Instructors' Classroom Assessment Practices as a Function of Training Background and Teaching Experience with Particular Reference to Final Examinations

Abatihun Alehegn Sewagegn[*]

Lecturer at Department of Psychology, Debre Markos University, Debre Markos, Ethiopia
*Corresponding author: abatihunalehegn@gmail.com

Abstract The purpose of the study was to analyze the assessment practice of Debre Markos University instructors as a function of training background and teaching experience with reference to final examinations. Furthermore, the study aimed at investigating the perception of students about the classroom assessment practices of their instructors and the perception of teachers on their classroom assessment practices. The subjects of the study were 280 students and 51 instructors from the five colleges. In addition to this 65 final exam papers were collected from the respective colleges. The instruments used for the study were document analysis (i.e., exam papers), and questionnaires. The data collected were analyzed using descriptive techniques and were tested through t-test and one-way ANOVA. The result suggested that there was significant difference in including the general information's of test construction principles, writing good multiple choice and short answer items as a function of training background. It was also observed that significant difference was observed in including the general information's of test construction principles and writing good short answer items as a function of teaching experience. Significant mean difference was not observed on the perception of instructors on the assessment practices. However, significant difference was observed in the perception of students about the assessment practice of instructors across different colleges. On the basis of the findings, conclusions were drawn.

Cite This Article: Abatihun Alehegn Sewagegn, "Instructors' Classroom Assessment Practices as a Function of Training Background and Teaching Experience with Particular Reference to Final Examinations." *American Journal of Educational Research* 1, no. 8 (2013): 300-306. doi: 10.12691/education-1-8-6.

1. Introduction

The Ethiopian ministry of education is committed to providing a high quality education for students at all levels of education [19]. The government has made substantial efforts to widen access, ensure enrolment and improve attendance in universities in its mission to achieve millennium development goals. However, its efforts to improve the quality of education have significantly lagged behind. This is evident not only from the poor achievement levels of students, but also from the poor quality of assessment taking place in universities. To achieve the mission and to reach at the goal there should be quality assessment of students' academic work. Quality assessment is a very essential element in the provision of quality education. This is because assessment provides a foundation for making sound evaluative judgments about students' learning progress in particular and about the effectiveness of the whole education system in general [17]. Therefore, assessment is an important element of the teaching learning process whatever the level of education

is. The assessment practice of an educational institution directly affects the quality of education in that institution.

The word "assessment" has taken on a variety of meanings within higher education. As it is defined by [16] assessment is the systematic collection and analysis of information to improve student learning. As it is defined by [5] classroom assessment is a continual activity for teachers to improve the quality of instruction and motivate students to learn. Airasian in [1] noted that classroom assessment is an assessment which is implemented or conducted by instructors to check the achievement of the learning outcomes and students' understanding of a certain lesson, topic, course or program. Every day in every classroom, instructors assess and make decisions about instructional success and pupils' learning. Airasian [1] has defined assessment as the process of gathering, interpreting, and synthesizing information to aid decision making in the classroom. This implies that the intent of collecting assessment information by instructors is to help them make decisions about their pupils' learning, success of the ongoing instruction and the social climate of their classroom.

Instructors use various methods of assessment to determine students' progress in learning and their academic achievement. According to [14] assessment methods refer to any of a variety of procedures used to obtain information about student performance. In the classroom, instructors usually use written test and performance assessment or authentic assessment such as observation and questioning to obtain information about students' learning [1]. Classroom assessment activities include constructing written test and performance assessment, grading, interpreting test score results, giving feedback on assessment results and using test results to make decision. When using written test and performance assessment, teachers need to be aware of the strengths and weaknesses for each assessment technique so as to select an appropriate technique to assess students' learning [21].

The purpose of gathering assessment information is thus to help teachers make decisions in the classroom; and assessment is not an end by itself, but a means to another end, namely, good decision making [12]. In simple terms, good assessment information is any information that helps teachers to make accurate decisions in their classrooms. To accomplish these activities, teachers or instructors may use various techniques of assessment. From these techniques, tests and exams are used most commonly in different grade levels; that is, starting from lower grade levels up to higher institutions [3].

Having the knowledge and skill of test development and principles of test construction is very crucial which is expected from all teachers or instructors. Before administering tests, the instructor has to prepare tests based on a table of specification. Regarding this [23] described that table of specification or test blue print helps the teacher to ensure that only those objectives actually pursued in instruction will be measured; that each objective will receive the appropriate weight and relative emphasis in the test; that by using subdivisions based on content and behavior. Strict observance of this principle, in addition to maintaining proportionality among contents and instructional objectives, can help in keeping the quality of items in post test analysis [20]. Moreover, instructors should follow the test construction guidelines suggested by educationalists. Before deciding to construct a test/exam, one needs to know what information is required, how quickly it is needed, and the likely actions that are to be taken according to the results on a test or exam [11]. When the tests or exams are developed it should be on the bases of test construction principles. As it is mentioned by [18], there should be general or basic information and the principles of multiple choice items, true false item, matching item, short answer and completion item and essay or work out item formats should be considered when we develop tests/exams.

Research concerning teacher-constructed tests has found that teachers lack understanding of measurement [18]. Another research has also shown that teachers lack sufficient training in test development, fail to analyze tests, do not establish reliability or validity, do not use a test blueprint, weight all content equally, rarely test above the basic knowledge level, and use tests with grammatical and spelling errors [9]. Technically their tests are simplistic and depend upon short answer, true-false, and other easily prepared items. Their multiple-choice items often have serious flaws-especially in distracters [18].

The more the teacher has obtained trainings on assessment, the more likely he/she would practice it in the classroom. Accordingly, if the training obtained is contextualized and if teachers are committed to practice what is obtained in trainings, trained teachers have high probability to implement assessment than untrained ones. In support of this view, Capper in [4] stated that teachers who are not well trained may have difficulty in using different approaches to assessment where as teachers who received appropriate training and have time to develop assessment would be able to develop pedagogically and technically more creative approaches.

Researchers have attempted to investigate teachers' perceptions of assessment in many different ways [6]. Chester and Quilter believed that studying teachers' perceptions of assessment is important in the sense that it provides an indication of how different forms of assessment are being used or misused and what could be done to improve the situation. They found that teachers' perceptions of classroom assessment affected their classroom assessment practices.

Teachers are responsible for assessing students' learning. All teachers must have assessment skills to implement the assessment strategies effectively. Teachers used various techniques in assessment even though they may not be given appropriate training on certain aspects of classroom assessment [15]. However, studies showed that most teachers lack effective assessment knowledge and skills when evaluating students' academic achievement [7]. Currently, not much is known about Ethiopian teachers' assessment practices and assessment skills. This study is carried out in an effort to identify teachers' assessment practice to enable appropriate actions be taken to enhance teachers' assessment skills in relation to the development of final examinations.

2. Methodology

2.1. Design of the Study

The purpose of this study was to analyze instructors' knowledge on classroom assessment and its practice as a function of the professional training taken in college days and teaching experience with particular reference to final examinations. To attain the desired objectives, the study followed a survey design and a quantitative approach in its nature.

2.2. Setting

The setting selected to conduct the research was Debre Markos University. This is chosen as the setting for this study because it is the researcher's work place. And the researcher observed problems in relation to the issue raised. Debre Markos University is a public university located in the town of Debre Markos, Ethiopia. The university is located two kilometers from the central square of the town. It is 300km far from Addis Ababa, the capital city of Ethiopia. It has a latitude and longitude of 10°20′N 37°43′E/ 10.333°N 37.717°E Coordinates and an elevation of 2,446 meters.

2.3. Participants of the Study

Instructors and students enrolled in Debre Markos University (DMU) were the participants of this study. Besides, final examination papers were included as data sources.

2.4. Sampling Techniques

Instructors and students were selected from the target colleges using simple random sampling technique. The actual number of participants from different departments of each college was selected on random basis. In 2010/2011 there were 1502 second (SSH, CBE, NCS and Agriculture) year and third year Technology students and 241 teachers. From these, 300 students and 65 instructors were selected. But 280 students and 51 teachers returned the questionnaires. That is, the response rate was 93% for students and 78.5% for teachers including the questionnaires which were completed correctly. In addition to this, 65 final exam papers prepared and administered in 2010/11 by the selected colleges were randomly taken from the participant colleges.

2.5. Tools and Procedures of Data Collection

Close ended questionnaires and documents (final exam papers) were employed to gather data pertinent to the study. The questionnaires were administered for students who were enrolled in different departments of the colleges in focus to gather data on their perception of how instructors practice classroom assessment and for instructors to know their perception about the classroom assessment they used. The questionnaires for both groups were 5-point likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A total of 22 perception items for instructors and 20 perception items for students were developed and used in this study. The minimum and maximum possible score for the instructor was 22 and 110 respectively. For the students, the minimum and maximum possible score for was 20 and 100 respectively.

Before collecting the main data, a pilot test was conducted to ensure reliability and validity of the questionnaires. In other words, the aim of the pilot testing was to find out ambiguities and omissions in each item of the questionnaire in order to avoid misunderstanding. Copies of the instruments were given to three educational measurement and psychology and one TEFL (Teaching English as Foreign Language) lecturers at Debre Markos University. Using the relevant comments and suggestions from these individuals, some corrections were made.

After the refinement of the instruments, pilot testing was carried out to test the reliability of the instruments taking 50 students and 11 lecturers those who will not be participated for the main study. The questionnaires were filled out properly and collected.

After the pilot study was conducted, the reliability was measured with the help of the internal consistency item analysis. The reliability for the pilot study and main study using Cronbach alpha was 0.61 and 0.71 respectively.

Based on the pilot study's result, content validity of the instrument was checked as follows:
- the items were thoroughly inspected for relevance and clarity;
- the content validity of the instruments, omissions, vague items and terminology were improved and made to measure what they were supposed to measure.

Document analysis (final exam papers administered in 2010/2011) was used as a source of data to know instructors' knowledge and skill about the basic principles of test construction and on their degree of practicing these principles when developing test items in relation to their training background and teaching experience. When the researcher collected the exam papers, he gathered data about the training background and teaching experience of the instructor who developed the exam. To evaluate the exam papers the researcher developed a rating scale having three levels (yes, to some extent and no) and a value of two, one and zero is given for each level respectively. If the instructor developed the test items based on the principle of test construction a value of two is given, if it deviates from the principle a value of zero is given and if it is in between, a value of one is given. A total of 56 items for the check list were used in the study.

2.6. Data Analysis Technique

Before analyzing the collected data effective data entry tasks were done and the analysis tasks were performed with the help of SPSS Windows 16.0. To analyze the data both descriptive and inferential techniques were used. Descriptively, the collected data were analyzed using percentages, mean values and deviation scores. In addition, inferential data analysis techniques namely independent t-test and ANOVA (one-way classification) were used for comparing mean differences between or among different groups considered in the study. The α (alpha) value for test of significance is set at 0.05 level.

3. Results

Under this section, instructors' background information, analysis of instructors' knowledge on classroom assessment as a function of the professional training and teaching experience with reference to final examinations and their perception about their assessment practices are presented. In addition, perception of students about the classroom assessment practices of their instructors' is also presented for the purpose of triangulation.

3.1. Instructors' Background Information

From the total sample of the respondents (i.e. instructors) 68.63% of them have not taken pedagogy courses during their stay in colleges/universities. Only 31.37% of the respondents have pedagogy background. However, most teachers that is, 70.59% have taken in-service pedagogical trainings. Regarding their teaching experience, half of the respondents (50.98%) have a teaching experience less than two years in higher institutions. On the other hand, 66.67% of the respondents have no teaching experience in general.

3.2. Analysis of Exam Papers based on Test Construction Principles as a Function of

Training Background and Teaching Experience

The collected final exam papers were analyzed based on test construction principles as a function of training background and teaching experience of instructors. This is to know whether there is a variation in instructors' knowledge on test construction principles on the bases of training background and teaching experience.

Table 1. Analysis of Exam Papers based on Test Construction Principles as a Function of Training Background

Item Formats	Training Background	N	Mean	Standard Deviation	df	t-value	p-value
Basic/General information	Pedagogy	24	18.96	2.56	63	3.78*	0.000
	Non-pedagogy	41	15.02	4.70			
	Total	65	16.48	4.45			
Multiple Choice	Pedagogy	20	18.95	1.90	38	2.82*	0.008
	Non-pedagogy	20	17.45	1.43			
	Total	40	18.20	1.83			
True False	Pedagogy	11	17.55	0.82	22	1.92	0.068
	Non-pedagogy	13	16.54	1.56			
	Total	24	17.0	1.35			
Matching	Pedagogy	13	9.15	2.44	25	2.35*	0.027
	Non-pedagogy	14	7.07	2.16			
	Total	27	8.07	2.50			
Short Answer	Pedagogy	12	11.83	0.39	18	5.41*	0.000
	Non-pedagogy	8	9.00	1.77			
	Total	20	10.70	1.81			
Essay/Workout	Pedagogy	15	7.60	0.91	53	0.59	0.560
	Non-pedagogy	40	7.43	1.01			
	Total	55	7.47	0.98			

* = there is a significant difference, $P < 0.05$

The result of this study revealed that there is a statistically significant difference in including the basic/general information of test construction principles in the preparation of exam papers across training background. That is, teachers who have pedagogy background used the basic/general information's of test construction principles than those teachers who have no pedagogy background ($t_{(63)} = 3.78$; $p < 0.05$). There is also significant mean difference in applying test construction principles for multiple choice, matching and short answer item formats between the two groups (i.e. Pedagogy and non–pedagogy). The t-test result was for multiple choice $t_{(38)} =$ 2.82, for matching $t_{(25)} = 2.35$ and for short answer $t_{(18)} = 5.41$; $p < 0.05$. But there is no significant mean difference in the application of test construction principles between instructors who have pedagogy and non-pedagogy background for the item formats true false and essay or workout.

3.3. Analysis of Exam Papers based on Test Construction Principles as a Function of Teaching Experience

Table 2. Summary of one-way ANOVA for the variations of general test construction principles across teaching experience

Item Formats	Variation	Sum of Squares(SS)	Df	Mean Squares(MS)	F	Sig.
Basic information	Between Group	412.97	3	137.66	9.79*	0.000
	Within Group	857.24	61	14.053		
	Total	1270.21	64			
Multiple Choice	Between Group	15.733	3	5.244	1.65	0.196
	Within Group	114.667	36	3.185		
	Total	130.400	39			
True False	Between Group	11.439	3	3.813	2.50	0.089
	Within Group	30.561	20	1.528		
	Total	42.000	23			
Matching	Between Group	38.302	3	12.767	2.38	0.096
	Within Group	123.550	23	5.372		
	Total	161.852	26			
Short Answer	Between Group	36.764	3	12.255	7.71*	0.002
	Within Group	25.436	16	1.590		
	Total	62.200	19			
Essay/Workout	Between Group	4.470	3	1.490	1.61	0.199
	Within Group	47.239	51	0.926		
	Total	51.709	54			

* = there is significant difference, $p < 0.05$

In including the basic information of test construction principles as a function of teaching experience, significant mean difference was observed (i.e., $F_{(3, 61)} = 9.97$, $p < 0.05$). The result of multiple comparison shows that experienced teachers considered the basic information while they write tests or exams than non-experienced teachers. In contrast, there was no significant mean difference in the application of test construction principles in writing essay or work out item formats (i.e., $F_{(3, 51)} = 1.61$; $p > 0.05$).

Table 3. Multiple comparison to see the variations by using Tukey Method

Item Format	Teaching Experience		Mean Difference (A-B)	Sig.
	A	B		
Basic/General Information	Less than 2 year	2 - 4 year	-5.58*	0.007
		4 - 6 year	-5.20*	0.004
		above 6 year	-5.51*	0.000
	2 - 4 year	4 - 6 year	0.39	0.997
		above 6 year	0.07	1.000
	4 - 6 year	above 6 year	-0.31	0.996

*= there is significant difference, $p < 0.05$
A and B indicates the categories/levels of teaching experience

3.4. Perception of Instructors' on Classroom Assessment as a Function of Training Background and Teaching Experience

Instructors were asked to respond for the same questions to see their perception about their classroom assessment practices. The result which is obtained from the data was presented in two categories. The first one was the perception of instructors' on classroom assessment practices as a function of training background. In this regard, there was no significant mean difference in the perception of instructors on classroom assessment practices across training background (i.e., $t_{(df = 49)} = 0.14$; $p > 0.05$). This shows that those instructors who have pedagogy background and those who did not take pedagogy courses during their stay in colleges or universities have similar perception about classroom assessment, even if the mean score of pedagogy group (M = 69.31) is greater than non-pedagogy group (M = 69.00) by fraction.

Table 4. Perception of Instructors' on Classroom Assessment Practice as a Function of Training Background

Training Background	N	Mean	Standard Deviation	df	t-value	p-value
Pedagogy	16	69.31	6.74	49	0.14	0.893
Non-pedagogy	35	69.00	9.61			
Total	51					

Note: The maximum possible score was 110 and the minimum was 22

The second category was the perception of instructors on classroom assessment practices as a function of teaching experience. In this regard, the result of descriptive statistics (mean value) indicates that the perception of instructors on their classroom assessment varies across different teaching experiences. The higher the mean value indicates the better perception of instructors on classroom assessment practices. That is, the mean values increases as the teaching experience increases. To see whether there is a significant mean difference between the four age groups one-way ANOVA was used and, it is indicated that there is no significant mean difference in the four categories of teaching experiences (i.e., $F_{(3, 47)} = 0.87$, $P > 0.05$).

3.5. Perception of Students on Instructors' Classroom Assessment Practices

Table 5. Perception of Students on Instructors' Classroom Assessment across Colleges

Colleges	N	Mean	Standard deviation
SSH	61	79.16	8.72
NCS	41	79.66	14.69
Agriculture	61	72.43	11.55
CBE	59	78.54	11.77
Technology	58	68.19	15.50
Total	280	75.36	13.21

Note: The maximum possible score was 100 and the minimum was 20
Key: SSH- Social Sciences and Humanities**, NCS** – Natural and Computational Sciences
CBE – College of Business and Economics

The students were asked to respond for same questions to see their perception on instructors' classroom assessment practices across different colleges. As the result of the descriptive statistics indicates, students of the three colleges (i.e., SSH, NCS and CBE) have almost similar mean score, that is they have similar perception about the assessment practice of their instructors. The mean score value was 79.16, 79.65 and 78.54 respectively. The perception of students about the classroom assessment practice of instructors of the two colleges (Agriculture and Technology) was less than the other three colleges. The mean score value was 72.42 and 68.19 respectively. The result of one way-ANOVA shows that there is a significant difference in the perception of students about the classroom assessment practice of instructors across different colleges (i.e., $F_{(4, 275)} = 9.20$, $p < 0.05$). In the Tukey post-hoc analysis (multiple comparison), the significant mean difference in perception of students about instructors classroom assessment practices was observed between SSH and Agriculture, SHH and Technology, NCS and Agriculture, NCS and Technology, CBE and Technology.

Table 6. Summary of one-way ANOVA for the perception of students about instructors' assessment practices across colleges

Sources of Variation	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5744.787	4	1436.197	9.20*	0.000
Within Groups	42936.056	275	156.131		
Total	48680.843	279			

*= there is significant difference, $p < 0.05$

4. Discussion

This study attempted to analyze instructors' knowledge on classroom assessment as a function of the professional training taken in college days and teaching experience with particular reference to final examinations. Thus, the discussion is based on the research question raised and the results obtained.

To see whether there is a variation or not in instructors' knowledge on test construction principles as a function of training background and teaching experience, final exam papers were collected and analyzed. The training back ground and teaching experience of instructors has its own contribution for their assessment practices. The result of this revealed that there is a statistically significant difference in including the basic information of test construction principles in the preparation of exam papers across training background. That is, instructors who have pedagogy background used the basic information of test construction principles than those teachers who have no pedagogy background ($t_{(63)}$ = 3.76; p < 0.05). There is also significant mean difference in applying test construction principles for multiple choice, matching and short answer item formats between the two groups (i.e. Pedagogy and non–pedagogy). The t-test result was for multiple choice $t_{(38)}$ = 2.86, for matching $t_{(25)}$ = 2.35 and for short answer $t_{(18)}$ = 5.41; p < 0.05. But there is no significant mean difference in the application of test construction principles between instructors who have pedagogy and non-pedagogy background for the item formats true false and essay or workout. In consistent with this study, a study conducted by [8] showed that, there was a statistically significant difference between teacher from teacher education and teacher from other faculty in the use of multiple choice test items. However, there no statistically significant differences among teacher educators and other groups of higher education faculty in their uses of short answer exams.

In including the basic information of test construction principles as a function of teaching experience, significant mean difference was observed (i.e., $F_{(3, 61)}$ = 9.97, p < 0.05). The result of multiple comparison shows that experienced teachers considered the basic information while they write tests or exams than non-experienced teachers. On the other hand, there was no significant mean difference in the application of test construction principles in writing essay or work out item formats (i.e., $F_{(3, 51)}$ = 1.61; p > 0.05). A study result of [2] revealed that there were statistically significant differences across teaching experience in analyzing test items, communicating assessment results, writing test items, using performance assessment, and grading. Scheffe's test indicated that teachers with more than 10 years of teaching experience reported on average higher levels of self-perceived skillfulness in analyzing test items, communicating assessment results, writing test items, using performance assessment, and grading than both teachers with 1 to 5 years of teaching experience and teachers with 6 to 10 years of teaching experience. But [10] study result shows that teachers with fewer than 8 years of experience developed items with better overall quality than those who had more experience. The less experienced teachers significantly outperformed their more experienced peers on seven of the quality factors studied: spelling, distractors, key accuracy, usability, validity, taxonomy, and overall quality. Although these results disagree with previous studies of instructors' assessment practices the current study results highlight the importance of teaching experience in that, assessment skills might best be mastered through practice and classroom experience.

In the perception of instructors on classroom assessment practices as a function of training background, significant mean difference was not observed (i.e., $t_{(df = 49)}$ = 0.14; p > 0.05). This shows that those instructors who have pedagogy background and those who did not take pedagogy courses during their stay in colleges or universities have similar perception about classroom assessment, even if the mean score of pedagogy group (M = 69.31) is greater than non-pedagogy group (M = 69.00) by fraction. This finding is consistent with [22] which reported that teacher's education program did not seem to contribute much to teachers' perceptions of classroom assessment. In relation to the perception of instructors on classroom assessment practices as a function of teaching experience, significant mean difference was not observed in the four categories of teaching experiences (i.e., $F_{(3, 47)}$ = 0.87, P > 0.05). This finding is also consistent with the findings of [22] who reported that teacher's experience did not seem to contribute much to teachers' perceptions of classroom assessment.

Finally, for the purpose of triangulation, looking students' perception on instructors' classroom assessment practices across different colleges was important. As the result of the descriptive statistics indicates, students of the three colleges (i.e., SSH, NCS and CBE) have almost similar mean score, that is they have similar perception about the assessment practice of their instructors. The mean score value was 79.16, 79.65 and 78.54 respectively. The perception of students about the classroom assessment practice of instructors of the two colleges (Agriculture and Technology) was less than the first three colleges. The mean score value was 72.42 and 68.19 respectively. The result of one way-ANOVA shows that there is a significant difference in the perception of students about the classroom assessment practice of instructors across different colleges (i.e., $F_{(4, 275)}$ = 9.20, p < 0.05). In the post-hoc analysis, the significant mean difference in perception of students about instructors classroom assessment practices was observed between SSH and Agriculture, SHH and Technology, NCS and Agriculture, NCS and Technology, CBE and Technology. This is supported by a study conducted by [12] in the University of Massachusetts. He observed that students' perception of classroom assessment differs significantly across the three departments (Social science, Natural science and Language).

5. Conclusions and Recommendations

Based on the above findings, the following conclusions are presented.

1. Statistically significant difference was observed in including the basic or general information's of test construction principles and writing good multiple choice, matching and short answer items while constructing exam items across training background. The result was in favor of teachers who have

pedagogy background. Therefore, it can be concluded that training background may have an influence in the application of test construction principles in the preparation of exam items.

2. Statistically significant mean difference was observed in including the basic information of test construction principles as a function of teaching experience. The result is in favor of experienced teachers. There is no significant mean difference in the application of test construction principles for essay or work out item formats. From this we can understand that teaching experience has no significant effect in the application of essay test construction.

3. Statistically significant mean difference was not observed in the perception of instructors on classroom assessment practices across training background and teaching experiences. From this we can conclude that all instructors have the same perception on their classroom assessment practices.

4. Statistically significant mean difference was observed in the perception of students about the classroom assessment practice of instructors across different colleges. The difference was observed between SSH and Agriculture, SHH and Technology, NCS and Agriculture, NCS and Technology, CBE and Technology. The result was in favor of SSH, NCS and CBE.

On the basis of the above, findings and conclusions, the following recommendations were made:

1. Continuous training should be given for the existing academic staff and for newly employed instructors about assessment specifically about test construction or preparation.

2. Specifically, instructors should apply or use test construction principles while they are developing tests and exams.

3. Instructors should use multiple item formats to assess students understanding in a better way.

Acknowledgement

Abbreviations

CBE	College of Business and Economics
MOE	Ministry of Education
NCS	Natural and Computational Sciences
SSH	Social Sciences and Humanities

References

[1] Airasian, P.W. (2002). *Classroom Assessment (Revised Ed.).*New York: McGraw-Hill, Inc.

[2] Alkharusi, H. (2010). Teachers' assessment practices and Students' perceptions of the classroom assessment environment. *World Journal on Educational Technology, 2, 27-41.*

[3] Angela,T.A. & Cross,P.K.(1993). *Classroom Assessment Techniques (2nd Ed.).* Sanfrancisco: Jossey Bass.

[4] Animaw T. (2009). "The Status, Gaps, and challenges of implementing Continuous Assessment: The Case of Second Cycle Primary Schools in Debre Markos Town." (Unpublished, MA Thesis). Addis Ababa University, Ethiopia.

[5] Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10, 161-180.*

[6] Chester, C., & Quilter, S.M. (1998). Inservice teachers' perceptions of educational assessment. *Journal for Research in mathematics Education, 33(2), 210-236.*

[7] Cizek, G., Fitzgerald, S., & Rachor, R. (1996). Teachers' Assessment Practices: Preparation, isolation, and the kitchen sink. *Educational Assessment. 3(2), 159-179.*

[8] Goubeaud, K. & Yan, W. (2004). Teacher educators' teaching methods, assessments, and grading: A comparison of higher education faculty's instructional practices, *The Teacher Educator, 40(1), 1-16.*

[9] 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.*

[10] Haynie, W. J. (*1992).*Post Hoc Analysis of Test Items Written by Technology Education Teachers. *Journal of Technology Education, 4(1), 26-38.*

[11] Izard, J. (2005). *Over view of test construction.* Paris: International Institute for Educational Planning.

[12] Javid, M. (2009). *Assessment Practices: Students' and teachers' Perceptions of Classroom Assessment*

[13] Lin, R.L., & Gronlund, E.N. (2005). *Measurement and assessment in teaching (8th ed.).* India: Baba.

[14] Linn, R. & Miller, M. (2005). *Measurement and Assessment in Teaching (9th Ed.).* Upper Saddle River NJ: Merrill-Prentice Hall.

[15] Marso, R.N. & Pigge, F.L. (1988). An analysis of teacher-made tests and testing: *Classroom resources, guidelines, and practices.* Paper presented at the annual meeting of the Mid-Western Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 291 781).

[16] Martha, L., Kathryn D. & Mya P. (2001). Handbook of Program-Based Review and Assessment: *Tools and Techniques for Program Improvement.* University of Massachusetts, Amherst.

[17] McMillan, J.M (2004). *Fundamental Assessment Principles for Teachers and School Administrators.* In Cauley, K.M., Linder, F., and McMillan, J.H. (Eds.). *Educational Psychology (Pp. 176-179).* USA: McGraw-Hill/Dushkin, A Division of McGraw-Hill Comp.

[18] Mehrens, W.A. & Lehmann, I.J. (1991). *Measurement and evaluation in education and psychology.* (3rd Ed.) New York: Holt, Rinehart and Winston.

[19] MOE (2011). *Education Statistics Annual Abstract.* Addis Ababa, Ethiopia.

[20] Nitko, A.J. (1983). *Educational Tests and Measurement: An Introduction.* New York: Harcourt Brace Jovnovich Press.

[21] Stiggins, R.J. (1992). *Student-Involved Classroom Assessment (3rd Ed.).* Columbus, OH: Merrill, an imprint of Prentice Hall.

[22] Susuwele-Banda W.J. (2005).*Classroom Assessment in Malawi: Teachers' Perceptions and Practices in Mathematics*: PhD Dissertation: Blacksburg, Virginia.

[23] Yalew, E. (2006). *Educational Measurement and Evaluation: Course Module for Distance Education Students.* Addis Ababa: Artistic Printing Press.