

# Quality Matters

Quality Matters, Vol. 7 No. 26 May 2013

A Quarterly Newsletter of the Center for Educational Improvement and Quality Assurance (CEIQA)

St. Mary's University College

## In this Issue

From the Editorial Desk.....	2
News .....	3
Research Corner.....	4
Interview.....	5
Perspective.....	7
Photo Gallery .....	11

## Quote of this issue

*Self-assessment is essential for progress as a learner: for understanding of selves as learners, for an increasingly complex understanding of tasks and learning goals, and for strategic knowledge of how to go about improving.*

*Sadler*

This newsletter is published every three months by the Center for Educational Improvement and Quality Assurance (CEIQA) of St Mary's University College (SMUC). The objective of the newsletter is to inform the SMUC community as well as interested public, private, non-governmental stakeholders about the activities and endeavors of the institution in fostering quality education and research in the Ethiopian Higher Education setting.

Tel.251-11-553 7999 / 8001

P.O.Box:1211

Email: [ceiqa@smuc.edu.et](mailto:ceiqa@smuc.edu.et)

<http://www.smuc.edu.et>

Addis Ababa, Ethiopia

## Agreement Signed

St. Mary's University College (SMUC) and ZTE (H.K) Limited Ethiopian Branch signed an agreement, to run ZTE Uni-Certification Program, on June 25th, 2013 here at SMUC Main Campus in Addis Ababa. The program focuses on training packages pertinent to the telecom industry, which requires skilled manpower in network architecture and related areas. ZTE is an international commercial telecom company, which has its own university. Representing the respective institutions, Ato Wondwosen Tamrat (Asst. Prof, Founder and President of SMUC) and Liu Xuefei (ZTE Training Manager) signed the agreement.

## 13th AAU General Assembly and Africa-EU Partnership Conference on Quality Assurance were Held

St. Mary's University College, represented by the Executive V/President, participated in the 13th General Assembly of the Association of African Universities (AAU), which took place in Libreville, Gabon, from May 28 to 31, 2013. The assembly, which gathers every four years, was held under the theme: "Transforming African Higher Education for Graduate Employability and Socio-Economic Development". Opening remarks and Keynote presentations were made, among others, by a representative of H.E. Ali Bongo Ondimba, Gabon's President, the Minister for Education, the Secretary-General of the AAU, Prof. Etienne Ehile and the Acting President of the AAU, Prof. George Magoha. During the three-day sessions, in addition to paper presentations, meeting of Sub-Regional Caucuses for Governing Board Elections were held and Executive Board resolutions were ratified.

... Continued on Page 4



## From the Editorial Desk

Assessment can take different forms. It can be teacher driven or student self assessment driven type. Teacher driven assessment is used to check whether students comply with the academic expectations of courses or not. In teacher driven tests, students do not judge their learning. Teacher driven tests can be classified into norm referenced and criterion referenced. Norm referenced tests are employed to rank learners on the basis of their achievement as high, moderate and low achievers, where as criterion referenced tests are conducted to know what learners can do and what they know. Norm referenced tests interpret performance with the aid of statistics. Experts criticize norm referenced tests saying that they encourage unnecessary competition, and are highly dependent on the normal distribution. The issue of competition may create unhealthy relationships among students, and it may have its own negative impact on the teaching learning process. Grading on the curve can also force teachers to group examinees into below average, average and above average when that is not practically the case. Besides, the use of objective tests in both norm referenced and criterion referenced tests are far from the real world in such a way that employees in the world of work are expected to do something not to choose or say either true or false. Again both tests give emphasis to validity and reliability. Valid assessment requires that we measure what we set out to measure, but, too often, scholars point out that teachers measure what they are able to measure. Similarly, multiple and double marking are used to test reliability, but experts contend that conformity between markers do not give cause for great confidence that marking is reliable, as multiple marking causes regression to the mean. To minimize the above problems, there is a need to promote authentic tests and student self assessment.

Authentic tests are realistic in such a way that they reproduce the circumstances to test the skills and

knowledge of examinees in real like situations. This usually involves doing in the examination what have been taught in courses. Authentic tests require examinees to apply multiple elements to pass judgments, solve problems or innovate.

The argument behind encouraging student self assessment is that it promotes lifelong learning and enables learners to assume responsibility. Self assessment develops the ability of learners to plan and direct their own learning without the assistance of their instructors, and this supply them various techniques, essential to appraise their success or failure of their learning independently so that they can take therapeutic measure whenever they sense their weaknesses without being told by another person. Learners are also expected to be responsible for their learning by not limiting their learning process to the higher education stay alone. Instructors are, therefore, expected to expose and train learners on how to evaluate and make judgments regarding the quality of work by helping them formulate explicit criteria. This trend, through time, can be part of their habit, and they may use the practice whenever they pursue their self directed learning. The use of self assessment also enables students to re-examine their progress of learning, and it also builds their self confidence in order to assume greater levels of responsibility in directing and evaluating their learning.

Hence, higher learning assessment should not be limited only to teacher or program driven measurements. There is a need to encourage authentic and self assessments. By doing so, learners can show true progress, equipping themselves with the required core skills, vital for real life task performance and lifelong learning.

This newsletter focuses on higher learning assessment. Enjoy reading it!



# Research Corner

**Title:** An Evaluation of the Implementation of Continuous Assessment (CA): The Case of St. Mary's University College

**Publication:** Published in the **5th Annual Research Proceedings of Saint Mary's in 2007 G. C.**

**Researcher:** Bekalu Atnafu

The traditional tests alone no more satisfy to evaluate teaching learning processes. Accordingly, the focus and methods of evaluation have shifted from summative to an ongoing continuous assessment classroom type of evaluation which enable to gather appropriate data for decisions to make regarding learners actual performance about the courses they take.

Assessment has many purposes: it appraises learners' achievements; provides feedback to both students and instructors about their learning and teaching; directs learning; motivates students to learn; serves to evaluate teaching methodology, etc. To realize the foregoing purposes, various continuous assessment methods can be applied such as tests, assignments, examination, quiz, projects, presentations, questioning, class participations, group reflections, discussions, portfolio assessment, term paper, self-assessment, observation, interview, peer assessment, attendance, role-playing, fieldwork, practical work, homework and the like.

For an effective implementation of continuous assessment, instructors should be equipped with the required knowledge and capability of designing various assessment techniques. They should plan the assessment activities, grounding on the actual condition, social factors, and pupils' level of knowledge of the class and the nature of instruction. Instructors should select and apply different continuous assessment methods and record the data for decisions. This type of evaluation can provide natural and generic picture about the progress and

development of students' performance by discouraging last minute cramming.

The researcher wanted to evaluate the practices of continuous assessment at St. Mary's University College. From 150 regular instructors, of which 85% were MA holders, 40 of them were selected randomly. The majority of the lecturers (75%) had education background. In addition to the lecturers, department heads took part in the study, and data were collected through questionnaire, document analysis and interview.

As to the findings, most instructors (90%) believed that continuous assessment is beyond testing, and they made clear that it is more than the traditional paper and pen testing. Most instructors (75%) also felt that continuous assessment improves pupils' learning by offering immediate feedback. The majority of them (80%) believed that continuous assessment motivates students to learn. They further pointed out that continuous assessment enabled them to review their teaching methods.

All instructors claimed that they have been frequently measuring their students' performance, applying continuous assessment. Practically, however, the findings revealed that two thirds of the instructors (77.5%) used few continuous assessment techniques. Tests, assignments, examination, quiz, projects, presentations and questioning were the most widely used tools of continuous assessment, and yet group reflections, discussions; portfolio assessment and term paper were used rarely as tools of continuous assessment. The use of few techniques in a semester could not be considered as continuous assessment. The traditional mode of assessment in which they passed through might have influenced instructors in practicing limited assessment techniques. While pinpointing the reasons in the open-ended responses, most instructors reported that shortage of time; lack of commitment; tight schedule; broad course content and teaching load were the factors which hamper the implementation of continuous assessment.



*Continued from page 1...*

In a similar development, St. Mary's also participated in a conference organized by the European Union in collaboration with the Association of African Universities (AAU) from May 31 to June 1, 2013 in Libreville, Gabon. The conference was conducted bearing the general theme: **Exploring Quality Assurance through the Africa-EU Partnership**. Under this major theme, two salient issues were presented for discussions: challenges and the interlinking roles in quality assurance and quality initiatives at institutional and national level; the way forward. With this framework, participants explored the relationship between government policy, institutional autonomy, accountability and quality assurance. It was also underlined the need for the establishment of a Continental Accreditation Agency for higher education in Africa focusing on possibilities for regional cooperation in setting up the Agency.

Paper presenters at the workshop came from diverse institutions: The European Commission, the African Union, University of Pretoria, Dublin Institute of Technology and University of Kwazulu-Natal. More than 60 participants were in attendance and most of them representing their respective national Quality Assurance Agencies.

#### **The 7TH National Student Research Forum**

The 7th National Student Research Forum was held here at St. Mary's Multi-purpose Hall on July 4, 2013. Since 2006, the forum has exclusively entertained research papers from undergraduate students, who are in their final years. This year's day-long event witnessed paper presenters, other than St. Mary's, from Jimma University, Debre Markos University, Admas University College and Addis Ababa University. Thirteen papers were presented covering topics like 'Entrepreneurship', 'Marketing and Financial Management,' and 'Human Resource Management'. Sponsored and organized by St. Mary's University College, welcoming and opening remarks were given by Wondwosen Tamrat (Asst. Prof.), Founder and

President of SMUC, and Tesfaye Teshome (PhD), Director General of Higher Education Relevance and Quality Agency respectively. The Research and Knowledge Management Office, SMUC, was in charge of organizing the event.

#### **Multi-disciplinary Seminar Held**

The 5th Multi-disciplinary Seminar was held on July 24, 2013. The thematic areas covered, at the annual event, include Human Resource and Institutional Management, Quality of Education in HEIs, and Gender and Culture. Sponsored and organized by St. Mary's, the Seminar had paper presenters, other than SMUC, from Adama Science and Technology University, Wollega University, the Ethiopian Investment Agency, the Ethiopian Management Institute and Bank of Abyssinia. Opening and closing remarks were made by the Executive V/President and the V/President for Research, Outreach and School of Graduate Studies, SMUC, respectively. St. Mary's organizes three annual research oriented events; the two others are the Student Research Forum and the conference on Private Higher Education Institutions. The 11th Conference on Private HEIs is to be held on August 24, 2013 at the UNECA Conference Center here in Addis Ababa.

Quality Matters

Quality Matters

Quality Matters



# Interview

*This column features interviews of people including government officials, policy makers, educators, and presidents of universities and colleges as well as students on quality related issues. In this edition of the newsletter, we interviewed Dr. Wubishet Shiferaw on the role of St. Mary's Testing Center activities and higher learning assessment. Dr. Wubishet Shiferaw earned his PhD from Bulgaria in Psychological Sciences from Academy of Social Sciences and Social Management in 1990. Presently, he is the Vice Director of Saint Mary's Testing center. Before joining St. Mary's, he served as Head of Department for Examination Development and Administration in National Organization, and as Acting Head for Board of National Examinations at the Ministry of Education. Enjoy the interview!*



**Quality Matters:** Saint Mary's University College has a Testing Center. When was it established, and what are some of the duties of the Testing Center?

**Dr. Wubishet:** The Testing Center was established in 2009, and its major duties include producing item assessment tools for the College of Open and Distance Learning (CODL), conducting item analysis for the regular division and CODL degree courses, maintaining test item banking and providing training for SMUC staff. Besides, the Center conducts research on problems of assessment and related matters, provides special purpose tests such as employment and placement, administers TOEFL iBT international test, registers new ACCA students, conducts tutorials for ACCA papers and provides Comprehensive Degree Exit Examinations, Institutional TVET Comprehensive Examinations, etc.

**Quality Matters:** How do you measure the contribution of the Center to realize quality education in the University College?

**Dr. Wubishet:** I think the Testing Center is contributing a lot to realize quality education by performing the aforementioned major activities as its duty. The above activities are one way or another linked to quality education.

**Quality Matters:** Experts in the field of testing attempt to classify tests into two: traditional and modern type. What do they mean when they say so?

**Dr. Wubishet:** Traditional assessment refers to test types which are forced choice measurer or supply type of pencil and paper tests such as multiple choice, fill in the blanks, true-false, matching and the like. In contrast, authentic/ performance assessment refers to tasks that are usually performed by students to demonstrate their mastery. Traditional assessment can commonly be distinguished from authentic assessment in terms of attributes such as selecting a response to performing tasks, contrived to real life, recall/ recognition of knowledge to construction/application of knowledge, teacher structured to student structured and indirect evidence to direct evidence.

**Quality Matters:** Some tests promote deep learning while others fail to do so. Which kinds of tests promote deep learning?

**Dr. Wubishet:** The acquisition of skills require deeper learning but not mere recalling of facts and figures and recitation. Deep learning is all about preparing students to be investigators and problem solvers and enabling



them to transform the knowledge and skills they acquired to real life situations. As a result, the assessment of deeper learning relies on developmental assessment of both content and skills. In assessing deeper learning outcomes, students are required to think critically and creatively and use higher learning competencies and problem solving tools to look for new meaning. Due to these, authentic assessment is usually considered as an appropriate tool to assess deep learning.

**Quality Matters:** Experts in the field of testing claim that students should engage actively in testing. It is not instructors alone who should judge learners knowledge but students themselves too. What do they mean when they argue so, and how can it be practiced?

**Dr. Wubishet:** Student self assessment occurs when learners assess their own performance. From self assessment, they learn to reflect and evaluate their own progress, understand gaps in their capabilities, discern on how to improve their performance, learn independently and think critically.

For effective self assessment to take place, students must first learn the technique of peer assessment and be introduced to the idea of self assessment in the process of formative evaluation. Teachers may require students to use a rubric to provide critique on the work of their peers, and further encourage to apply the same criteria to their own work. Self assessment can be incorporated to any assessment task either at or after assignment submission time. Combining self and peer assessment with teacher assessment, the decision maker can get a more reliable result about learners learning.

**Quality Matters:** Some feedbacks promote learning while others do not so. What kinds of feedback promote learning, whereas which ones do not facilitate learning?

**Dr. Wubishet:** Feedback is usually given on all assessed work- both formative and summative- and uses a range of approaches which are deemed as best practices in each discipline. There are different ways of giving feedback. These include traditional written comments

on feedback sheet about transferable skills, seminar or tutorial session, class room group feedback, verbal feedback for entire class, laboratory and practical work feedback, exam and final year project and dialogue between students and peers and between students and academics.

Similarly, group feedback aiding staff members to evaluate teaching as well as informing academic staff about students' progress, areas of success and of general comments are useful to facilitate the learning teaching process. With this in mind, effective feedback needs to be constructive, forward looking, timely, comprehensible, relevant and encouraging. Otherwise, it can be considered as less useful which may not facilitate learning as desired.

**Quality Matters:** Thank you very much for sharing your scholarly experience!

**Dr. Wubishet:** Thank you!

### Virtual Links

Ethiopian Ministry of Education  
Web site: [http:// www.moe.gov.et](http://www.moe.gov.et)  
Higher Education Relevance and Quality Agency (Ethiopia)  
Website: [www.higher.edu.et](http://www.higher.edu.et)  
PROPHE- Programme for Research on Private Higher Education  
Website: [www.albany.edu/dept/eaps/prophe/](http://www.albany.edu/dept/eaps/prophe/)  
International Network for Quality Assurance Agency in Higher Education (INQAAHE)  
Website: [http:// www.inqaahe.org](http://www.inqaahe.org)  
Quality Assurance Agency for Higher Education (UK)  
Website: <http://www.qaa.ac.uk>  
Center for International Research on Higher Education  
Website: [http://bc.edu/bc\\_org/avp/soe/cihe](http://bc.edu/bc_org/avp/soe/cihe)  
Quality and Standards Authority of Ethiopia  
Website: <http://www.qsae.org/>  
International Institute for Capacity Building in Africa  
Website: <http://www.eric.ed.gov>  
International Network for Higher Education in Africa (NHEA)  
Website: <http://www.bc.edu>  
European Association for Quality Assurance in Higher Education  
Website: <http://www.enqa.eu>  
Asian Pacific Quality Network  
Website: <http://www.apqn.org>



# Perspective



## RETHINKING ASSESSMENT IN HIGHER EDUCATION

By : ATLABACHEW GETAYE  
getayeatlabachew@ymail.com

Assessment directs learning in such a way that students apply either a deep approach by understanding the underlying meaning of their courses or a surface approach by reproducing what they have learned without essentially understanding the meaning of their courses. Ramsden (2003, p. 182), for example, states that “from our students’ point of view, assessment always defines the actual curriculum”. Biggs (2003, p. 182) also points out that “students learn what they think they will be tested on”; similarly, Gibbs (2006, p. 23) claims that “assessment frames learning, creates learning activity and orients all aspects of learning behavior”. In other words, assessment plays a pivotal role in determining students learning.

Assessment has been defined by different writers differently. Sadler in Joughin (2009,p.16), for instance, defines assessment as follows: “The act of assessment consists of appraising the quality of what students have done in response to a set task so that we can infer what students can do, from which we can draw an inference about what students know.” The above definition involves three factors: knowing what students can do about the task; passing judgment about the quality of the work done by students and making inference regarding students’ knowledge, attitude and practice.

There are two types of standard tests: norm referenced tests and criterion referenced tests. Norm referenced tests are distinct from criterion referenced tests in terms of objectives, content selection and score interpretation.

### 1. Norm Referenced Tests

Norm referenced tests are designed to categorize learners into strata, highlighting achievement differences across the continuum. They compare learners score against the achievement of learners to:

*rank pupils in order of achievement, from high to low so that decisions based on relative achievement can be made with greater confidence. Thus, a key feature in constructing norm referenced test is the selection of test items that provide a wide range of scores. This is done by eliminating those items that all pupils are likely to answer correctly and by favoring items at the 50 per cent item level of difficulty. Such items tend to maximize differences in performance and thus provide the most reliable ranking of pupils (Gronlund, 1981, pp.19-20).*

### 1.2 Criteria for Norms

For adequate norms, representativeness, relevance and recency should be met. Norms should also be comparable and sufficiently described. Representativeness takes into account two factors: the size of the sample and the method of sampling. The size of the sample should be representative enough so as to give stable scores and stable statistics. Similarly, sampling method should be done using a random selection to lesson biases. Relevance refers to the comparability of the norm group with the group under consideration. Otherwise, an irrelevant norm groups distort test results. Recency which refers to the up-to-datedness of the test is a central issue when it comes to achievement tests.

### 1.3 Types of Norms

In educational testing there are different types of norms. Two of them are widely known: national norms and local norms. When the tests are intended to be used at the national level such as standardized achievement and aptitude tests, they require national norms. The



norms should be set taking into account variables like age, geographic region, ethnic background and socio economic factors. Equally important, there is a need to use a representative sample. Having complete lists of the norm groups used for the norm referenced scores; the population should be defined along with the sampling procedure. On the other hand, to carry out intra school distinct comparisons, local norms are extremely convenient. Mean norms should be used to make comparisons between mean performance of school.

#### 1.4 Uses of Norm Referenced Tests

Most tests are norm referenced type. In many ways, the role of testing is to rank learners with reference to some group performance. For instance, after learning a subject, students may not understand courses equally for various reasons. In this case, testing is an effective tool to recognize their achievements, identifying their position within the continuum of the strata- saying below, average and above average. It is also equally important to recruit contestants within the same profession, if, for example, two competitors are applying for a job, mastering the basics of the given profession- one with better performance, the other with less performance, the better one will get encouragement. Therefore, NRTs are useful to pass decisions.

#### 1.5 Interpreting Norm Groups

Statistics is the tool which enable examiners communicate about the numbers that test procedures produce. Frequency distribution, measures of central tendency, norm referenced scores are some of the statistical elements essential to interpret norm groups. Frequency distribution lists down the values and the corresponding frequency. Measures of central tendency are points on the scale of measurement. The mean, among others, is the most commonly used measures of central tendency.

The other useful tool is norm referenced scores. The rationale for interpreting an individual's score using the norm group is to identify the position of the

individual. There are several ways of test scores. Percentiles, stanines, standard normal distribution and normalized T-scores are few to mention. Percentages show the percentage of students in the norm group who are at or below a particular rank. The percentile rank of a score is a whole number which ranges from 1 to 99 that shows the percentage of individuals in the norm group who scored at or below that particular score. For instance, if Abebe achieved a score of 65 on a 100 item test and told that his percentile rank is 80, this means that 80 percent of the people in the norm group had scores of 65 or less. The 80th percentile of this distribution is the score 65, but 65 has a percentile rank of 80. Stanines scores are another type of standard score used to report students' performance. The stanines break down the distribution into nine parts, providing equal units of measurement- each band equal to one half standard deviations in width. Stanines are calculated using standard score. Of the 9 scores, a score of 1 shows low performance, but a score of 9 reflects high performance. Normal curve equivalent scores, as its name implies, the score bases on the normal curve. It has a fixed mean of 50 and a standard deviation of 21.06. Scores range from 1 to 99. Scores from 1 to 29 are considered below average; where as those which range from 29 to 71 are average. 71 and above scores reflect above average performance.

#### 1.6 Validity and Reliability of Norm Referenced Tests

Validity refers to the degree to which a test measures what it is intended to measure. While considering norm referenced tests, weight and care should be given to content, criterion and construct validity. Many techniques can be used for estimating the reliability of norm referenced tests. Values for reliability co-efficient range from 0 to 1:00, and the greater the co-efficient, the more reliable is the measurement.

#### 2. Criterion Referenced Tests

Criterion referenced tests are not concerned with ranking learners in line with their achievement. Criterion referenced tests ascertain "... what test takers can do and what they know, not how they compare to others" (Anastari, 1988, p.102). They are designed to yield



scores that are interpretable with reference to a specified domain of content. Test items are selected based on their ability of reflecting a certain learning tasks being measured. CRTs usually are set to determine whether a student has learned the material taught in a specific grade of a course.

## 2.1 Domain Specification

To generalize results beyond the specific items on the test, there is a need to identify the domain of content circumspectly.

*If the domain consists of the 26 letters of the alphabet and we wanted to know whether a child could name the letters when they were pointed to, we could randomly select 5 of the letters, point to them, and see whether the child could correctly identify them. Another child would be presented with a different random sample of 5 letters. In this way we are developing randomly parallel tests and we can infer that a child who correctly identifies 2 of the 5 letters that are presented probably knows about 40 percent of the letters of the alphabet (Wiersman and Jurs, 1990, p.212).*

During domain specification, there is a need to consider the item form carefully. The other option to set a well defined domain is to use a specific instructional objective. In other words, tests can be developed by using instructional objectives and item specifications.

## 2.3 Uses of CRTs

The tenet behind criterion referenced tests is that everyone can learn anything. CRT, thus, avoids unhealthy competition among learners, measuring specific performance in an individualized manner. It gives due emphasis to behavioral objectives not to the individual's position, so it is not harmful to low

achievers. After formulating the cut of scores, testers can easily discriminate the mastery or non mastery of domains by a given learners.

## 2.4 Methods of Setting Standards

Comparing the performance of learners against a standard help to identify whether a student has internalized a domain or not. Performance standard can be formulated as follows:

**2.4.1 Professional Judgment.** Mostly standards are set by instructors particularly in higher institutions with the view that they are familiar with the content that is to be mastered.

**2.4.2 Nedelsky Method.** According to Nedelsky (1954), to set an appropriate standard of minimally acceptable performance test, it is advisable to do the following:

- ▶ Urge experts to deliberate on the content area
- ▶ Solicit each expert to examine every item on the test in order to eliminate a minimally competent item
- ▶ Calculate the guessing probability for the correct answer. It is done by dividing 1 by the number of alternatives
- ▶ The sum of the probabilities across the set of items is the expected score of a minimally competent examinee as determined by that expert
- ▶ The mean of the expected scores of all the experts is then the yardstick of minimally acceptable performance

Nedelsky (1954) also suggested contrasting method, identifying two groups of students: those who have scored above average and those who have scored less than average.

## 1.5 Item Statistics for Criterion Referenced Tests

### 1.5.1 Item Level Analysis

Many methods have been designed for item analysis of criterion referenced tests. The analysis of student performance on the basis of each item is, for example, useful to know the level of that specific item contribution to the total test score.



Information regarding the performance of students on a given test is a feedback so as to adjust decisions about the students, the instruction, and the test items. It is possible to get item level information through a content review and analysis of the kinds of error that students commit.

### 1.5.2 Statistical Item Analysis

For criterion referenced tests, to know the contribution of an item in order to upgrade the level of a test, there is a need to know the item difficulty and item discrimination power of tests. To identify whether a test is difficult or not, there is a need to determine the percentage of learners who answered the items correctly for item difficulty inform a lot. In fact, there is a need to examine difficulty indexes very carefully when criterion referenced tests are standards to check the attainment of instructional domain. Test items should discriminate learners into classes, identifying those who have mastered the domain of content from those who have not internalized the behaviors that the criterion referenced tests require. Brennan (1972) has come up with a mastery/ non mastery index as follows:

Item Correct	a	b	Discrimination index (B)
Incorrect	c	d	
	a+c	b+d	$B = \frac{b}{b+d} - \frac{a}{a+c}$
	No mastery	Mastery	

The result of the index ranges between -1 and +1. A positive result shows that the test's power of discrimination is good. By contrast, if the value is negative, the item fails to discriminate the two groups.

### 1.6 Reliability and Validity of Criterion Referenced Tests

Testing the same group of learners twice and comparing their results of the first with the second would help to determine the consistency of their results. When their achievements are similar, achieving mastery, it reflects that the test is reliable. Thus, if scores are consistent, they are then proofs of test reliability.

The role of validity is to see whether a test is able to measure what it is intended to measure. Of the types of validity, content validity is the prime concern for criterion referenced tests which enable to specify the relevant and representative domain so as to assess learners' performance appropriately. Content validation can be done through panel discussion. Experts can rate items appropriateness using objectives as standards.

Scholars in the field of testing are proposing to rethink our mode of assessment by shifting the focus from comparison and objective to authentic type of tests so as to promote deep learning. The world of work require tangible performance, and therefore, "students need not only to learn, but to learn in order to be able to act, and to be able to act in practice-like contexts where they can experience themselves as responsible agents and be aware of the consequences of their actions"(Gary and Thomas, 2010,p.217). Hence, the new mode of assessment, according Gary and Thomas (2010, pp.219-220), should be framed in line with the subsequent points:

**Managing Editor**  
Atlabachew Getaye  
**Layout Designer**  
Robel Tegene

Center for Educational Improvement and Quality Assurance (CEIQA)  
St. Mary's University College (SMUC)

Tel. 251-115-53 79 99 / 53 80 01 ext. 145  
P.O.Box:1211

Email:ceiqa@ smuc.edu.et  
Addis Ababa, Ethiopia.

If you have comments on this issue of the newsletter or would like to contribute to future issues, please contact our office on 011 553 79 99, 011 553 80 01 Ext 145  
Email: ceiqa@smuc.edu.et



*\_ Assessment tasks should incorporate the characteristics of practice, including the contextual and embodied nature of practice, requiring the engagement of the student as a whole person.*

*\_ Assessment tasks need to both develop and reflect students' generic abilities – for example, to communicate effectively within their discipline, to work collaboratively, and to act in ways that are socially responsible – as well as developing discipline specific knowledge and skills.*

*\_ Assessment tasks should require responses that students need to create for themselves, and should be designed to avoid responses that can be simply found, whether on the Internet, in the work of past students, or in prescribed texts or readings.*

*\_ Assessment tasks should become the basis of learning rather than its result, not only in the sense of students' responses informing ongoing teaching ...but perhaps more importantly assessment as a process whereby students produce and appraise rather than study and learn.*

Both students and instructors need to ensure that they have taken the above points into account, and by doing so, assessments fulfill the standards of work. Standards are inbuilt in the fields and practices, and there is no way to flee from such standards, and students should judge their learning quality through self, peer and instructors' assessments in order to confirm whether they have met the standards of work whenever they take a course.

## References

- Anastasi, A. (1988). **Psychological testing** (6th ed.). New York: Macmillan.
- Biggs, J. B. (2003). **Teaching for quality learning at university** (2nd ed.). Maidenhead: Open University Press.
- Brennan, R. (1972). Estimating the dependability of the scores. In R.A. Berk (Ed.), **A Guide to criterion referenced test construction** (pp.292-334). Baltimore: Johns Hopkins University Press.
- Gary, L.K and E. M, Thomas. (2010). "Putting Students First as Partners in the Learning Enterprise." In Gary, L and L, Randy. (Eds). **Higher Education Assessments Leadership Matters**. New York: Rowman & Littlefield Publishers, Inc.
- Gibbs, G. (2006). How assessment frames student learning. In C. Bryan & K. Clegg (Eds.), **Innovative assessment in higher education** (pp. 23–36). London: Routledge.
- Gronlund, N. (1981). **Measurement and Evaluation in Teaching** (4th ed) .Newyork: Macmillan Publishing Comp.
- Nedelesky, L. (1954). "Absolute grading standards for objective tests." In **Educational and Psychological Measurement**, 14, 3-19.
- Ramsden, P. (2003). **Learning to teach in higher education** (2nd ed.). London: Routledge.
- Sadler, R. (1990). "Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning". In Joughin, G (Ed.), **Assessment, Learning and Judgement in Higher Education**. Wollongong: University of Wollongong.
- Wiersman, W and S.G, Jurs. (1990 ). **Educational Measurement and Testing**. Boston: Allyn & Bacon





# St. Mary's University College

---

Center for Educational Improvement and  
Quality Assurance (CEIQA)

Tel: (251) 115537999

P.O.Box: 1211

Addis Ababa

