



THE KENYA NATIONAL EXAMINATIONS COUNCIL

THEME: VALIDITY, VALIDITY, VALIDITY

**SUB -THEME: IMPROVING TEST DEVELOPMENT PROCEDURES TO IMPROVE
VALIDITY**

**Title : Factors influencing Validity of Classroom Test Construction and Challenges
experienced by trainers in Early Childhood Development and Education Programmes,
Nairobi County, Kenya**

Presented by

ELIZABETH A OBADE

eobade@kneec.ac.ke

**During the 41st -International association for Educational Assessment (IAEA)- Conference
held in Lawrence Kansas, USA 11th and 15th October 2015**

ABSTRACT

The study sought to examine the factors influencing the validity of classroom test construction and challenges experienced by trainers in Early Childhood development and Education programmes, Nairobi County, Kenya. The specific objectives of the study were four fold namely; to explore the content and face validity of ECDE classroom examinations; trainers experience in test construction; use of table of specification and bloom's taxonomy in classroom tests to improve validity and to examine the challenges facing trainers in classroom test construction .The study was guided by the cognitive learning theory. Classroom tests can be designed by the teacher to determine or monitor progress of the learners in a classroom. Most trainers have limited assessment training and even less time to develop and evaluate their own tests. Trainers routinely develop tests without considering validity issues. The Kenya National Examinations Council is mandated to assess the learning objectives of ECDE trainees at Certificate and Diploma level and award Certificates to successful candidates. However, the average performance of some candidates in National Examinations alleged by Chief Examiners is attributed to poor syllabus coverage. The Study adopted a multiple case study research design and was conducted in 2 ECDE training institutions. Purposive sampling was used to select a sample of 15 trainers. The main research instruments were Questionnaires, Structured interview and Document analysis. Data was analyzed using descriptive statistics and the information presented in figures and percentages in tables. The study revealed that the continuous assessment tests administered to the ECDE trainees had low content and low face validity. Despite their experience in training of ECDE trainees, they still had insufficient skills in test construction, syllabus, inadequate resources and test quality related challenges affected test validity. Intensive item writing training recommended.

Key words: Validity, Bloom's taxonomy, test construction, Early Childhood Development and Education

Introduction and background

The provision of quality Education and training to all Kenyans is fundamental to the success of the Government's overall development strategy .Kenya vision 2030 articulates the development of a middle income country in which all citizens will have embraced entrepreneurship ,able to engage in lifelong learning (Republic of Kenya,2007).

Early Childhood Development and Education (ECDE) is the bedrock and foundation of all learning in the life of any child. With the advent of the implementation of the constitution the devolution process heralded the devolution of preschool. The Fourth Schedule of the Constitution places pre-primary Education and childcare facilities under the County Governments (Constitution of Kenya, 2010). ECDE teacher training is done at the National and District systems. The training is offered for Certificate and Diploma levels by the District Centre's for early childhood and also the registered Private training institutions by the Ministry of Education, Science and Technology (MoEST). On completion of training the successful candidates are awarded Certificates (KNEC, 2007).The Curriculum for training ECDE teachers was developed by the Kenya institute of Education(KIE, 2006).At the core of ECDE curriculum is the endeavor to address the total needs of children(NACECE,1999).The aim of ECDE is to develop the whole personality encompassing physical, social, intellectual, spiritual and cultural that provides holistic education, particularly in the formative stage of the child(KIE,2006).

Validity is an important quality to consider when constructing or selecting a test. Test development process begins with Curriculum development (KIE, 2006).The ECDE curriculum constitutes 23 units for Diploma and 24 units for certificate level. The units have been merged by the Kenya National Examinations Council (KNEC) to constitute examination papers. The units for Certificate examinations include;

- Child development& psychology 7011
- Foundations and management in ECDE 7012
- Child Health, nutrition, guidance & counseling 7013
- Curriculum & methodology 7014
- English and Kiswahili 7015
- Curriculum activities 7016
- Teaching practice 7017

The units for Diploma level are;

- Foundations of ECDE and curriculum development 6011
- Psychology, personality development, guidance and counseling 6012
- Child growth and development 6013
- Child health, nutrition and rights 6014
- Management of ECDE, community development and research 6015
- Methodology and curriculum activities 6016
- Research Project 6017
- Teaching practice 6018

One of the most basic and difficult task that teachers face in their work is the process of assessment. Classroom assessment includes all the process involved in making decisions about students learning progress. It includes the observation of student’s written work, their answers to questions in class and performance on teacher-made and standardized tests. According to Koyalik(2002),

classroom assessment provides valuable information that allows teachers to adapt instructional procedures to the learning needs of their students.

Gibbs (1994) defines validity as “the extent to which an assessment measures what it purports to measure. If the assessment does not measure what it is designed to measure then its use is misleading”. Classroom assessment is an integral part of teaching (Chase, 1999, Popham, 2002; Trice 2000; Ward & Murray – Ward 1999) and may take more than one third of teachers’ professional time (Stiggins, 1991). Most classroom tests involve tests that teachers have constructed themselves. Teachers regularly use tests they have constructed themselves (Boothroyd, Mc Morris & Pruzek, 1992; Marso & Pigge, 1988). Most teachers believe that they need strong measurement skills (Wise et al, 1991). While some report that they are confident to provide valid and reliable tests (Oescher & Kirby 1990). Others report a level of discomfort with the quality of their own tests (Stiggins, 1991) or believe that training was inadequate (Wise et al, 1991). Most state certification systems and half of all teacher education programs have no assessment course requirement or even an explicit requirement that teachers have received training in assessment (Boothroyd et al: Stiggins, 1991; Trice, 2000). In addition, teachers have historically received little or no training or support after certification (Herman & Dorr- Bremme, 1984). The formal assessment training teachers do receive often focuses on large scale administration and standardized test score interpretation rather than on the test construction strategies or item writing rules that teachers need (Stiggins, 1991). A quality teacher made tests should follow valid item-writing rules. However, empirical studies establishing the validity of item-writing rules are in short supply and often inconclusive, and “ item writing rules are based primarily on common sense and the conventional wisdom of test experts” (Millan & Greene, 1993).

Hermanowicz (1980) argued that a major component in teacher pre service education ought to be training in the development of measurement and evaluation proficiencies. Practicing teachers themselves report that assessment of learners is a key element in the instructional process, and measurement specialists such as Stiggins, Conklin and Bridgeford (1986) and Dorr – Bremme (1983) have provided information describing how classroom teachers do integrate testing within day to day instructional practices. The major concerns about teachers having little or no pre service teacher training in testing and whether such training is appropriate, several researchers have reported that in- service teacher training in testing is almost non-existent (Dorr-Bremme, 1983; Gullickson, 1984), and Marso and Pigge (1988) found that neither teachers’ ratings of their own testing proficiencies nor the quality of their teacher-made tests improved with the teachers years of teaching experience. Further, what little in-service training teachers receive in testing and evaluation is commonly perceived by teachers as not being helpful.

Tests can be important parts of the teaching and learning process if they are integrated into daily classroom teaching and constructed to be part of the teaching learning process not just the culminating event. They allow the students to see their own progress and allow teachers to make adjustment to their instruction on a daily basis. But one of the most serious problems of evaluation is the fact that a primary means of assessment the test itself is often flawed or misused (Hills, 1991).The measurement of achievement is a critically important part of efforts to improve student learning. It is imperative that these tests be technically as sound as possible. Classroom teachers know the learning needs of their own students and the content in which the students explore and apply specific skills and concepts. Therefore, they are the people in the best position to develop fair and effective tests for their students. Well constructed and fairly administered teacher –made

tests can provide evidence of quality learning and teaching. The process of assessment and evaluation doesn't start and stop with a single test.

Constructing a good teacher-made test is time consuming and difficult. It is hard to understand why something so essential to the learning process have been virtually ignored in teacher pre service or in-service training. Veteran teachers have relied on commercially made tests in workbooks or on their own often inadequate teacher-made tests for most of their evaluation. Teachers have often neglected addressing this aspect of instruction because they were not trained to write effective tests and few administrators would offer guidance.

One of the problems with teacher-made tests is their emphasis on lower-level thinking. A study by (Fleming and Chambers, 1983) examined over 300 teacher-made paper and pencil tests. The results of the study found that teachers appeared to need training in how to:

1. Plan and write longer tests.
2. Write unambiguous paper and pencil test item; and
3. Measure skills beyond recall of facts (Stiggins, 1985).

Literature review

A test is a question or a task or a series of such designed to elicit some predetermined behavior from the person being tested. The word test implies a paper and pencil instrument, administered under pre-specified conditions that are consistent across students. Its traditional usage by teachers has implied a written series of tasks to which students responded in writing for example essay or short-answer items or marked their response choices with a pen or pencil for example true-false, multiple choice or matching items (Gallagher, 1998).

Test items are the most critical part of any test. Unfortunately, writing understandable, useful, appropriate and valid test questions is very difficult. It is essential that you make sure that test items assess the most important and relevant concepts and skills you have taught. In addition, your test items should be consistent with the instructional strategies you used to help your students learn (Hogan, 2007). Although it is not appropriate to “teach to the test” you can consider whether it is appropriate to match the content and format of test questions on your teacher-made tests with the high-stakes tests your students will take. That way you can assess students performance while at the same time helping students become more familiar with the conditions they will encounter when taking high-stakes tests (Savage, Savage & Armstrong, 2006).

Classroom tests provide teachers with essential information used to make decisions about instruction and student grades. A table of specification (TOS) can be used to help teachers frame the decision making process of test construction and improve validity of teachers’ evaluations based on tests constructed for classroom use(Notar, Zuelke, Wilson & Yunker, 2004).

Frequently there is a real and perceived mismatch between the content examined in class and the material assessed on an end of chapter/unit test. This lack of coherence leads to test that fails to provide evidence from which teachers can make valid judgments about students’ progress (Brookhart, 1999). One strategy teachers can use to mitigate this problem is to develop a Table of Specifications (TOS). A TOS, sometimes called a test blue print, is a table that helps teachers align objectives, instruction and assessment (Notar, Zuelke, Wilson & Yunker, 2004). This strategy can be used for a variety of assessment methods but is most commonly associated with constructing traditional summative tests. When constructing a test, teachers need to be concerned that the test measures an adequate sampling of the class content at the cognitive level that the material was taught. The TOS can help the teacher’s map the class of time spent on each objective with the

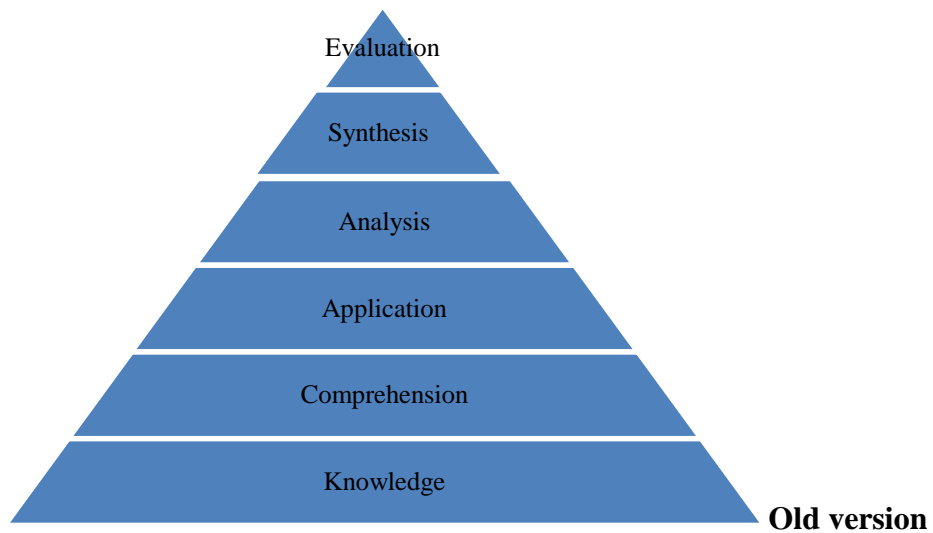
cognitive level at which each objective was taught thereby helping teachers to identify the types of items they need to include on their tests (Anderson et al, 2006).

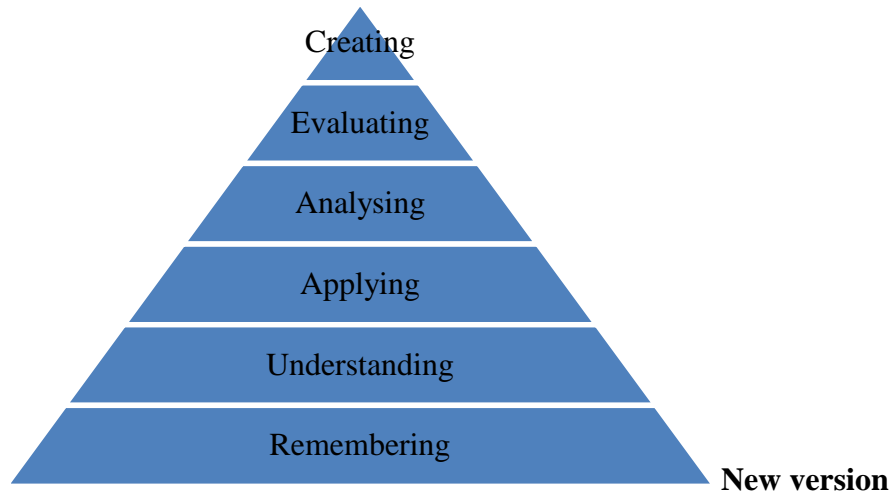
In order to understand how to best modify a TOS to meet your needs, it is important to understand the goal of this strategy: improving validity of teachers' evaluations based on a given assessment. Validity is the degree to which the evaluations or judgments we make as teachers about our students can be trusted based on the quality of evidence we gathered (Wolming & Wilkstrom, 2010). For classroom assessments two sources of validity evidence are essential: evidence based on test content and evidence based on response process (APA, AERA, NCME, 1999). Evidence based on test content underscores the degree to which a test or any assessment task measures what it is designed or supposed to measure (Wolming & Wilksrom, 2010). Response process evidence is the second source of validity evidence that is essential to classroom teachers. Response process evidence is concerned with the alignment of the kinds of thinking required of students during instruction and during assessment (testing) activities.

Six levels of thinking were identified by Bloom in the 1950's and these levels were revised by a group of researchers in 2001 (Anderson et al). The revised version of the taxonomy is intended for a much broader audience. Emphasis is placed upon its use as a "more authentic tool for curriculum planning, instructional delivery and assessment" . The structure of the revised taxonomy table matrix "provides a clear, concise visual presentation" (Krathwohl, 2002) of the alignment between standards and educational goals, objectives, products and activities. The taxonomy is to reflect relevance to 21st century work. Note the changes from nouns to verbs associated to each level. The new version of Bloom's taxonomy includes the following cognitive process dimensions:

- **Remembering:** Can the student recall or remember the information?
- **Understanding:** Can the student explain the ideas or concepts?
- **Applying:** Can the student use the information in a new way?
- **Analyzing:** Can the student distinguish between the different parts?
- **Evaluating:** Can the student justify a stand or decision?
- **Creating:** Can the student create new product or point of view?

Today's teachers must make tough decisions about how to spend their classroom time. The pyramids show a diagrammatic representation of the old and new version of Bloom's taxonomy.





To build assessment tasks for tests, match each skill or concept with a type of question or performance task and will create the best opportunities for students to demonstrate their understanding. Bloom's taxonomy offers a basic framework for generating a variety and range of assessment tasks. This range of question types can provide students opportunities to answer questions at various levels of thinking .It can generate rich assessment information not only what students know but what additional learning they need.

Test validity can be interpreted as usefulness for the purpose. Validity is an integrated evaluative judgment on the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment (Messick, 1993). Validity of classroom assessment depends on;

- Analyzing the intended learning and all its embedded elements.
- Having a good match among the assessment approaches, the intended learning and the decisions that teachers and learners make about learning ensuring that the assessment adequately covers the targeted learning outcomes including content, thinking processes, skills and attitudes.

- Providing learners with the opportunity to show the knowledge of concepts in many different ways and with multiple measures to establish a composite picture of learner learning. It ensures the central question; does assessment measure what it purports to measure (Wragg, 2001).

Face validity is the most common criterion which addresses the question; does the test look as if it does the job it is intended to do? (Wragg, 2001). Lacity et al (1994) defines face validity as making common sense and being persuasive and seeming right to the reader. That is validity taken at face value.

Content validity refers to the extent to which a learner's response to a given assessment reflects the learner's knowledge on the content of interest (Moskal & Leydens, 2000). Content validity is also concerned with the extent to which the assessment samples the content domain. This requires affirmation from the expert. The expert should look into whether the content is representative of the skill that are supposed to be measured, that is test objectives, syllabus content and the test contents (Maizan, 2005). Content covered and the cognitive skill level should conform to a set syllabus (Black, 1997). Classroom teachers need the requisite skills for them to be experts. Lack of these might result in assessment problems. Robert Linn and David Millar (2005) propose four major considerations that arguments concerning validity should take into account. These are content considerations, construct considerations, criterion relationships and consequential considerations.

What the considerations entail and how they can help us evaluate validity:

- Do the tasks match the learning intentions we are interested in?
- Does the test cover a wide enough range of content?

- Are there enough items or tasks to cover the scope of what is being addressed?
- Do the tasks require use of the desired skills and reasoning processes?
- Is there an emphasis on deep, rather than surface knowledge?
- Are the directions for the assessment task clear?
- Are the questions unambiguous?
- Are the time limits sufficient?
- Do the tasks avoid favoring groups of students more likely to have useful background knowledge for instance, boys or girls?
- Is the language used suitable?
- Are the reading demands fair?

Our ability to make valid interpretations and decisions based on assessment data can be weakened by many factors. Being aware of these can help us frame questions that inform our decision-making about validity claims (Millar, 2005).

Theoretical framework

This study has been guided by the Cognitive learning theory. According to cross K.P et al (1996) this theory connects classroom assessment to learning. Cognitive theory focuses on what is going on in the mind of the learner rather than thinking of learning as a simple stimulus – response connection. The learner is an active participant in the learning process. The theory describes learning as a building of connections between a learner’s prior knowledge and experience and the new information or skill that is being learned. A successful learning episode results in the assimilation of new information into the long-term memory structure of the learners via these connections. Later, when the new information is needed, the learner will activate any one of a

number of these connections and the information will be retrieved for use. There are several key processes in the building of connections during learning. The three most important are:

- **Attention:** For information to be learned, the learner must focus attention on it.
- **Encoding (deep processing):** The learner transforms the information in many ways that make it more meaningful, more connectable and therefore more retrievable.
- **Meta-cognition:** The learner is aware of and in active control of his or her own learning. This is manifest in goal setting (understanding or deciding why something needs to be learned and setting the standard by which progress will be measured); comprehension monitoring (recognizing when one isn't learning and why); strategy selection (being able to select from an array of learning strategies those most likely to achieve the goal); and resource management (being aware of and able to bring a wide range of resources into play in achieving the goal).

ECDE trainers should reflect on all aspects of learning when constructing continuous assessment tests for the trainees.

Statement of the problem

Assessment of students is very critical because effective teaching decisions are based on the ability of teachers to understand their students and to match actions with accurate assessments (Mc Millan, 2008). Question papers remain one of the most critical instruments assessing knowledge and skills acquired by learners and therefore the assessment instruments that are used in the examination must be of the highest quality and standard. Most trainers have limited assessment training and even less time to develop and evaluate their own tests. Trainers routinely develop tests without considering validity issues. Assessment for Diploma and Certificate levels requires that

the trainers construct continuous assessment tests to be administered to the trainees and the scores be submitted to Kenya National Examinations Council (KNEC). This constitutes 30% of their final examination paper marks to enable them meet the course requirement. Poor syllabus coverage reported by Chief Examiners affects the candidates' performance in National Examinations and test validity at institutional level. It is against this background that a study was conducted to examine the factors influencing validity in classroom test construction.

The purpose of the study

The purpose of this study is to examine the factors influencing validity in classroom test construction and challenges experienced by trainers in test construction in ECDE programmes.

Research objectives

The specific objectives of the study were four fold:

1. To explore the content and face validity of ECDE classroom examination.
2. To find out the trainers experience in test construction.
3. To establish the use of table of specifications (TOS) and bloom's taxonomy in test construction.
4. To examine the challenges experienced by trainers in test construction.

Research design and area of study

The study was conducted 2 ECDE training institutions namely; the City Centre for Early childhood Education (CICECE) and Nairobi East District Centre for Early Childhood Education (DICECE) in Nairobi County. These are Government training institutions registered by MoEST and KNEC as Examination centre's offering Diploma and Certificate examinations.

The study adopted a multiple Case study research design. A Case study is an approach to research that focuses on gaining an in-depth understanding of a particular entity or event at a specific time. Carla Willig (2008) asserts Case studies focus on a particular unit of analysis.

The study population

The Study population targeted 15 trainers handling Certificate and Diploma levels. 5 trainers from Nairobi East DICECE and 10 trainers from CICECE were selected and were all involved in the study. Purposive sampling was used in selection of the respondents. In this form of sampling, the investigator relies on his/her expertise or expert judgment to select units that are representative or typical of the population. The general strategy is to identify important sources of variation or criteria in the population, and then to select a sample that reflects this variation (Orodho, 2012).

Data collection instruments

The main research instruments used in the study included Questionnaires, structured interview and document analysis. The Questionnaire sought information on the use of a Table of Specification, Bloom's taxonomy and trainers experience in test construction. The structured interview sought information on the challenges experienced by the trainers in test construction. The Document analysis was used to explore the content and face validity of the tests. Schram (2003) identifies an advantage of using document analysis as documents can be accessed at a time convenient to the researcher at a minimum cost. The trainers were requested to provide one continuous assessment test and a marking scheme for each subject.

Data analysis

Data analysis was done using descriptive statistics and the information was presented in figures and tables using percentages.

Findings and Discussion

Table 1: Content and face validity

Content and Face validity	Frequency	Percentages%
Adequate Syllabus coverage	10	66.7
Objective identification	10	66.7
Content identification	10	66.7
Appropriate paper length	12	80
Time allocation	7	46.7
Appropriate language use	8	53.3
Test items errors	10	66.7
Availability of Rubric	5	33.3
Balancing of skills	5	33.3
Availability of a marking scheme	15	100

The study findings revealed that majority (100%) of the trainers developed marking schemes while more than a half (80%) developed test papers of appropriate length with sufficient test items while more than a half (66.7%) were able to identify and make use of the specific syllabus

objectives, content and adequately covered the syllabus . It also emerged that more than a half (66.7%) had test item errors while less than a half (46.7%) did not allocate time in the test papers while (33%) did not have the rubric and had imbalanced skills which was generally skewed on the lower levels of Bloom’s taxonomy. These study findings are not supported by a study done by Wragg (2001) on face validity which addresses the question does the test look as if it does the job it is intended to do? And concurs with Maizan (2005) on content validity in which the expert should look into whether the content is representative of the skill that are supposed to be measured, that is test objectives, syllabus content and the test contents.

Table 2: Trainers experience in test construction

CATEGORY	DURATION (YEARS)	FREQUENCY	PERCENTAGE (%)
ECDE TRAINERS	1-5	03	20
	6-10	07	46.7
	11 & ABOVE	05	33.3
TOTAL		15	100

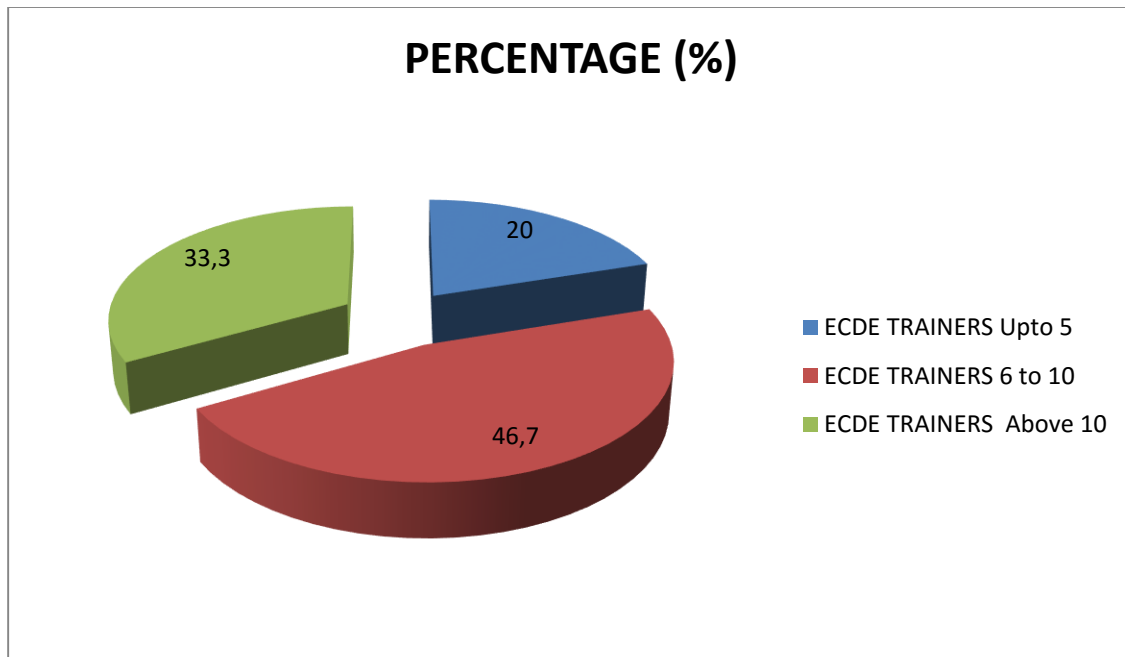


Figure1: Trainers experience in test construction

The study revealed that more than a half (80%) of the ECDE trainers had experience of more than six (6) years in ECDE training. Despite the longer training experience, they were unable to effectively develop valid tests. This is in line with the studies done by Marso & Pigge (1988) who found that neither teacher's ratings of their own testing proficiencies nor the quality of their teacher made tests improved with the teachers years of experience.

Table 3: Participation in training for assessment

TRAINING	RESPONSES	FREQUENCY	PERCENTAGE%
TRAINING IN ITEM WRITING	YES	0	0
	NO	15	100
	TOTAL	15	100
	YES	10	66.7

PARTICIPATION IN ANY TRAINING PROGRAM ON ASSESSMENT	NO	05	33.3
	TOTAL	15	100

It emerged that none (100%) of the trainers had been trained in item writing while more than a half (66.7%) of the trainers had been trained in assessment as examiners by KNEC. However, the examiner training is insufficient to empower the trainers with skills to adequately construct classroom tests. This is in tandem with study done by Stiggins (1991) in which the assessment training teachers do receive often focuses on large scale administration rather than on the test construction.

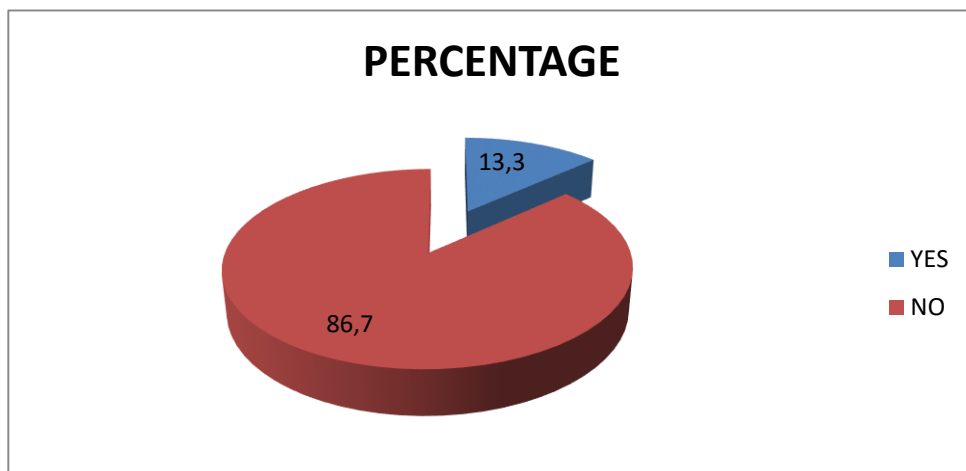


Figure 2: Awareness about table of specifications

The study findings revealed that the trainers conceptualized planning for a test in their own understanding. Majority of the respondents (100%) mentioned that they were engaged in planning for the administration of the test while more than half of the respondents (86.7%) had no idea of what a Table of Specification is in test construction. This implies that the trainers did not plan for the development of the continuous assessment test by drawing a Table of specification. The implication is that most of their test items were skewed to the lower level cognitive skills. This does not concur with studies done by Notar et al (2004) which requires that teachers develop a Table of Specification which helps teachers align objectives, instruction and assessment.

Table 4: Application of Bloom’s taxonomy in test construction

SN	PAPER CODE	NO. OF ITEMS	LOW ORDER SKILLS		HIGH ORDER SKILLS		TOTAL (%)
			FREQUENCY	PERCENTAGE%	FREQUENCY	PERCENTAGE%	
1	6011	15	14	93.3	1	6.7	100
2	6012	20	13	65	7	35	100
3	6013	16	13	81.3	3	18.7	100

4	6014	26	20	77	6	23	100
5	6015	20	16	80	4	20	100
6	6016	20	18	90	2	10	100

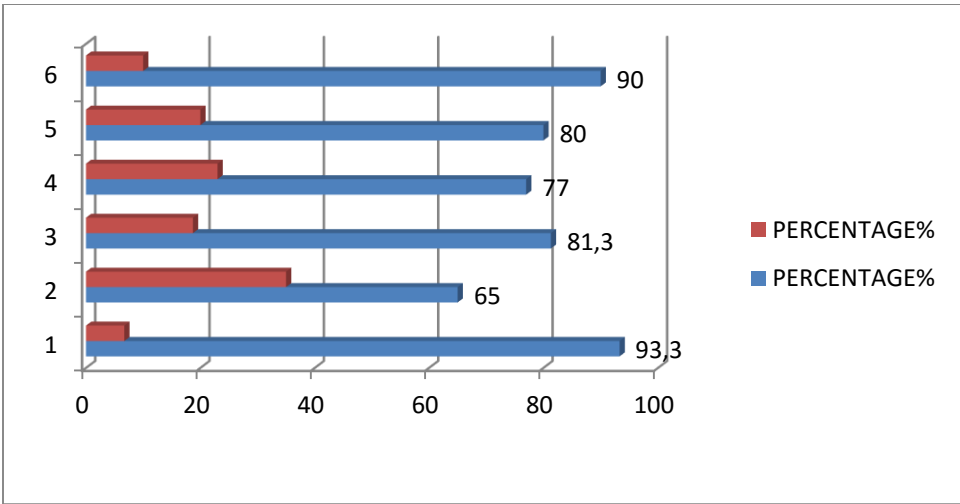


Figure 3: Application of Bloom's taxonomy in test construction

The study findings revealed that more than a half (93.3%) for 6011,(90%) for 6016 (81.3%) for 6013, (80%) for 6015 ,(77%) for 6014 and (65%) for 6012 applied the low order skills of Bloom's taxonomy while slightly less than a half (35%) for 6012,(23%) for 6014, (20%) for 6015,(10%) for 6016 and (6.7%) for 6011 applied the higher order skills respectively. It is evident that more emphasis is on the lower level skills. This concurs with a study done by Stiggins(1985) in which one of the problems with teacher –made test is their emphasis is on lower-level thinking skills which affects test validity.

Table 5: challenges experienced by trainers in test construction

Category	Reported challenges	Frequency	Percentages
----------	---------------------	-----------	-------------

KICD ECDE SYLLABUS	Outdated content	13	86.7
	Mismatch of objectives with content	12	80
	Content without objectives	7	46.7
	Objectives not SMART	7	46.7
	Broad syllabus	13	86.7
	Content overlaps	10	66.7
RESOURCES	Lack of variety reference materials	12	80
TEST QUALITY	Ambiguous items	8	53.3
	Language level/high pitched	8	53.3
	Many items measuring knowledge	10	66.7
	Inadequate Sampling content	13	86.7
	Application of action verbs	10	66.7
	Originality of test items	12	80

The findings revealed that majority (86.7%) reported that the syllabus content was outdated, inadequate sampling of content attributed to poor syllabus coverage and experienced a challenge of matching specific objectives with appropriate content. More than a half (80%) lacked variety of reference materials and had a challenge in developing original test items by lifting items from KNEC past papers while (66.7%) reported content overlaps within and across syllabus units making it difficult for the trainers to adequately sample content, developed many items measuring knowledge skills and experienced difficulty in application of action verbs used in Bloom's taxonomy. It also emerged that slightly more than a half (53.3%) developed ambiguous items and

high pitched test items while slightly less than a half (46.7%) revealed that some syllabus content lacked objectives and some objectives were not measurable or SMART. This has an implication results into construction of tests with low content and low face validity. The findings is in tandem with the findings of Black (1997) who stated that content covered and cognitive skill level should conform to a set syllabus.

Conclusion

Considering the findings of the study, it is clear that the trainers constructed tests which were of low content and low face validity. Despite their longer experience in ECDE training, they still had insufficient skills in test construction and inability of planning for test development. The syllabus, inadequate resources and test quality related challenges affected test validity.

Recommendations

The following recommendations were made based on this study:

- The Ministry of Education, Science and technology (MoEST) through Education Standards Quality Assurance council (ESQAC) should monitor ECDE curriculum implementation in the training institutions to ensure adequate syllabus coverage by the trainers. This would ensure adequate sampling of syllabus content for test construction.
- The County Director (MoEST) should work closely with the ECDE trainers at the county level and monitor the ECDE teacher training component to ensure quality training. They should also organize for training workshops and invite resource persons to facilitate on item writing to improve validity.
- The Kenya institute of curriculum development (KICD) should fast track the review of the ECDE Syllabuses for Certificate and Diploma levels. This would eliminate the challenge

of content overlaps, outdated syllabus content and formulate SMART and measurable specific objectives .This would provide a basis of developing quality test items that would improve validity of ECDE continuous assessment tests and keep abreast with the best assessment practice in the 21st Century. A Unit in measurement and evaluation should also be incorporated in the Diploma ECDE Syllabus.

- The Kenya National Examinations Council (KNEC) should train more trainers in item writing in order to build their capacity in assessment to enable them effectively carry out classroom assessment. The knowledge and skills in assessment would be beneficial at institutional level as the trainers would construct valid continuous assessment test items and in future participate in depositing quality test items in the KNEC item bank.
- The Dean of Curriculum should effectively coordinate assessment at institutional level to ensure professionalism in the development of continuous assessment tests. The trainers should also ensure adequate syllabus coverage to enable them sample the syllabus content appropriately.

Reference

Boothroyd, R.A., Mc Morris, R.F., and Pruzak, R.M. (1992). What do teachers know about measurement and how did they find out? Paper presented at the annual meeting of the National Councils on Measurement in Education, San Francisco, C.A.

Hills, J.R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan*, 72(7); 540-545

Kenya Institute of Education (2006), *Teacher Certificate in Early Childhood Development and Education Syllabus*, Nairobi; KIE

Lacity M. and Wildcock L. (1994) *Information System and Outsourcing: A transaction cost interpretation of empirical evidence OX // M working paper*. Templeton College. Oxford University, Oxford

Linn, R.L., and Gronlund, N.E. (1995). *Measurement and assessment in teaching*. New York: Prentice Hall

Maizan A. (2005). *Assessment of Learning Outcome: Validity and reliability of classroom tests*. World Transaction on Engineering and Technology Education. Vol.4 No.2

Marso, R.N., and Pigge, F.I. (1988) *An analysis of teacher-made tests: Testing practices, cognitive demands, and item construction errors*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA

Messicks. (1993), *Validity*, in R.L Linn (Ed) *Educational measurement* (2nd Ed: pp 13-104) Phoenix: American Council of Education and Oryx press

Moskal, B.M and Leydens, J.A (2000) *Second development: Validity and Reliability*. Practical assessment, *Research and Evaluation* 7(10). Retrieved 22,2012 from <http://pareonline.net/getun.asp>.

National Centre for Early Childhood Education (1999), Guidelines for Early Childhood Development in Kenya Nairobi; KIE

Notar,C.E.,Zuelke,D.C,Wilson,J.d and Yunker B.d (2004).The table of Specifications:Insuring accountability Edvisions approach teacher made tests.Journal of instructional Psychology,31:115-129

Oescher,J.,and Kirby ,P.C(1990)Assessing Teacher-made tests in Secondary Math and Science Classrooms.Paper presented at the annual meeting of the National Council on Measurement Education,Boston

Orodho, J.A, (2012).Techniques of writing research proposals and reports in education and Social Sciences.Nairobi:Kanezja publishers

Stiggins,R.J and Bridgeford N.J.(1985).The ecology of Classroom assessment.Journal of Educational measurement 22: 271-286

Republic of Kenya (2010).The Constitution of Kenya. The National Council of Law. Nairobi; Government Printer

Republic of Kenya (2007). The Vision 2030, Nairobi, Government printer

Schram, T.M. (2003). Conceptualizing qualitative inquiry. New Jersey, Merrill Prentice-hall

Stiggins, R.J.(1991)Assessment literacy.Phi Delta Kappan.72;534-539

Wise,S.L.,Lukin,L.E and Roos,L.L(1991).Teacher beliefs about training in testing and measurement.Journal of Teacher Education,42:37-42

Wragg, E.C (2001). *Assessment and Learning in the Primary School*. Rout Ledge Publisher,
Falmer