and planning classroom activities. Table 15 summarizes the responses to this research question.

Table 15

Frequency Distribution by Test Purposes of Students Continuous Assessment (CA) Scores

Use of Students' test scores	Frequency	(%)
Improving teaching & learning	220	70.9
Receding purposes	28	9.0
Selection & placement	23	8.4
Organising remedial teaching	16	5.3
Certification purposes	14	4.5
Planning classroom activities	6	1.9
Total	**310**	**100**

Most tutors, 220 (70.9%) tutors indicated that they used the continuous assessment of the students for improving teaching and learning. Nine per cent (9%) reported that they used the students' continuous assessment scores for recording purposes, and 8% indicted they used the students' Continuous Assessment (CA) results for selection and placement.

Items on the questionnaire which were not related to any of the research question but have significant importance to the study have been analyzed. Means and standard deviations, and frequency distributions have been used in analyzing such items of the questionnaire.

Planning Schedule of Assessment

Item 14 of the questionnaire sought to find out how tutors planned schedule in the assessment of students in the teacher training colleges. The results are indicated in Table 16.

Table 16

Frequency Distribution by Planning of Schedule for Assessing Students

Time	Frequency	%
Monthly	158	51.0
Fortnightly	89	28.7
Weekly	54	17.4
Yearly	9	2.9
Total	**310**	**100**

The results from Table 16 indicated that, out of the 310 tutors for the study, 158 (51%) reported that they planned their schedule in assessing the

students monthly while 89 (28.7%) tutors indicated that the schedule for assessing students was done fortnightly. With the management of assessment practices, one very important area is the handling of the continuous assessment component of the students.

Storing Continuous Assessment Scores

Item 15 of the questionnaire requested tutors to indicate where continuous assessment scores of students were being kept in the college. The results are shown in Table 17.

Table 17

Frequency Distribution of Officer in Charge of Students' Continuous Assessment Scores

Officer	Frequency	%
College Assessment Officer	207	66.8
Vice - Principal	90	29.0
Head of Department	13	4.2
Total	**310**	**100**

Table 17 shows the results of the responses of the tutors. Sixty-seven per cent of the tutors in the teacher training colleges indicated that students' continuous assessment scores were kept by the College Assessment Officer while 90 (29.0%) tutors said that students' continuous assessment scores were kept by the Vice –Principal.

College Assessment Committee

For effective management of assessment practices in the teacher training colleges, it behoves on every college to have a College Assessment Committee which would ensure that assessment practices are effectively being practiced in the colleges. To this, item 19 on the questionnaire requested tutors to indicate whether the college had such committee in the colleges.

In finding out to what extent the College Assessment Committees' were effective, item 20 on the questionnaire asked respondents to indicate how effective the College Assessment Committees were in the teacher training colleges.

Table 18

Frequency Distribution on Effectiveness of College Assessment Committees

Response	Frequency	%
Very effective	46	14.8
Effective	142	45.8
Not effective	122	39.4
Total	**310**	**100**

As indicated in Table 18, 188 (59.6%) tutors said that the assessment committees in the colleges were effectively being managed. One hundred and twenty-two tutors (39.4%) indicated that the assessment committees in the

colleges were not effective in the colleges. Judging from the results, one may conclude that there is a little more work should to be done by college management to make the assessment committees more effective than the situation is now in the teacher training colleges.

The last section of the questionnaire asked tutors to state two major problems faced by them in the management of assessment practices in the teacher training colleges. Respondents were also requested to state two major recommendations that they thought would improve the management of assessment practices in the teacher training colleges. The result is indicated in Table 19

Table 19

Frequency Distribution of Major Problems Tutors Face in the

Management of Assessment Practices

Problem	Frequency	Percent (%)
Large class size.	100	32.26
Lack of skills in testing practices.	45	14.52
Difficulty in computation of test scores.	32	10.32
Lack of motivation.	21	6.80
Late submission of assignments by students.	17	5.48
Extra load on tutors.	16	5.16
Lack of offices for assessment officers in the colleges.	15	4.84
No clear cut- policy on assessment practices.	13	4.19
Frequent interferences of co-curricular activities.	12	3.87
Delay in submission of assessment materials.	11	3.5
Late submission of test scores to Vice-Principals.	10	3.22
Tutors perceive testing as an additional responsibility	8	2.58
Loaded syllabuses	6	1.93
Takes too much time of tutors	4	1.29

About thirty-three percent (33%) and fifteen percent (15%) indicated

that the major issues concerning problems that tutors faced in the management

of assessment practices in the teacher training colleges were large class size and lack of skills in test construction, administration and scoring of teacher – made tests in the colleges respectively. Responses to major suggestions that tutors were asked to state in the improvement of assessment practices in the colleges are shown in Table 20.

Table 20

Frequency Distribution of Major Suggestions in Improving Assessment Practices

Suggestion	Frequency	(%)
Responsibility allowances to be given to tutors for invigilation and setting/marking of exams.	100	32.29
Regular in-service training for tutors on testing practices in the colleges.	56	18.05
Special personnel to be made responsible in computing assessment records in the colleges.	41	13.22
Need for a standardized assessment approach for all colleges.	26	8.3
Training of heads of departments, vice-principals, assessment officers and tutors in ICT to facilitate the recording/computation of assessment results in the colleges.	20	6.45
Need to reduce large class size.	15	4.84
Need to provide a computer to each department.	15	4.84
Assessment materials should be provided to tutors early.	14	4.51
Regular monitoring of assessment practices by Teacher Education Division and the Institute of Education	13	4.19
Need to increase tutor-student ration in the colleges		
Total	**310**	**100**

As regards major suggestions, about thirty-three percent (33%) indicated that there is the need for tutors to be motivated by giving them extra allowances for invigilation and setting/marking of examinations. Tutors did not see testing as part of teaching process. It is heart-warming that about eighteen percent (18%) reported that there should be regular in-service training programmes on assessment practices for tutors in the colleges.

Discussion of Findings

This section discusses the research findings in relation to:

1. Principles to be followed when constructing teacher-made tests in the teacher training colleges;

2. Differences that exist between tutors who have taught for between 1-3 years and those who have taught above three years in the teacher training colleges;

3. Principles followed in the administration of teacher-made tests and scoring of essay-test type tests items;

4. Differences that existed between tutors in relation of experience in the teacher training colleges with regards to administration of teacher-made tests, and scoring of essay-test type test items;

5. Methods used in the grading of essay-test type items in the teacher training colleges;

6. Differences that exist in level of testing between tutors who had a course in testing (measurement and evaluation) and those who had on course in testing during their pre-service training;

7. Differences that existed between tutors who had in-service training in testing practices and those who had no in-service training;

8. Test formats often used in testing students in the colleges; and

9. How tutors put students' test score to use.

Research question one sought after principles tutors followed when constructing teacher-made tests in the teacher training colleges. The basic principles of constructing teacher-made tests have beeb developed over the years by a number of scholars like Cunningham, 1986; Ebel, 1972; Gronlund, 1985; Gullickson, 1982 and Mehrens and Lehman, 1991 to mention a few. For example, Gronlund (1985) identified five basic principles in the construction of teacher-made tests. These are:

    1. Determining the purpose of testing.

    2. Developing the test specifications.

    3. Selecting appropriate item types.

    4. Preparing the relevant test items and

    5. Assembling the test for administration.

Evidence has been found that in-service training in assessment for tutors in Ghana have contributed to little competency in the knowledge of test construction and administration of classroom teach-made tests as well as scoring of essay-type tests in the colleges (Amedahe, 1989; Curriculum Research

Development Division, 1999). The lack of skills and knowledge in the construction and administration of classroom/teacher –made tests as well in the scoring of essay-type tests has

Experience in teaching in the teacher training colleges is of utmost importance, as compared between tutors have taught in the teacher training colleges between 1-3 years and those above three years, in finding out whether there was any statistically significant difference in the construction of test items by tutors .in the colleges. The result of the study indicated that there was no statistically significant difference between tutors who had taught between 1-3 years and those above three years.

Items 1-7 of question 12 of the questionnaire sought to find out principles tutors followed in the administration of test items Item 12 was used to find out whether tutors under study maintained good physical conditions or not when testing their students. The physical conditions are important because as discussed in the literature review, they affect the performance of students in one way or the other on any test, (Amedahe, 1989). The findings of the study appear to suggest that tutors ensure good physical conditions in administering their tests. This situation is crucial to any test taking because as aptly pointed out by Ebel and Frisbie (1986) "the general goal of good test administration is to present and maintain the conditions that would permit all examinees to demonstrate their true level of achievement" (p. 210).

Moreover, the result of the study indicated that there was no statistically significant difference between tutors who had taught between 1 − 3 years and

above three years in the college in the knowledge of the administration of classroom/teacher-made tests in the colleges. This could partly be attributed to the lack of skills and competency in the knowledge   assessment practices by tutors in the teacher training colleges (Amedahe, 1989; Etsey, 2003).

It is not sufficient to assume that because a scoring key/scheme has been prepared for scoring an essay test, consistency in scoring would be achieved. Consistency in scoring would only be achieved when the marking scheme is followed rigorously. Anything short of this would lead to either overstoring or underscoring of the essay (Anamuah-Mensah & Quagrain, 1998).

Item six of the questionnaire sought to answer research question seven. The research question sought to find out whether any difference existed in the level of knowledge in assessment practices between those who had in-service training in assessment and those who did not have in–service training in assessment practices. The results of the study indicated that there was a statistically significant difference between tutors who had in-service training in assessment and those who had no in-service training in assessment The result implies that those who had in-service training in assessment might have been exposed to better assessment practices than those who did not have in-service training in assessment in the knowledge of construction, administration of classroom/teacher-made tests as well as in the scoring of essay-type tests in the colleges

Item eight, which answers research question eight, was to find out which of the testing formats were often used by tutors in testing the students in the

colleges. The result indicated that tutors equally used both objective and essay test formats in testing their students in the colleges. Majority, (about 83%) indicated that they often used short answer/supply test format while about 80% said that they often used essay test format in testing their students in the colleges.

Item 17 of the questionnaire was used to answer research question nine. The question sought to find out how tutors put the students assessment scores to use. Result from the data showed that most tutors, about 71% used the assessment scores of the students for improving teaching and learning. The result indicated that about 2% used the student assessment scores in planning activities, while about 5% reported that they used the students' assessment scores for organizing remedial teaching in the colleges.

Summary

The descriptive statistics (frequencies, percentages, means and standard deviation and means) and independent samples t-test tools were used in analyzing the nine research questions. The findings are presented as follows:

1. Tutors in the teacher training colleges fairly followed the basic principles in constructing teacher-made tests in the colleges.

2. There was no significant difference between tutors who took a course in assessment and those who did not take a course in assessment (measurement and evaluation).

3. The findings suggest that tutors ensured good physical conditions in the administration of teacher-made tests in the colleges.

4. There was no significant difference between tutors who had taught between 1-3 years and above three years in the administration of test items in the colleges.

5. Overall, the results indicated that the main method often used by tutors in the scoring of essay-type tests was the analytic method.

6. Overall, about 83% reported that they often used short/supply test formats while about 80% reported that they often used essay-type tests in testing students in the colleges.

7. Most tutors, 71% indicated that they used students' test scores for improving teaching and learning.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview of Research Problem and Methodology

The study sought to investigate the management of assessment practices with regards to basic principles in the construction, administering and scoring of classroom or teacher-made tests with particular reference to the scoring of essay test items in the teacher training colleges. The study was, therefore, the goodness of fit between prescription and practice of educational achievement testing in the teacher training colleges.

A descriptive sample survey was conducted in 20 selected teacher training colleges using 310 respondents and questionnaire. Statistical procedures used in data analysis were mainly means and standard deviations, independent t-test, frequency distribution and percentages. The following were the summary of the main key findings of the study:

Summary of Key Findings

The following were the main findings of the study:

1. There was no significant difference between the procedures used in

constructing classroom or teacher-made tests by those who had taken a course in measurement and evaluation and those who have not taken any course in testing.

2.   Tutors in the teacher training colleges followed some basic principles in the constructing of classroom or teacher-made tests.

3.   There is no significant difference between tutors who have taught between 1-3 years and above 3 years in the teacher training colleges in the knowledge in the construction of classroom or teacher-made tests in the teacher training colleges.

4.   There was no significant difference between tutors who have taught between 1-3 years and above 3 years in the teacher training colleges in the knowledge in the administration of classroom/teacher-made tests in the teacher training colleges

5.   Tutors in the teacher training colleges used mainly the analytic method in scoring their essay tests.

6.   There was statistically significant difference between tutors who had in-service training in assessment and those who did not have an in-service training in assessment.

Conclusions

It is evident from the study that, on the whole, tutors in the teacher training colleges had limited skills and competence in the knowledge of construction, administration of classroom/teacher made tests and scoring of essay-types tests.

That assessment of student performance is an important skill for tutors was not disputed by those in either the teacher education or the measurement community. Whether teachers get their knowledge and skills in testing practices before they begin teaching or in an in service or professional development setting may be an issue. Most tutors had exposure to test and measurement training.

It is evident from the study that, on the whole, tutors in the teacher training colleges have little knowledge in testing practices, even though most of them have been exposed to a course in testing (measurement and evaluation) at the pre-service stage of their training. In addition, they lacked skills in test construction. This supports statement made by Ashworth (1982) that "teachers often prepare tests at the last minute; the test preparation is slipshod, unreliable and poor in most respects. It often fails to motivate and direct pupils' future learning when that may have been its main purpose" (p. 31).

On the basis of the results obtained in this study, the following conclusions can be drawn. Teacher-made tests have been of low quality; and that items might be generally of low order types and tutors are not well grounded in the construction, administration of teacher-made tests. However, they are expected to be knowledgeable in testing practices. This supports the observation made by Izard (1996) that although teachers are expected to be measurement-competent educators, significant deficits in teachers' knowledge concerning testing have been found, and that most teachers acknowledge that testing and measurement come from trial and error learning in the classroom. Given this situation, tutors should follow the basic principles underlying construction,

94

administration and scoring of teacher-made tests in the colleges in improvement of testing practices in the colleges, as well as in the quality of education.

Recommendations

In view of the following conclusions, and based on the findings resulting from the study, the following recommendations are made for the improvement in the management of assessment practices by tutors in the teacher training colleges:

1. As regards to putting skills and competencies in testing practices to use what tutors learn from training sessions, heads in the teacher training colleges should ensure that tutors implement what they learn from workshops in the teacher training colleges by often observing the tutors in the classrooms when teaching.

2. The colleges should organise in-service training on testing practices for tutors in the teacher training colleges in order to improve their skills in the construction, administration of teacher-made tests and scoring of essay-type test in the colleges, and also, the Teacher Education Division of the Ghana Education Service in collaboration with the Institute of education of the University of the University of cape Coast should take steps take steps to organise in-service training course, seminars or workshops for tutors in the teacher training colleges. This should involve both tutors who took a course in testing practices and those who did not to improve their skills in test construction, administration of teacher-made tests and scoring of essay-type tests

Suggestions for Future Research

The following are suggested for future research:

1. The study should be carried out to determine to what extent validity and reliability of teacher-made tests contribute to the testing practices, in the teacher training colleges.

2. Study in the teacher training colleges would be useful to determine the interpretation of tests scores of students, and how tests takers are informed of such interpretation.

3. If the study of testing practices in the teacher training colleges is to be complete, then, there is also the need to replicate the study, by researching on testing practices in the areas of construction of objective test items.

REFERENCES

Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple–choice and free response writing test. *Applied Psychological Measurement, 12,* 117-128.

Adamolekun, J. (1985). Effective Test Construction: A case for the junior senior secondary school mathematics curricula. *Journal of the School of Pure Sciences.* 1(3), 197-224.

Aidoo-Taylor, N. (1993). Practical assessment in the school system. *Journal of the Institute of Education,* 2 (2*), 67 -74.

Akyeampong, A. K. (1997). *Continuous assessment in post-secondary teacher training colleges in Ghana: A case study evaluation.* Unpublished Ph.D. Thesis. University of Nottingham, Nottingham: United Kingdom.

Amedahe, F. K. (2000). Issues in combining continuous assessment scores with external examination scores for certification at the basic and secondary school levels. *Journal of Educational Management* 3, 112-123.

Amedahe, F. K. (1989). *Testing practices in secondary schools in the Central Region of Ghana.* Unpublished Masters thesis, University of Cape Coast, Cape Coast.

Amedahe, F. K. & Etsey, Y. K. A. (2003, October). *Assessment criteria, feedback and evaluation.* Unpublished paper presented at the University of Cape Coast, Ghana.

Anamuah-Mensah, J. & Quagrain, K. A. (1998). Teacher competence in the use of essay -tests. A study of secondary schools in the Western region of Ghana. *The Oguaa Educator.* 12 (1), 31-43.

Anastasi, A. (1976). *Psychological testing.* New York: Macmillan Publishing
Company.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities.
American Psychologist, 36, 1086-1093.

Andrews, T. E. & Barnes, S. (1990). Assessment of teaching. In W. R. Houston
(Ed.), *Handbook of research on teacher education: A project of the
Association of Teacher Educators* (pp. 569 – 598). New York: Macmillan.

AREA/APA/NCME. (1985). *Standards for educational and psychological
testing.* Washington, DC: American Psychological Association.

Asamoah-Gyimah, A. (2002). *An evaluation of the practice of continuous
assessment in the senior secondary schools in the Ashanti Region of
Ghana.* Unpublished thesis, University of Cape Coast, Cape Coast, Ghana.

Ashworth, A. E. (1982). *Testing for continuous assessment: A handbook for
teachers in schools and colleges.* Ibadan: Evans Brothers Limited.

Bartels, E. K. (2003). The practice of continuous assessment in teacher
training colleges in Ghana. *Journal of Educational Development   and
Practice,* 1(1), 59-72.

Bejar, I. (1984). Educational diagnostic assessment. *Journal of Educational
Measurement*, *21*, 175-189.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives. Handbook1:
Cognitive domain.* New York: David Mckay and Co. Inc.

Boothroyd, R. A., McMorris, R. F., & Pruzek, R. (1992, April). *What do
teachers know about measurement and how did they find out?* Paper
presented at the annual meeting of the National Council on Measurement
in Education, San Francisco.

Capper, J. (2007). Testing to learn-Learning to test. *International Reading
Association. Academy for Educational Development. New Washington:*
Delaware Publishing Company.

Charmey, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18,* 65-81.

Chase, C. I. (1979). *Measurement for educational evaluation.* Philipines: Addison –Wesley Publishing Company Inc.

Clawson, T. W, Firment, C. K. & Trower, T. L. (1981). Test anxiety: Another origin for racial bias in standardized testing. *Measurement and Evaluation in Guidance,* 13 (4), 210-215.

Culler, R. E., & Hollohan, C. J. (1980). Anxiety and academic performance. *Journal of Educational Measurement, 72,* 16-20.

Cunningham, G. K. (1986). *Educational and psychological measurement.* New York: Macmillan Publishing Company.

Curriculum Research and Development Division (CRDD). (1999). *Investigation into student assessment procedures in public junior secondary schools in Ghana.* Accra: Ghana Education Service.

Diederich, P. B. (1967). Cooperative preparation and rating of essay tests. *English Journal, 56,* 145-150

Diederich, P. B., French, S., & Carlton, S. (1961). *Factors in judgments of writing ability* (Research Bulletin, 61-65). Princeton, N. J: Educational Testing Service.

DuBois, P. H. (1966). A test dominated society: China 115 B. C. – 1905 A. D. Proceedings of the 1964 Invitational Conference on Testing Problems. In A. Anastasi (Ed.). *Testing Problems in Perspective,(p. 56).* Washington D. C: American Council on Education.

DuBois, P. H. (1970). *A history of psychological testing.* Boston: Allyn and Bacon Inc.

Ebel, R. L. (1972). *Essentials of Educational Measurement.* Englewood Cliffs. New Jersey: Prentice Hall Inc.

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* Englewood Cliffs, NJ: Prentice Hall Inc.

Etsey, Y. K. A. (2003). Pre-Service Teacher's Knowledge of Continuous Assessment Techniques in Ghana. *Journal of Educational Development and Practice*, 1 (1), 1-18

Frankel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education* New York: McGraw Hill.

Frankel, J. R. & Wallen, N. E. (1993). *How to design and evaluate research in education.* New York: McGraw Hill.

Frary, R. B. (1981). Cheating? *Teaching and Learning, 2,* 3-4.

Freedman, S. (1981). Influences on evaluators of expository essays: Beyond the text. *Research in the Teaching of English, 15,* 245-255

Freedman, S. (1979). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology, 71,* 328-338.

Freedman, S. W,. & Calfee, R. C. (1983). Holistic assessment: Experimental design and cognitive theory. In P. Mosenthal, L. Tamor & S. W. Walmsley (Eds.). *Research on writing: Principles and methods (pp. 75-98).* New York: Longman.

Galton, F. (1967). Classification of men according to their natural gifts. In J. A. Wiseman. (Ed.), *Intelligence and ability.(pp.34-40),* Baltimore: Penguin.

Gay, R. L. (1992). *Educational research: Competencies for analysis and application.* New York: Macmillan Publishing Company.

Gipps, C. (1992a). *National testing at seven: What can it tell us?* London: Hodder and Stoughton Publishing Company

Gipps, C. (1992b). *What we know about effective primary teaching.* London: Tufnell Press.

Gronlund, N. E. & Linn, R. L. (1990). *Measurement and evaluation in teaching,* New York: Macmillan.

Gronlund, N. E. (1985). *Measurement and evaluation.* New York: Macmillan.

Gronlund, N. E. (1982). *Constructing achievement* tests. Englewood Cliffs NJ: Prentice-Hall.

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. (1982, August). *The practice of testing in elementary and secondary schools.* Paper presented at the Rural Education Conference. Mahattan K. S : Kansas State University

Gullickson, A. R.,. & Ellwein, M. C. (1985). Post hoc analysis of teacher–made tests: The goodness–of-fit between prescription and practice: *Educational Measurement: Issues and Practice.* 4(1*),* 15-18.

Hales, L. W., & Tokar, E. (1975). The effect of quality of preceding responses on the grades assigned to subsequent responses to an essay question. *Journal of Educational Measurement, 12,* 115-117.

Harris, W. (1977). Teacher response to student writing: A study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. *Research in the Teaching of English, 11,* 175-185.

Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta, 72,* 540-545.

Impara, J. C., Divine K. P., Bruce, F. A., Liverman, M. R., & Gay, A. A. (1992). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10* (4), 16-18.

Infantino, R. L. (1976). *Testing and accountability: A survey of the knowledge and attitudes of New York state secondary school English teachers.* Doctoral dissertation, State University of New York at Buffalo. (University Microfilms, No. 77-6143).

Izard, J. (1996, September). *Quality assurance in educational testing.* Paper presented at the 22nd. International Atomic Energy Agency (IAEA) Conference, Beijing.

Kirsch, I. S.,. & Guthrie, J. T. (1980) Construct validity of functional reading tests. *Journal of Educational Measurement, 17,* 81-93.

Kline, P. (1986). *A handbook of test construction.* New York: Methuen.

Kuo, P. W. (1915). *The Chinese System of Public Education. Contribution to Education.* New York: Columbia University Teacher College.

Lein, A. J. (1976). *Measurement and evaluation of learning.* Dubuque, IA: William C. Brown.

Lindon, K. W., & Lindon, J. D. (1968). *Modern mental measurement: A historical perspective.* Boston: Houghton Miffin Co.

Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist, (Ed.). *Educational Measurement* (pp. 119-158). Washington DC: American Council on Education.

Mardell-Czudnowski, C. D. (1980). *Learning Disability Quarterly, 3* (1). 73-83.

Markham, L. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal, 13,* 277-283.

Marso, R. N., & Pigge, F. L. (1989, April). *The status of classroom teachers' test construction proficiencies: Assessment by teachers, principals, and supervisors validated by analyses of actual teacher-made tests.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco.

McCallum, B., McAlister, S, Brown, M., & Gipps, C. (1993). Teacher assessment at key stage one . *Research Papers in Education,* 20 (*3*), 305-327.

McColly, W. (1970). What does educational research say about judging of writing ability? *Journal of Educational Research, 64,* 147-156.

Mehrens, W., A. & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practice, 8* (1), 14-22

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology.* New York: Harcourt Brace College Publishers.

Mehrens, W. A. (1984a). National tests and local curriculum: Match or mismatch? *Educational Measurement: Issues and Practice, 3 (3),* 9-15

Mullis, I. V. S. (1984). Scoring direct writing assessment: What are the alternatives? *Educational Measurement: Issues and Practice, 14,* 16-18.

Murchan, D. P. (1989, April). *Essay versus objective testing in the context of large-scale assessment programs.* Paper presented at the annual meeting of the National Council of Measurement in Education. San Francisco.

National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing.* Washington, D. C: Author.

Nold, E., & Freedman, S. (1977). An analysis of readers' responses to essays. *Research in the Teaching of English, 11,* 164-174.

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

Nitko, A. (1983). *Educational tests and measurement: An introduction.* New York: Harcourt Brace Jovanovich.

Nunnally, J. C. (1978). *Psychometric theory,.* New York: McGraw-Hill.

Nunnally, J. C. (1972). *Educational measurement and evaluation.* New York: McGraw-Hill.

Osuala, E. C. (2001). *Introduction to research methodology .* Onitisha, Nigeria: Africana EEP Publishers Ltd.

Pavet, W., Diener, E., Colvin, C. R., & Sandvick, E. (1991). Further validation of satisfaction with life scale: Evidence for the cross method convergence of well being measures. *Journal of Personality Assessment. 57,* 149-161.

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.

Polit, D. F. & Hungler, B. P. (1995). *Nursing research: Principles and methods (5th. ed.).* Philadelphia: J. B. Lippincott Company.

Quagrain, A. K. (1992). *Teacher–competence in the use of essay tests: A study of secondary schools in the Western Region of Ghana.* Unpublished thesis. University of Cape Coast, Cape Coast, Ghana.

Reynolds, M. C. (Ed.). (1989). *Knowledge base for the beginning teacher.* New York: Pergamon Press.

Roberts, D. M. (1987). Limitations of the score-difference method in detecting cheating in recognition test situation. *Journal of Educational Measurement*, *24* (1), 77-81.

Rudman, H. C., Mehrens, W. A. & Wanous, D. S. (1983). *Integrating assessment with instruction:* East Lansing Institute for Research on Teaching. Michigan State: University College of Education.

Sarantakos, S. (2005) *Social research.* New York: Palgrave Macmillan.

Sarason, H,, Hill, K., & Zimbrado, P. (1964). A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. *Monographs of the Society for Research in Child Development, 29, 28-34*

Scanell, D. P., & Marshall, J. C. (1966). Effect of selected composition errors on grades assigned to essay examinations. *American Educational Research Journal, 3,* 125-130.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education,* 38 (3), 57-63.

Seifert, K. L., & Hoffnung, J. R. (1994). *Child and adolescent development* . Boston: Houghton Mifflin Company.

Spandel, V., & Stiggins, R. J. (1988). *Writing assessment in the classroom. .* White Plains, New York: Longman.

Spandel, V. (1981). *Classroom application of writing assessment: A teacher's handbook.* Portland, OR: North–west Regional Education Laboratory.

Stanes, D. L. (1992). Analytic responses to conceptual style test as a function of instructions. *Child Development,* 44, 389-391.

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

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22(4), 271-286.*

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Issues and Practice, 8(2)*, 5-14.

Takeuchi, R. T. (1977). Attitudes of elementary teachers toward testing: Use and abuse of standardized tests in California, 1976-77. *Dissertation Abstracts International,* 38, 4500A. (University Microfilms No. 7731120).

Tamakloe, E. K., Atta, E. T., & Amedahe, F. K. (1996). *Principles and methods of teaching.* Accra: Black Mask Ltd.

Thorndike, E. L. (1903). *An introduction to the theory of mental and social measurements.* New York: The Science Press.

Thorndike, R. L., & Hagen, E. (1977). *Measurement and evaluation in psychology.* New York: Wiley.

Thorndike, R. L., & Hagen, E. (1969). *Measurement and evaluation in psychology and education,* New York*:* John Wiley & Sons Inc.

Tiedt, I. M. (1983). *The language arts handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Trentham, L. L. (1975). The effect of distractions on sixth grade students in a testing situation. *Journal of Educational Measurement,. 12,* 13-17.

Trochim, W. (2000). *The research methods knowledge base* Cincinnati, OH: Atomic Dog Publishing.

Tyler, R. W. (1950). *Basic principles of curriculum and instruction.* Chicago: University of Chicago Press

Vacc, N. B. (1998, September). Writing evaluation: Examining four teachers' holistic and analytic scores. *The Elementary School Journal, 90*(1), 87-95

Ward, J. G. (1980). Teachers and testing*: A survey of knowledge and attitudes.* In L. M. Rudner (Ed.) *Testing in our schools.* (pp. 15-24), Washington, D. C: National Institute of Education.

Ward, W. C., Kline, R. G., & Flaugher, J. (1986). *College board computerized placement tests: Validity of an adaptive test of basic skills (*Research Report 86-29). Princeton NJ: Educational Testing Service.

APPENDIX A

ZONAL CENTRES OF TEACHER TRAINING COLLEGES IN GHANA

ZONE 1:

1. Komenda Training College

2. Foso Training College

3. Holy Child Training College

4. OLA Training College

5. Enchi Training College

ZONE 2:

1. Kibi Training College

2. SDA Training College

3. Presbyterian Women's Training College

4. Accra Training College

5. Presbyterian Training College

6. Mount Mary Training College

7. Ada Training College

ZONE 3:

1. Wesley College   .

2. St. Louis Training College

3. Akrokerri Training College

4. Agogo Women's Training College

5. Offinso Training College

6. St. Joseph's Training College

7. Berekum Training College

8. Abetifi Training College

9. Wiawso Training College

10. Mampong Technical Training College

11. St. Monica's Training College

ZONE 4:

1. St. Francis Training College

2. St. Theresa's Training College

3. Jasikan Training College

4. E. P. Training College, Amedozfe

5. Dambai Training College

6. Peki Training College

7. Akatsi Training College

ZONE 5:

1. Bagabaga Training College

2. Tamale Training College

3. Bimbilla Training College

ZONE 6:

1. St. John Bosco's Training College

2. Gbewaa Training College

3. Tumu Training College

4. N.J.A. Training College

SOURCE: INSTITUTE OF EDUCATION, UNIVERSITY OF CAPE COAST, CAPE COAST. 2008)

APPENDIX B

SAMPLED TRAINING COLLEGES FOR THE STUDY

Serial No.	College	Zone
1.	OLA Training College	1
2	Foso Training College	1
3.	Komenda Training College	1
4.	SDA Training College	2
5	Accra Training College	2
6	Presbyterian Training College	2
7	Kibi Training College	2
8	St. Monica's Training College	3
9	Agogo Women's Training College	3
10	Atebubu Training College	3
11	Berekum Training College	3
12	Offinso Training College	3
13	Wesley College	3
14	Dambai Training College	3
15	E.P. Training College, Amedozfe	4
16	Akasti Training College	4
17	Bagabaga Training College	4
18	Tamale Training College	5
19	St. John Bosco's Training College	6
20	NJA Training College	6

QUESTIONNAIRE FOR TUTORS

Assessment Practices of Teacher Training College Tutors

The purpose of this questionnaire is to obtain information for a study in the assessment practices in the teacher training colleges. It is hoped that the results of the study will help in policy decisions concerning teacher education in Ghana. Kindly complete it anonymously. Please respond to **all** items and do it honestly. No attempt will be made to associate your name or institution with the completed instrument. All information will be kept confidential.

SECTION A

Background information//Personal data

DIRECTIONS: Please tick [√] the box that best describes your response (s) where applicable or write in the space provided.

1. Gender Male   [ ]

   Female   [ ]

2. Number of years you have been teaching in TTC

   1 – 3 Years   [ ]

   4 – 7 Years   [ ]

   8 – 11 Years   [ ]

   12 Years and above   [ ]

3. Highest Academic Qualification

   Bachelor of Education (B.Ed)   [ ]

BA/BSc	[ ]
Bachelor of Arts/Social Science	
with PGDE	[ ]
Master of Education (M.Ed)	[ ]
Master of Philosophy (M.Phil)	[ ]
MA/MSc	[ ]
Others	[ ] Please specify

4. Subject (s) you are teaching

English Language	[ ]
Mathematics	[ ]
Science	[ ]
Education	[ ]
Environmental Social Studies	[ ]
Vocational Studies	[ ]
Technical Skills	[ ]
Others	[ ]

5. Status in college

Vice - Principal	[ ]
Head of Department	[ ]
Assessment Officer	[ ]
Tutor	[ ]

CONSTRUCTION OF ACHIEVEMENT TESTS

6. Have you ever taken a course in testing (measurement and evaluation)?

Yes   [ ]

*No   [ ]

7. Have you ever had any inservice training in testing (measurement and

evaluation)?

Yes   [ ]

No   [ ]

8. Indicate with a tick ([√]) **how often** you use the following test formats in

assessing your students

Test Format	Very often	Fairly often	Often	Not often	Never
1. Essay					
2. Multiple choice					
3. Short-answer/Fill in- the blanks/Supply					
4. Matching					
5. True & False					

9. Indicate with a tick [√] your level of practice on the following activities

regarding test construction

Activity	Always	Very often	Often	Not often	Never
1. I write individual test items at least two weeks before the date of testing.					
2. I prepare a marking scheme before students take the.test					
3. I review test items by letting another tutor in the subject area read over them.					
4. I write directions/instructions for the test clearly					
5. I copy questions from past teacher training colleges examinations questions set by the Institute of Education					
6. I develop test items only when it is time to assess students					
7. I use a test specification table when writing test items					
8. I copy test questions from textbooks					
9. I match instructional objectives with test items					
10. I arrange test items in order of increasing difficulties					

Indicate with a tick [√] your level of practice on the following activities regarding test construction

11. I write items based on information that students know					
12. I arrange objective test answers in a pattern to make scoring easy					
13. I refer to the principles of test construction when I write test items					
14. I consider the purpose of the test before developing test items					
15. The institution conducts in-service training in test construction for tutors					
16.I allow students to select a specific number of items from a given set of items in an essay test					
17. I prepare more items than needed before I review and select in the examination					
18. I evaluate the test as a whole before I submit it for typing					

10. What is your **current** level of knowledge on the following concepts and terms in educational testing and practices?

Term and Concept	I remember exactly what it means	I remember fairly what it means	I have forgotten most of what it means	I never learnt about it
1.Test specification table				
2. Item difficulty				
3. Taxonomy of educational objectives				
4. Distracter analysis of multiple –choice items				

SECTION C

TEST ADMINISTRATION AND SCORING

10. Which of the following methods do you adopt **MOST** when scoring essay tests?

- Analytic Scoring (Scoring point by point)   [ ]

- Holistic Scoring (Reading through before giving a total score)   [ ]

11. Indicate the extent of influence with a tick, [√] on the following factors when marking essays

Factor	Very large extent	Fairly large extent	Moderate extent	Slight extent	Not at all
1. Handwriting					
2. Number of scripts					
3. Grammar and language expression					
4. Knowledge of students' previous knowledge					
5. The gender of student					
6. The length of student's essay					

12. Indicate your level of practice with a tick [√] on the following activities regarding test administration and scoring

Activity	Always	Very often	Often	Not often	Never
1. After examination questions are typed, I proofread for mistakes before students take the test					

Indicate your level of practice with a tick [√] on the following activities regarding test administration and scoring

2. I ensure good sitting arrangements to prevent students from copying from each other					
3. During examinations, I tell students that if they do not write fast, they will fail					
4. During examinations, I prepare for and expect emergencies					
5. I give hints to students when they ask about individual test items during examinations					
6. I read novels, newspapers or mark questions when I invigilate tests/examinations					
7. I inform students in advance about content/topics that the tests/examinations cover					

8. I promptly mark examination papers					
9. I score essay tests, question by question					

10. I prepare marking schemes for essay tests after students have taken the test				
11. I promptly record test results				
12. I reshuffle essay scripts before I score the next question				
13. I periodically rescore previously scored paper to check consistency in scoring				
14. I score answer scripts with the names of the students known to me				
15. I keep previously scored items out of sight when scoring the rest of the items				
16. I provide comments on essay tests for students to facilitate learning				
17. The first few essays I score influence the rest of the scores I give				
18.. I give a separate mark for the mechanics of writing such as correct grammar, flow of expression etc when scoring essays				
19.I promptly submit test results to the head of department				

SECTION D

GENERAL ISSUES RELATED TO ASSESSMENT PRACTICES IN THE

TEACHER TRAINING COLLEGES

13. How often do you plan the schedule for assessing the students?

Weekly   [ ]

 Monthly   [ ]

 Fortnightly   [ ]

Yearly   [ ]

14. Indicate (with a tick [√]) the number of tests administered to students for the

number of credit (s) course (s) you taught during the last semester.

Activity	More than 4 times	4 times	3 times	2 times	Once
1. Class test/					
2. Quiz					
3. Project work					

15. Indicate with a tick [√] the person who keeps continuous assessment scores in

the college

Vice – Principal (Academic)   [ ]

Head of Department   [ ]

College Assessment Officer   [ ]

16. Indicate with a tick [√] the appropriate place where continuous assessment records are kept in the college

College cabinet   [ ]

Exercise books   [ ]

College Assessment's office   [ ]

Vice Principal's office   [ ]

On the computer   [ ]

17. To what use do you put the students' continuous assessment scores? (Tick as many as possible)

To improve teaching   [ ]

To organize remedial teaching   [ ]

To plan classroom activities   [ ]

For record purposes   [ ]

Others (Please, state) …………………………………………………………..

18. Do you have Assessment Committee in the college?

Yes   [ ]

No   [ ]

19. How effective is the role of the College Assessment Committee in relation to the assessment practices in the college?

Not effective   [ ]

Effective   [ ]

Fairly effective   [ ]

Very effective   [ ]

20. State **TWO** major problems that you face in the management of assessment (eg. tests and test scores) in the college.

…………………………………………………………………………………………

………………………………………………………………………………………

…………………………………………………………

21. Suggest **TWO** recommendations to improve assessment practices in the college.

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

………………………………………………………………………………

THANK YOU VERY MUCH FOR YOUR TIME, CONTRIBUTIONS AND HELP

APPENDIX D

Cronbach Alpha Reliability Test

Reliability Statistics

Case Processing Summary

		N	%
Cases	Valid	310	100.0
	Excluded	0	.0
	Total	310	100.0

a. List wise deletion based on all variables procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.775	81