

# LECTURERS' COMPETENCY IN THE CONSTRUCTION OF MULTIPLE-CHOICE AND ESSAY-TYPE TESTS AT THE UNIVERSITY OF EDUCATION, WINNEBA – GHANA

ESHUN PETER

Department of Psychology and Education  
University of Education, Winneba-Ghana

## ABSTRACT.....

*This study assessed whether lecturers at the University of Education, Winneba, followed the laid down principles of the construction of multiple-choice and essay-type test items. The study adopted the descriptive survey design for its execution. The stratified sampling technique was used to select 195 lecturers at the Winneba Campus of the University of Education. The data collecting instrument used in the study was a questionnaire. The main statistical method used in the analysis of the data was analysis of variance (ANOVA) at an alpha level of 0.05.*

*Findings of the study showed that: (a) lecturers at the Winneba campus of UEW generally did not apply the prescribed principles for constructing both multiple-choice and essay-type test items; (b) the number of years one lectured did not have significant influence on lecturers at UEW when it comes to applying the prescribed principles of constructing both multiple-choice and essay-type test items.*

*It was recommended among other things that, lecturers should put into practice prescribed principles of test construction when constructing their tests.*

.....



## INTRODUCTION

The importance of testing in education makes it an important topic of continuing research. According to Davis (1999), many teachers dislike preparing and grading exams, and most students dread taking them, yet tests are powerful educational tools that serve at least four functions. First, tests help the teacher to evaluate students and assess whether they are learning what the teacher is expecting them to learn. Second, well-designed tests serve to motivate and help students structure their academic efforts. Carter (1984), McKeachie (1994), and Wergin (1988) reported that students study in ways that reflect how they think they will be tested. If they expect exams that focus on facts, they will memorize details and if they expect a test that requires problem solving or integrating knowledge, they will work toward understanding and applying information. Third, tests can help the teacher understand how successfully he/she presents material. Finally, tests can reinforce learning by providing students with indicators of what topics or skills they have not yet mastered and should concentrate on. Despite these benefits, testing is also emotionally charged and anxiety producing.

Educators often measure performance and attribute of individuals with a test, but it is irresponsible to randomly select a test. Before one selects a test, or possibly construct one's own test, one should consider, (a) if criterion referenced measurement or norm referenced measurement should be used. With the norm-referenced test scores are interpreted on a relative basis in terms of the performance of a "test" or sample group (called a norm group), while with the criterion-referenced, test scores are interpreted on the basis of some absolute performance criterion that describe performance in terms of the kinds of tasks a person with a given score can do, and (b) the criteria for determining a good test - validity, reliability, objectivity and administrative feasibility (Miller, 1998).

In the current debates about poor educational outcomes and the need for education reform in Ghana, testing has been viewed as both part of the problem and part of the solution. Tests have been regarded as the most tangible cues and the most crucial yardstick in determining the attainment of the objectives of any learning experience. It is from the outcome of testing that teaching and learning objectives are appraised and refined. Teaching and testing can, therefore, be considered as the two sides of a coin. Without testing the teacher would not be in a strong position to know whether the objectives set out to be achieved at the outset have been attained. Similarly, its diagnostic function ensures identifying strengths and weaknesses of students, determining the form and kinds of remediation, counseling services, selection, placement, classification and general facilitation of learning.

Sinclair (1970) was of the view that decisions of major consequence to the individual are increasingly being made on the basis of his/her performance in tests by teachers. Yet, little if any thought, is given to the qualification or skill of the teacher as an evaluator. Decisions such as declaring that a student has passed or failed his or her test and consequently qualified to enter the next stage of the educational process or has to repeat a course are cases in point.



The fact that lecturers including even those lecturers who have not received any formal instruction in educational measurement or testing, teach and test their students have led some people to underrate the importance of using testing practices that measurement or testing experts deem appropriate and useful. For instance, Ebel (1979) recorded that:

There are some professors who do not take the responsibility of evaluating students' achievement seriously. They regard testing and grading as nonessential administration 'red tape' largely divorced from the essential process of education, since in their view, the whole process is unimportant, it is unnecessary for them to expend much time or effort to achieve high validity or precision in the process (p. 226).

Even though this idea was expressed some forty-two years ago about university professors on achievement testing, it could be observed to be the situation among some Ghanaian lecturers today. In Ghanaian universities, lecturers are given the duty of giving instructions in the courses they handle, develop and construct test items, administer, score and interpret the scores without proper supervision as to what he or she does. These lecturer-made tests focus on the specific objectives the lecturer has set and helps the lecturer immensely to clarify and refine these objectives where necessary.

Most Ghanaian universities rely on lecturer-made tests to assess students' performance. This implies that, the enormous task of test construction, administration, scoring and interpretation has to be shouldered by the lecturer.

The multi-faceted question is posed as to whether the Ghanaian lecturer is adequately prepared and professionally armed to measure up to the task of constructing good test items. The lecturer's expertise in handling tests will be necessary to justify his/her crucial role as an assessor or a judge in his/her course. Whatever decision he/she takes may have far reaching consequences on the student. It stands to reason therefore, that prospective lecturers must have to demonstrate competence in test construction and use.

Most lecturers in Ghana use both the objective and essay type items in their tests and not much research work has been done with regards to how competent these lecturers are in the construction of multiple-choice and essay-type tests following the basic scientifically prescribed principles of testing.

### **Statement of the Problem**

The popular use of multiple-choice and essay-type tests by lecturers does not necessarily guarantee that they adhere to the laid down principles in the construction of these tests. It is a well known fact that most teachers have not been well trained in test construction, whether subjective or objective (Kusiszyn & Borich, 1987). Although most teachers including lecturers construct tests, little has been done in



terms of research to find out whether classroom teachers actually follow the principles laid down by test specialists in writing test items (Gullickson & Ellwein, 1985)

Studies by Amedahe (1989) and Quaigrain (1992) have revealed that most secondary school teachers lack the skills required 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. In cases where teachers underwent a course of instruction in testing and assessment, studies revealed that few teachers depend upon their knowledge in test construction (Amedahe, 1989; Quargrain, 1992).

Some lecturers prefer the objective-type tests because of the ease of scoring, and also the large number of students that they handle, without considering what goes into the construction of such items. Gronlund (1986) perceived that teachers like the ease at which they compose the questions to suit their conviction without due regard to standard testing techniques and that salient points bordering on sampling of course content, preparation of questions, scoring and control of students' responses are not accorded the due attention.

Most lecturers construct both multiple-choice and essay-type of test items. The test as a mechanism of measurement suffers condemnation, if the user is ignorant in the use of test. The need for the ingenuity of lecturers in dealing with both multiple-choice and essay-type tests need not be overemphasized.

Both multiple-choice and essay-type tests have in-built problems and these problems are either compounded or well-contained and reduced to the barest minimum, depending on the competence of the test user. Looking at the important role that tests play in our educational system, it is very essential to know how competent lecturers at the University of Education, Winneba are, in the construction of multiple-choice and essay-type test items, so as to have confidence in the results of these tests.

### **Research Questions**

1. Which principles do lecturers at UEW use when constructing their multiple-choice test items?
2. Which principles do lecturers at UEW use when constructing their essay-type test items?

### **Research Hypotheses**

1. There is no significant difference between UEW lecturers' years of lecturing and applying principles in constructing multiple-choice test items.
2. There is no significant difference between UEW lecturers' years of lecturing and applying principles in constructing essay-type test items.



## **METHODOLOGY**

### **Research Design**

The research design used was the cross-sectional descriptive survey. The major purpose of surveys is to describe the characteristics of a population. In essence, what researchers want to find out is how the members of a population distribute themselves on one or more variables (Fraenkel & Wallen, 2000). According to Doyle (2004), surveys are good for asking people about their perceptions, opinions and ideas though they are less reliable for finding out how people actually behave.

Since the study was to make lecturers report on the procedures they use to construct objective-type and essay-type tests for the researcher to measure them against standard testing techniques, the descriptive survey design was deemed appropriate.

### **Sample and Sampling Technique**

A sample of one hundred and ninety-five (195) full-time lecturers from nineteen departments from five faculties at the Winneba campus of the University of Education, Winneba, participated in the study. A stratified sampling technique was employed for the selection of the lecturers to ensure that each of the nineteen departments is represented. In all, 173 male lecturers and 22 female lecturers were involved in the study. 90 (46.2%) of the respondents have lectured between 1 and 5 years, 63 (32.3%) have lectured between 6 and 10 years, and 42 (21.5%) have lectured more than 10 years.

### **Instrument**

The instrument used for the study was a questionnaire. The questionnaire was designed by the researcher using the scientifically prescribed principles for constructing multiple-choice and essay-type test items. The questionnaire was in three sections. Section A sought information on the principles for constructing multiple-choice item, section B on the principles for constructing essay-type item, while section C tapped the background information of the respondents. The instrument was given to some experts in the area of testing for validation before its administration. The internal consistency of the questionnaire items was determined using the Cronbach coefficient alpha. The co-efficient alpha obtained for the study was 0.897.

### **Data Collection and Analysis**

The data collection was done by the researcher. In all, 195 questionnaires were administered and collected within two weeks. The response rate was hundred percent. The data collected were subjected to frequency counts, mean, and ANOVA tested at 0.05 level of significance.



## RESULTS

**Research Question One:** Which principles do lecturers at UEW use when constructing their multiple-choice test items?

This question sought to find out how frequent lecturers at UEW apply the principles of constructing multiple-choice test items when constructing their multiple-choice test items.

The question was assessed on a 4-point scale level ranging from 1 to 4. Positive items were assessed using 1 = Not at all, 2 = Not often, 3 = Often and 4 = Very often. Negative items were assessed using 1 = very often, 2 = often, 3 = not often and 4 = Not at all.

Means and standard deviations were calculated for each of the principles used. A cut-off point of 3.5 was set for the analysis. This cut-off point is the midpoint of the interval between applying the principle often (3) and very often (4) for positive items, the principle not often (3) and not at all (4) for the negative items. This cut-off point was used because lecturers were expected to apply all the principles at all times.

The overall mean score for the general application of the prescribed principles for constructing multiple-choice item by lecturers at UEW, Winneba campus was 3.30 which is less than the cut-off point. Six (6) out of the fifteen (15) principles used for the study representing 40% had means which were above the cut-off point of 3.5. These principles are;

1. Write items with the important issues in the stem.
2. Use simple vocabulary and sentence structure.
3. Arrange the alternatives/options in a vertical form, rather than horizontal.
4. Make the distractors grammatically correct with respect to the stem.
5. Make the placement of the correct options not to follow a pattern.
6. Not to use **“all the above”** as an option to a single best response item.

Nine (9) out of fifteen (15) principles for constructing multiple-choice items used for the study, representing 60% had means which were below the cut-off point. These principles are;

1. Not constructing items which measure students' opinions.
2. Not copying sentences directly from text books or past test items.
3. Not repeating words and phrases in the options.
4. Avoid overlapping alternatives/ options.
5. Not write item whose answer is dependent on the knowledge of the answer to a previous item.
6. Not write words that can give clues to the best/correct options.
7. Emphasize the word **not** by underlining it or writing it in capital form for negative items.
8. Ensure that the responses/options are about the same length.
9. Write items from a test specification table.

Analysis from the responses showed that, on the whole, lecturers at UEW, Winneba campus, do not frequently apply the prescribed principles for the construction of multiple-choice test items.



**Research Question Two:** Which principles do lecturers at UEW use when constructing their essay-type test items?

The question intended to find out how frequent lecturers at UEW apply the principles of constructing easy-type test items when they are constructing their essay-type test items.

This question was analyzed following the same principle as in question one.

The overall mean score for the general application of the prescribed principles for constructing essay items by lecturers in UEW was 2.96 which is less than the cut-off point.

Only three (3) out of the twelve (12) principles used for the study representing 25% had means which were above the cut-off point. These principles are:

1. Construct items that test important aspect of the learning target.
2. Prepare a scoring key (marking scheme) at the time the item is prepared.
3. Construct items based on the instructional objectives for each content unit.

Nine (9) out of twelve (12) principles for constructing essay items used for the study, representing 75% have means which are below the cut-off point. These principles are:

1. Specify the limit of the problem so that the student knows exactly what to do.
2. Do not construct items that require only recall of information.
3. Do not write items that give room for a broad range of responses.
4. Construct items that require students to apply their knowledge and skill to a new or novel situation.
5. Give preference to a large number of items that require brief answers.
6. Indicate the value of each question.
7. Do not use words such as list, mention, give, and state, as much as possible at the start of essay items.
8. Write items from a test specification table.
9. Do not write items of varied difficulty if students are to select from a given number of items.

Analysis from the responses showed that, on the whole, lecturers at UEW, Winneba campus, do not often apply the prescribed principles for the construction of essay test items.

**Research Hypothesis One:** There is no significant difference between UEW lecturers' years of lecturing and applying principles in constructing multiple-choice test items.

The purpose for this research hypothesis was to find out whether the number of years one has lectured has any influence on ones effort to apply the prescribed principles when constructing multiple-choice test item. A one-way ANOVA test was conducted to test this hypothesis.



Table 1

**One-way ANOVA for the number of years of lecturing experience and applying principles in constructing multiple-choice test items.**

No. of years	No. of respondents	Mean	SD	F	p-value	$\eta^2$
1 – 5	90	49.23	5.23	1.170	0.313	0.012
6 – 10	63	49.13	4.80			
Above 10	42	50.50	4.45			

From Table 1, the ANOVA was not significant at  $\alpha = 0.05$ ,  $F(2,192) = 1.170$ ,  $p = 0.313$ . The strength of the relationship between the number of years one has lectured and applying the principles for constructing multiple-choice item, as assessed by  $\eta^2$ , was weak, with the number of years accounting for 1.2% of the variance of the application of these principles.

**Research Hypothesis Two:** There is no significant difference between UEW lecturers' years of lecturing and applying principles in constructing essay-type test items.

This research question sought to find out whether the number of years one has lectured has influence on applying the scientifically prescribed principles in constructing essay-type test item. A one-way ANOVA test was conducted to test this hypothesis.

Table 2

**One-way ANOVA showing the relationship between years of lecturing experience and level of applying principles in constructing essay-type test items.**

Number of years	Number of respondents	Mean	SD	F	p-value	$\eta^2$
1 – 5	90	35.70	3.469	.383	.682	.004
6 – 10	63	35.59	2.775			
Above 10	42	35.19	2.890			

From Table 2, the main effect was not significant at  $\alpha = 0.05$ ,  $F(2,192) = 0.383$ ,  $p = 0.682$ . The strength of the relationship between the number of years one has lectured and applying the principles for constructing essay item, as assessed by  $\eta^2$ , was weak, with the number of years accounting for 0.4% of the variance of the application of these principles.

**DISCUSSIONS**

It is important to write items from a test specification table (blueprint). Anastasi (1988) cautioned that a test constructor who plunges directly into item writing is likely to produce a lopsided test. It means that some areas will be over-represented while others may remain untouched. The test constructed without a blueprint is likely



to be overloaded with impermanent and less important material. To avoid any lopsidedness and overrepresentations of tests it necessitates test constructors to go back into the statement of objectives made for the term. The objectives will serve as the direction on what specific task is to be measured. To guard any fortuitous imbalances and disproportionate item distribution, test constructors draw up a table of specifications before any items are prepared. Such specifications should begin with an outline of both the instructional objectives of the course, the subject matter to be covered, and the cognitive skills measured – a three-way grid (Gronlund, 1986).

On the construction of essay test items, the study revealed that, few of the respondents do indicate the value of each question on the test while majority do not. This lack of score point for each item might deny the students some vital examination instruction. Hanna (1966), supports the importance of time and score points on the strength that students must be adequately informed of the ground rules in testing. Unfortunately it is a common spectre to find lecturers constructing and scoring essay questions and arbitrarily assigning scores without the student knowing what would be given him or her. Sax (1974), cautioned that, "Tell the student what you expect, how much you expect, and how many points will be accorded each part of the question" (p. 125).

Various criticisms have been leveled against essay tests as providing a small sample of the students' knowledge and potentials, that is, limited items posed usually lacked content validity and focus on spotty sampling of examinees' global knowledge (Ebel, 1979). Lecturers, in the attempt to strike a compromise between increasing the adequacy of the sampling by asking many different questions and also ensuring that questions are probing in nature in the face of time constraints, are compelled to use options. Gronlund (1986), opposed options on the strength that options tend to increase the excellence in calculating the probability of certain items appearing in the test. This might reinforce the habit of preparing material for memorization or even sending them to the examination room. The obvious consequence is that there is the distortion of measurement of students' actual ability which renders the test invalid. Unless students run the same race by answering the same questions, valid comparisons of achievement among them becomes impossible. It is therefore conclusive that options in essay tests need not be encouraged if content matters.

The results from the analysis of research hypotheses one and two indicated that, both hypotheses were retained, meaning the number of years one has lectured do not matter when it comes to applying the prescribed principles for constructing both multiple-choice and essay-type test items. These findings are in line with the findings from a research conducted by Mogno (2003) to determine the profile of the professors' level of appropriateness in test construction in University of Perpetual Help Laguna. The results revealed that there was no significant relationship between the level of appropriateness in test construction and the actual years of teaching experience.



## **CONCLUSION**

Lecturers at UEW, Winneba campus do not put into practice the prescribed principles for test construction they learnt in educational measurement and evaluation (testing). This situation accounted for the non frequent use of the principles by lecturers. This may be as a result of too much work load on the lecturers, which made them not to have enough time to follow rigidly the prescribed principles of test construction. It follows therefore, that, some of the principles of test construction seem to them to be cumbersome and not easy to adopt in the classroom situation. For instance, in following the principles of test construction a lecturer has to establish his objectives, outline his subject matter content and build a table of specifications. These steps require adequate time from the lecturer which most lecturers appear not to have got hence the present state of the art.

It was observed that lecturers could not take advantage of the testing experience to enhance their competence in the art of testing. One would have thought that as lecturers continued to use tests, they would gradually improve upon their competence.

## **RECOMMENDATIONS**

1. Lecturers should at all times put into practice the prescribed principles for test construction to enable them construct quality test items.
2. Practical-oriented periodic in-service training on testing skills should be organized by the university authorities for all lecturers. This should be organized regularly for lecturers to appreciate the importance of applying the prescribed principles of test construction to improve their competency in test construction.
3. Lecturers should review their test items before administering them. They can do this by writing their test items early, read over them or give them to colleagues in the subject area to vet. This will help them to identify some weaknesses in their own test items which will improve their competency in test construction.



## REFERENCES

- Amedahe, F. K. (1989). *Testing practices in secondary schools in the Central Region of Ghana*. Mater's thesis, University of Cape Coast, Cape Coast, Ghana.
- Anastasi, A. (1988). *Psychological testing*. (6<sup>th</sup> ed). New York: Macmillan.
- Carter, K. (1984). Do teachers understand the principles for writing tests? *Journal of Teacher Education*, 35, 385-397.
- Davis, F. B. (1999). Item selection techniques. *Educational Measurement*, 23, 266-328.
- Doyle, E. (2004). *Process of community health education and promotion*. London: Waveland Pr. Inc.
- Ebel, R. L. (1979). *Essentials of educational measurement*. (3<sup>rd</sup> ed). Englewood Cliffs, NJ: Prentice-Hall.
- Fraenkel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education* (4<sup>th</sup> ed.). New York: McGraw-Hill, Inc.
- Gronlund, N. E. (1986). *Constructing achievement tests*. (3<sup>rd</sup> ed). Englewood Cliffs, NJ: Prentice-Hall.
- Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Journal of Educational Measurement: Issues and Practices*, 4 (1), 15-18.
- Hanna, P. L. (1966). Improving classroom testing: Suggestions from the 1965 national council for the social studies yearbook. *Social Education*, 30, 321-324.
- Kusiszyn, T., & Borich, G. (1987). *Educational testing and measurement*. (2<sup>nd</sup> ed.). U.S.A.: Scott. Foresman and Co.
- McKeachie, W. (1994). *Teaching tips: A guidebook for the beginning college teacher*. (9<sup>th</sup> ed). Lexington, MA: Heath.
- Mehrens, W. A., & Lehmann, I. J. (1987). Using teacher-made measurement devices. *NASSP Bulletin*, 71(496), 36-44.



- Mogno, C. C. (2003). The profile of teacher-made test construction of the professors of University of Perpetual Help Laguna. *UPHL Institutional Journal Volume 1, Issue 1*, pp. 48-55.
- Osuala, E. C. (2001). *Introduction to research methodology*. Onitsha: African FED Publishers.
- Quaigrain, A. K. (1992). *Teacher-competence in the use of essay tests: A study of secondary schools in the Western Region of Ghana*. Master's Thesis, University of Cape Coast, Ghana.
- Sax, G. (1974). *Principles of educational measurement and evaluation*. Belmont: Wadsworth Pub. Co. Inc.
- Thorndike, R., Cunningham, G., Thorndike, R. L., & Hagen, E. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York: Macmillan Publishing Company.
- Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, 75 (3), 200-214.