

NIGERIAN JOURNAL OF EDUCATIONAL RESEARCH AND EVALUATION

Volume 10, No.1, 2011



**A PUBLICATION OF THE NIGERIAN ASSOCIATION
OF EDUCATIONAL RESEARCHERS AND EVALUATORS**

VOLUME 10, NUMBER 1, 2011 ISSN 0795-3607

**NIGERIAN JOURNAL OF
EDUCATIONAL RESEARCH
AND EVALUATION**

UNIVERSITY OF IBADAN LIBRARY

**A PUBLICATION OF
THE NIGERIAN ASSOCIATION OF
EDUCATIONAL RESEARCHERS AND EVALUATORS**

EDITORIAL BOARD

Prof. E.R.I. Afolabi (Editor),
Department of Educational Foundations and Counselling,
Obafemi Awolowo University,
Ile-Ife.

Dr. N. J. Odili
Faculty of Education,
Delta State University,
Akraka

Dr. O. S. Akinsola
Nigerian Educational Research and Development Council
Sheda, Abuja.

Prof. S. O. Bandele
Institute of Education,
University of Ado-Ekiti
Ado-Ekiti

Dr. A. A. Adediwura
Department of Educational Foundations and Counselling,
Obafemi Awolowo University,
Ile-Ife.

Editorial Advisers

A. S. Olorundare	Professor of Science Education, University of Ilorin, Ilorin.
J.B. Babalola	Professor of Educational Management, University of Ibadan, Ibadan.
S.O. Ayodele	Professor of Language Education, University of Ibadan, Ibadan

TABLE OF CONTENTS

Teacher-student and Parent-Child Relationships as Determinants of Academic Achievement in Senior Secondary School Commerce in Ibadan, Oyo State.....	1-11
Adams O. U. Onuka, Ph. D. & Esther O. Durowoju	
Standard Setting: Innovative Methods for Objectivity in Tests and Performance Evaluation In Public Examinations.....	12-26
Dr E. A. Okwilagwe	
Language Education Research Problems of Tertiary Institutions in Nigeria.....	27-35
Okoye Mary Ekeata (Ph.D)	
Relationship Between Students Motivations for Schooling and Human Capital Acquisition.....	36-45
Dr. (Mrs.) Goodness W. Orluwene	
Education for Human Capital Development: Implication for School Guidance and Counseling Services.....	46-50
Njoku Obiageli. C And Okafor Nkem. C	
Entrepreneurship Education: A Sine Qua Non For Human Capital Development.....	51-59
Dr. (Mrs.) Elizabeth I. D. Offor	
Comparative Analysis of Students' Performance in Two Formats of a Chemistry Test.....	60-67
Osuala Rita Chigozie, Ph.D. and Prof Ogomaka P.M.C	
Abolishing of the Post-UTME Screening Test: Implications for Manpower Development in Nigeria.....	68-78
Romy O. Okoye (Ph.D) & Ngozi N. Agu (Ph.D)	
Evaluation of Mathematics Teachers in Human Capital Development in Central Senatorial District of Delta State.....	79-86
Mrs Bertha, E. Omoruan	

- Teaching of Character Education in Primary Schools:
Implications for Human Capital Development.....87-95
Nwoke Chinyere
- An Investigation into Science Teacher's Access to,
and Utilization Of Research Findings.....96-108
Egbujuo, Chima Jonas
- Will Self-concept Impact Self-disclosure of Sexual
Health Problems of Nigerian Undergraduate Student Clients?.....108-115
Oluwatosin S. A.
- Evaluation and Human Capital Development: The Role of WAEC
and NECO Through Multiple Choice Test Items in Biology.....116-126
Dr (Mrs) Joy Nkeiruka Onyeka
- The Effects of Inquiry and Conventional Teaching
Methods on Students' Performance in Biology.....127-132
Ajuonuma, Juliet O. Ph.D
- Causal Determinants of Attitude Toward Mathematics Among
Nigerian Junior Secondary School (JSS) Students.....133-149
Akintomide Akinjide Gabriel & Akintomide Jochebed Seun
- A Measure of Teacher Efficacy in Nigerian
Secondary Schools.....150-160
Bamidele Abiodun Faleye
- Quality of Mathematics Textbooks Recommended to Students:
Implication for Human Capital Development.....161-169
Dr. (Mrs.) Anugwo, M. N.

STANDARD SETTING: INNOVATIVE METHODS FOR OBJECTIVITY IN TESTS AND PERFORMANCE EVALUATION IN PUBLIC EXAMINATIONS

By
Dr E. A. Okwillagwe
Institute of Education
University of Ibadan.

Abstract

This paper examined the concept and methods of standard setting and obtaining the corresponding cut score using the Angoff (1971) and Contrasting Group Methods, two methods that have been established as standard methods for test and performance standards. The paper presented procedural application of each of the methods, providing illuminating discussion on key issues to be addressed in the process of setting standards and cut scores.

Introduction

Examination is the ultimate means of ensuring that a curriculum has been executed in a particular subject area and in assessing the competency of candidates who enroll for the subject. Such competency is measured in terms of knowledge and skills possessed by the candidates. Public examinations are conducted in thrust or on behalf of a nation or state by certified examination bodies like the West African Examination Council, National Examination Council, JAMB, the Cambridge Schools Examination Board and the College Entrance Examination Board. Access to these examinations is defined by predetermined criteria that candidates should meet. Public examinations are those mainly used for student's certification (e.g. WASSCE), selection for further studies (e.g. JAMB) and selection for job opportunities or employment (e.g. Civil Service examinations, and other professional examinations into law, accountancy and the medical professions).

Public examinations are highly competitive and are usually taken by a large number of candidates whose usefulness to these candidates, is highly dependent on their performance in them. They are, therefore, veritable tools for ensuring accountability in the education system of a nation since their outcomes are used in taking high stake decisions that affect these candidates. In view of the usefulness of public examinations, their conduct and determination of results should be assessed with utmost care and within reasonable standards.

Current Methods of Determination of Results by Public Examination Bodies

Most public examination bodies in Nigeria conduct external examinations twice a year, set apart at about six months duration. Following closely to the conduct of these examinations is their scoring and grading in order to determine the results. This is usually done by the application of predetermined formula which is individual

examination body specific to obtain the number of candidates who are passing or failing. The WAEC, for instance, use standard nine called Stanine to grade candidates. While these methods may have been in use for a long period of time by the respective examination body, it is the contention in this paper, those better and more objective methods which are empirically determined, could be adopted to achieve higher standards and to be in conformity with other foreign examination bodies.

Standard Setting: Meaning and Purpose

Standard setting has been a major step of providing accountability in the education sector for stakeholders such as policy makers and other consumers of evaluation results. In schools and the education sector, standard is set to determine the minimum score which is the minimum level of competence for the award of a certificate, diploma, or promotion of students from one class to another, setting the examination's performance standard in students' assessment and to license a teacher.

Standard setting is defined by Roach and Elliot (2004) as the process of determining the appropriate cut scores for specified levels of performance. This process determines the knowledge, skills, and other competencies students are expected to "understand and demonstrate" at each level of performance. It also requires that the test scores that correspond to these expectations are determined. These scores are referred to as cut scores. According to these authors, standard setting requires making a great deal of valid judgement and as such, in their opinion, the cut scores that are linked to the proficiency levels in each content area, are not the most important outcomes of standard setting, rather it is the descriptors (indicators) of what students who achieve the various performance levels know and are able to perform.

A cut score is a point on a score scale that separates one performance standard from another. According to Brandon (2005) cut scores are the operationalized versions of performance standards. It serves as a yardstick or benchmark for making high stake decisions on students' performance. A performance standard is a statement defining and describing the knowledge or skills that students must exhibit at a performance level. According to Hansche (1998) performance standard is a point on the score scale that separates one level of achievement e.g. 'pass' from another e.g. 'fail', and is identified through a technically sound process. Usually, performance standards are set before cut scores are set, and are most often set by policy makers or academic board of an organization. They also see to it that these standards are being achieved and improved. However, performance standards can sometimes be developed by panel of judges that set scores.

Reasons for Setting Standard and Cut Scores

The purposes of setting standards are numerous and they include:

- i. to drive innovation in education that engender reforms which stress or emphasise outcome-based accountability;

- ii. the need to conduct National Assessment of Educational Progress (NAEP) to gauge the level of performance by students at critical points in their schooling and the direction of education generally;
- iii. need to engender performance assessment by nations gravitating towards evaluating higher-order skills.
In schools, cut scores are used for the purposes of:
 - i. separating students into learning groups based on their level of competency or proficiency;
 - ii. for licensure of professionals (such as teachers);
 - iii. for promotion from one class to another; and
 - iv. for selection. Also, cut scores can be used for the selection of other personnel (e.g. lawyers, and medical personnel, armed forces and civil servants).

Suggested Methodologies for Setting Standards in Public Examinations

There are different methods of setting standards in examinations. These vary from examinee-centred methods such as the contrasting group method, bookmark, the objective method and Rasch model to Computer adaptive methods. The test-centred methods include those of Nedelsky, Ebel and Angoff and the Modified Angoff. Of the later group, the Angoff method is easier to use, straight forward and direct, even though it is one of the traditional methods of setting standards which uses the judgemental means and test items to set standards. The contrasting group method is an objective and empirical method that easily amends itself to the method of test construction (i.e. item specific) of present day examination bodies in Nigeria. It is one of the modern standard setting methods evolved to tackle the inadequacies of the traditional methods. These are, therefore, recommended in this paper and are further discussed.

Practical Steps in Standard Setting with the Angoff Method

The Angoff method was developed by Angoff in 1971, and according to Livingston and Zieky (1982) the method can be used to set standard with multiple choice questions and essay tests. The cut score (passing score) is usually computed from the expected scores for the respective test items. The following are the procedures for setting standards and determining the corresponding cut score using this method.

Steps in Using the Angoff Method

Five basic steps have been identified by Livingston and Zieky (1982). However, the last two steps are more of summary views suggested by Hansche (1989) who further amplified these steps. They are:

- i. Select the judges (large and representative of stakeholders)
- ii. Define 'borderline' knowledge and skills;
- iii. Train the judges on how to use the method;

- iv. Obtain Judgement;
- v. Combine the judgements to choose a passing score.

In step ii above, the judges (panelists) review the test item by item and provide estimations of the proportion of a sample of the examinees (candidates) who answer each item correctly. These experts judge the performance characteristics of the “just competent” examinee also called 'borderline basic'. This process of estimating to determine the minimum raw scores for 'passing', or 'honour', or 'succeeding', according to Angoff (1971) entails that “the experts/panelists have to go through the test item by item and decide whether a hypothetical person (of assumed academic qualities) could answer correctly each item being considered. If a score of 1 is allotted for each item answered correctly by the hypothetical person, and a score of zero is given for each item answered incorrectly, then, the total of the items scores will equal the raw scores earned by the minimally acceptable person” or borderline person.

vi. The judges/panelists standards can be averaged to obtain a final standard.

vii. Repeat this process for each performance standard of interest, that is, if more than one standard or cut score is required (See example below). Note, this instruction is interested in establishing more than one cut score.

A typical example of the Angoff's method instructions as used in performance standards on the NAEP Mathematics Assessment obtained from the Handbook for Panelist by Hansche (1998) is an example of how the Angoff method have been modified for more relevance, is as follows:

For the Borderline Basic student, your task is to specify the probability that this borderline student should answer each item in the assessment correctly. This chance or probability for each test item can range from zero (where you would be specifying that the borderline student should have no chance of giving a correct answer) to 1.00 (where you would be specifying that the borderline student should, without a doubt, answer the item correctly). After specifying the performance level for the Bordeline Basic student on an item, you should provide estimates on the same item for the Borderline Proficient and Borderline Advanced students (Hambleton & Bourque, 1991, p. 114).

A typical description of a 'borderline student' or the 'just competent student' (Giraud, Impara & Plake, 2000, p.6) is presented here for a 9th grade mathematics student:

The barely master student is able to solve some mathematical problems that require multiple applications using numerical method, but has difficulty using traditional algebraic methods. The barely master student can solve most algebraic, geometry, simple arithmetic applications that are not complex for

example, percent, proportion, probability, mean. The 'typical' barely master student in 9th grade algebra can demonstrate the skills necessary to earn a grade of 'C' or 'C+'.

Table 1: Calculations of Performance Standards for a Single Panelist Using the Angoff Method for Two Sets of Ratings

Item	Basic		Proficient		Advanced	
	R-1	R-2	R-1	R-2	R-1	R-2
1	.30	.35	.70	.65	.80	.80
2	.40	.40	.65	.65	.85	.85
3	.25	.28	.50	.45	.70	.65
4	.60	.55	.70	.70	.95	.90
5	.70	.65	.80	.80	.90	.90
6	.30	.30	.40	.45	.75	.80
7	.20	.20	.40	.45	.70	.70
8	.50	.50	.60	.65	.85	.85
9	.60	.55	.70	.75	.90	.90
10	.45	.45	.75	.80	.85	.85
Performance Standard	4.30	4.23	6.20	6.35	8.25	8.20

(Source: Hansche, 1997; p. 96)

(For the description of these performance levels adapted after Hambleton (1998) see Appendix I.

Livingston and Zieky (1982) and Hansche (1998), explained that with an instruction as the above, panelist would then work through the complete set of test items and at times may be asked to think of 100 borderline students, and then estimate the number of these borderline students who should answer an item correctly. Many panelists seem to find the last instance an easier task than estimating the probability of correct performance on an item by the borderline student. A common practice of the judgemental method is to repeat the probability assignment process, followed by discussions among the panelists about their assigned probabilities. Sometimes, panelists are provided with item statistics or information on the consequences of their actions (i.e. passing and failing rates estimated) of various standards set. This aids them in the standard setting process and an overall effect on the resulting standards.

Obtaining Judgements for Angoff Method

The Angoff method of obtaining judgement is typically that of the 'borderline' examinee or what can be called the 'basic' level of proficiency. Livingston and Zieky (1982,p.15) contend that making judgement is not in terms of making evaluative judgements about someone's performance but they are decisions on how a borderline examinee would likely respond to each of the questions on the test. These scholars contend that the judgement method though relatively convenient, is 'much more difficult to make'.

Guidelines for Selection of Judges for Standard Setting

- i. Judges must be experienced instructors and educators.
- ii. They must be knowledgeable in all the aspects of the subject content area of interest (curricular expertise)
- iii. They must be familiar with the needs of students (knowledge, skills and abilities of those who would or would not succeed).
- iv. They must be experts in standard setting (even though some can be trained).
- v. Need to achieve representativeness (geographical, type of school, sex and status).
- vi. Besides, they must undertake a training process, where they are trained on how to identify the abilities and characteristics of the sample drawn from the population used in setting the cut scores.

Advantages of the Angoff Method

- i. It is easy and simple to implement relative to other methods in its category.
- ii. It is suitable for setting cut scores for traditional tests (MCQ and Essay).

Disadvantages of the Angoff Method

Some of the weaknesses leveled against the Angoff method, even though some seem to lack validity are:

- i. It is not appropriate for making judgement about item content because it is a contrived test.
- ii. It is not suitable for establishing achievement levels for NAEP because of the many factors identified by users.
- iii. It lacks practical utility because it only allows knowledgeable and experienced educators as judges to the exclusion of policy makers and other stakeholders. Some psychometricians do not agree on this flaw though. An example is Brandon (2005).
- iv. Not appropriate in setting multiple cut scores for performance assessment. This again is not true as the process can be repeated as many times as the intended number of cut scores (Hansche, 1998).
- v. Do not allow judges to consider the validity of the items in term of

their importance and relevance when arriving at the performance standard.

The Contrasting Group Standard Setting Method

This method can be applied to traditional tests such as MCQ and select response formats. It is also useful in standard setting on complex performance-based measures such as assessment of writings, physical task performance and tasks that require demonstration in which a judge is to decide whether the performance is above or below a certain criterion. In contrasting group method, experts are expected to make a dichotomous judgment about examinees as the case demands that is, as: 'masters/non-masters', 'good/bad', 'competent/not competent', 'certify/no certification', 'qualified or unqualified', 'passing or failing'.

Contrasting group method is based on individual examinee and not on the test. Two information is required on each examinee. These are: the 'examinee's test score' and a judgement of the adequacy of the examinee's knowledge and skills. As identified by Livingston and Zieky (1982) four requirements must be met for making judgements. These are that the judgements:

- i. must be made by qualified persons (e.g. teachers) who can determine each examinee's knowledge and skills, and the level of knowledge and skills the examinee passing the test should have. Otherwise, the trainer should provide them with this information.
- ii. Must be judgement on knowledge and skills the test is intended to measure, and they should not be influenced by other irrelevant characteristics of the examinee such as participation in class activities or punctuality in assignment submission etc. Judges should be properly informed on where to focus their judgement.

Where judges are not familiar with the examinee's knowledge and skills, they should be given opportunity to observe a demonstration, or an example of the product of each test examinee's knowledge and skills. For example, on a test of musical performance, portfolio, or typing or foreign language speaking, the judges can observe actual performance or listen to recorded performance or parts of it and form their judgements. Where the products are large, a selected sample of the examinee can be arranged for observation.

If the actual test is the best available indication of the examinees' skill, then, the judges should base their judgement on the observation of the examinees actual test performance but not the test scores. For instance, a test on essay writing which test examinees writing skills, the judges can read these essays and form their judgements.

iii. must reflect the examinee's skills at the time of testing not at a future date. This means that judgements should be made as close to the time of testing as possible, except where the performance is recorded.

- iv. must reflect the judges true opinions. By implication, judges should not be

influenced by possible personal gains that they make lenient or strict judgements of examinees' skills.

Basically, passing scores form the basis for marking two or more places on a score scale where important decisions are to be made. For instance, in (NAEP) achievement level can be 'basic', 'proficient' and 'advanced'. Livingston and Zieky (1982 p.35) contend that once the examinees have been divide into two or three groups based on the judgement of their knowledge and skills, then you can consider all the examinees with a particular test score and ask, "Are the majority of them qualified or unqualified?" Most of the examinees with very high scores will be in the "qualified" group. As you go down the score scale, the proportion of the test-makers who are "qualified" will decrease. At the lowest score level, the "unqualified" test-makers will outnumber the "qualified" test-makers.

Basic Steps in Using the Contrasting Method

The following are the essential steps when using this method as explained by Cizek and Husband (1997):

1. Select the judges
2. All the judged examinees take a test based on the domain of interest.
3. Define adequate and inadequate levels of the knowledge and skills tested.
4. Select the sample of the examinees whose skills will be judged (skip step if the population is small and if you can get judgement on all the examinees).
5. Obtain the test scores and the judgements of the test-takers you have selected. Do not let the judges know the test-takers scores.
6. Divide the test-takers at each score level into "qualified" and "unqualified" groups on the basis of the judgements. Compute the percentage of the test-takers at each score level who are in the "qualified" group. Indicate the percentage judged to be passing at each score point or interval. (If you do not have several test-takers at each score level, it is suggested that you combine score levels into larger intervals before you do this calculation).
7. Use a "smoothing" method to adjust the percentages you have computed if n is small (see Table 3).
8. Choose the passing score on the basis of the "smoothed" percentage. Note that if a large population is used, the percentages obtained may most likely rise steadily.

Table2: How to Smoothing the Percentages Obtained Using Life Data

Test Score	Number of Test Takers (<i>Examinees</i>)			Percent Qualified
	Qualified (<i>Passing</i>)	Unqualified (<i>Failing</i>)	Total	
96-100	5	0	5	100
91-95	3	1	4	75
86-90	6	2	8	75
81-85	18	1	19	95
76-80	17	3	20	85
71-75	15	10	25	60
66-70	20	9	29	69
61-65	7	8	15	47
56-60	6	17	23	26
51-55	2	9	11	18
46-50	6	8	14	43
41-45	2	4	6	33
36-40	2	12	14	14
31-35	0	7	7	0
0-30	0	3	3	0

Source: Kastrinos W., and Livingston S. A.. The Development of a Proficiency Examination for Dental Auxiliaries (Princeton. N.J Educational Testing Service. 1979). p 64.

Smoothing the Percentages

There are several methods for smoothing the obtained percentages, while some are simple, others are complex and need the application of a computer and assistance of a statistician. It has been found that some of these methods give almost the same results and in the above method, (Table 2) the data set was used by Kastrinos and Livingston (1979) cited in Livingston and Zieky (1982) is a good example of a small sample size. Smoothing the percentages obtained is essential according to Cizek and Husband (1997) if the sample size is small where N is less than 30. Note that in step 6 of the outlined procedure, the percentage of the examinees at each score level of those qualified is computed. You will observe that the percentages as computed of the number of qualified Table 2 are high and low i.e. 'zigzag'. The number drops from 100 to 75, rises to 95, drops to 60 etc. These unsteady percentages have to be smoothed. Smoothing means adjusting the percentages obtained. This is best done by the graph

method called 'graphic smoothing.' This is shown in Figure 1 as a follow-up of the data in Table 2, and it indicates a graphic smoothing of the percentages obtained.

This graphing method makes the smoothing method easy and as explained by Livingston and Zieky (1982), the basic idea about methods of smoothing is that the judgments of the test-takers (examinees) at each test score level tell something about the knowledge and skills of test-takers at near-by test score levels. To use this smoothing method, plot a distribution of the obtained percentages with the score on the X-axis and the percentage qualified on the Y-axis (see fig. 1). To adjust the graph draw a smooth curve along that of the observed percentages (zigzag pattern) by getting the curve close to each point as possible. Where there are large variations of observed scores, the curve should be drawn close as much as possible to those points indicating large numbers of examinees.

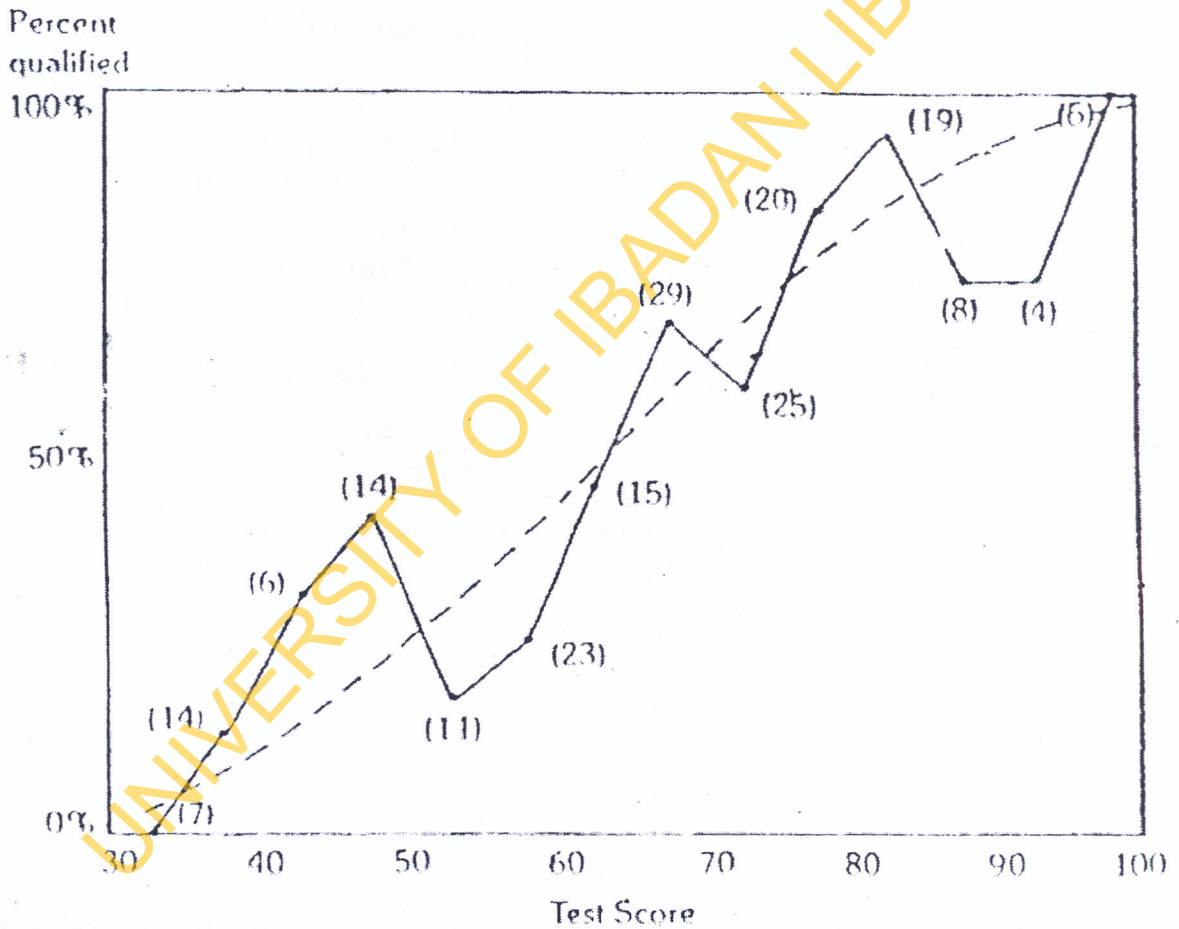


Figure 1: Example of a Graphic Smoothing

(Numbers in parentheses indicate the number of test takers at each test score level)

Another method of smoothing observed percentage is called 'moving average'. Using the same data in Table 2, the score interval and qualified number are extracted (see Table 3). This method involves replacing the "observed percentage of each test score

level (interval) with the average of the percentages for that score level and two other (adjacent) score levels” (Livingston & Zieky, 1982). For instance, the smoothed percent qualified at score level 86-90 give the average of the percentage for test scores level 81-85, 86-90, and 91-95 which when computed give averages of 95, 75, 75, respectively (see Table 2) and a smoothed percentage qualified of 82%. Refer to Table 4 for how this was arrived at and also the others. You will observe that the smoothed percentages computed with this method are 'smoother' than with the earlier method. One weakness of this method, however, is the inability to obtain estimates with it for percent qualified or pass at the highest or lowest score levels, (e.g. 96-100) even though this does not invalidate its use (Livingston & Zieky, 1982). As a rule, these authors say it is normal to indicate the method of smoothing at the outset so as to be within safe waters.

Table 3: Smoothing by “Moving Average”

Smoothed Test Score Percent Qualified	Number of test -takers		
	Qualified	Total	
96-100	5	5	
91-95	3	4	82%
86-90	6	8	87%
81-85	18	19	87%
76-80	17	20	78%
71-75	15	25	70%
66-70	20	29	61%
61-65	7	15	and
so on			

Note: This method can not be used to estimate the percent qualified at the lowest and highest test score levels.

The contrasting group methods can be used to set standards on multiple cut scores. If for instance, one is interested in setting standards for the 'basic', 'proficient' or 'advanced' students or candidates, all one needs to do is to provide judgement on the probability that each of the three examinees would answer the questions correctly. The remaining processes are the same as earlier explained.

Selecting the Cut Score with the Contrasting Group Method

As earlier mentioned in the paper the cut score is the passing score. In choosing the score for which the percent succeeding or passing has been smoothed psychometricians agree that it should be exactly at 50 percent i.e. where the number of qualified examinees are just equal to the number of unqualified (Cizek & Husband, 1997 and Livingston & Zieky, 1982). At any score lower than this, an examinee may most probably be judged as failing or not advancing than passing or advancing and the reverse is most likely to be the case. In the example used in this paper the cut score is 65 approximately, (refer to Fig.1). Explaining the reasoning behind choosing 50% chance as the yardstick Livingston & Zieky (1982) contend that because the two types of possible wrong decision about an examinee are equally serious. This decision rule is similar to statistical decision theory of the type I and type II errors in accepting a null hypothesis instead of rejecting it, and vice-versa.

Selecting the Sample

To obtain the passing score or cut score; it is at the point where the number of succeeding or passing score is just equal to the number of not succeeding score or failing score. Psychometricians agree that it is better to use the whole population for standard setting but occasions may demand that the population of examinees is so large that a sub-set (sample) is taken. The sample could be chosen randomly before test is written, or better still after the test is written so that the selected sample will be guided by an even spread of scores to cover all the possible range of scores, Livingston and Zieky (1982) suggest that in a 100 question test, 10 examinees should be chosen from each 5 point score interval e.g. 31-35, 36-40, 41-45 etc. They cautioned that whatever sample is selected at each score level should be representative of all the examinees at their score level.

Advantages and Disadvantages of Contrasting Group Method

One advantage of the contrasting group method is that it is not difficult to use because people judge other people's skills in the society especially for educational or occupational purposes. However, one disadvantage is that classifying students on the basis of an external criterion (teacher) independent of a test, creates the problem of classification error due to non-representative sampling frame.

Guidelines for Setting Performance Standards

Some researchers have suggested guidelines to follow in setting and/or reporting performance standards (Cizek 1996a, 1996b; Hambleton & Powell, 1983; Livingston & Zieky, 1982; Plake, 1997 in Hansche, 1998). An updated list of guidelines to follow for setting performance standards when using test based methods as outlined by Hansche (1998) are:

1. The importance attached to the classification of students to performance categories should substantially drive the effort that is committed to producing defensible standards and the effort should include compiling evidence to support the validity of the standards.
2. The design of the standard setting process should be influenced by the panelists and their backgrounds; test length and test item formats. For example, inexperienced panelists may require substantial amounts of training, long tests may require that sub-groups of panelists be formed, with each group assigned a different portion of the test etc. (Hambleton, Jaeger, Plake & Mills, 1998).
3. With important standard setting initiatives, the full process should be field tested prior to using it operationally to eliminate errors through careful conduct and evaluation field tests of the standard setting process.
4. The selection and the number of panelists should be given considerable attention.
5. The panelists should take the test (or a part) under test like condition to familiarise themselves with the test and its administration, and to enhance the validity of the resulting standards.
6. The panelists should be thoroughly trained in the standard setting process and be given practice exercises. This will boost their confidence in the process and make the standards that are produced acceptable.
7. It is often desirable to provide an opportunity for panelists to discuss their first sets of ratings with each other prior to providing a final set of ratings. This lead to more informed and more defensible standards because many sources of error due to misunderstandings, carelessness, inconsistencies and mistakes can be removed.
8. The full process of standard setting should be documented so that it is available if challenges to the performance standards arise. Every detail, from who determined the composition of the panel, to the choice of method, to the resolution of differences among the panelists etc should be documented for possible use later.

Conclusion

This paper examined two methods of standard setting and obtaining cut-score for achievement tests and performance assessment. These are the Angoff (1971) method

and the contrasting group method. These methods are been proposed in view of their objectivity in obtaining defensible standards and have been much researched into and they have established themselves in the standard setting arena. These methods are been proposed to the public examination bodies based on the afore-cited characteristics. These could be used in addition to the grade fixing methods already in use, as such, the proposal is not entirely advocating for a scrap of the existing grading system ~~be~~ but as complements. This will make for better objectivity of the outcome of the results made available to the stakeholders and the general public.

References

- Angoff, W.H. (1971). *Scales, Norms and Equivalent Scores*. In R.L Thorndike (Ed), *Educational Measurement* (2bd ed., pp.508-600). Washington D.C.: America Council on Education
- Brandon, P.R. (2005). Using test standard setting methods in educational program evaluation: Addressing the issue of how good is good enough: *Journal of Multi-Disciplinary Evaluation*. (JME 3). <http://evaluation.wmich.edu/jmde/2007>
- Cizek, G.J. (1996a). Setting passing scores. *Educational Measurement: Issues and practice*, 15, 20-31.
- Cizek, G.J. (1996b). Standard setting guidelines. *Educational Measurement: Issues and practice*, 15, 13-21.
- Cizek, G. J. and Husband, T. H. (1997). A Monte Carlo investigation of the contrasting groups standard setting method. Paper presented at the Annual meeting of the American Educational research association. (Chicago, IL. March, 24-28).
- Hambleton, R. K. (1998). Setting performance standard on achievement tests: Meeting the requirements of title one in Hansche, L.N. (1998). *Handbook for the Development of Performance Standards: Meeting the Requirements of Title I*. U.S. Department of Education and the Council of Chief State School Officers.
- Hambleton, R.K and Powell, S. (1983). A framework for reviewing the process of standard-setting. *Evaluation and the Health Professions*, 6, 3-24.
- Hambleton, R.K., Jaeger, R.M., Plake, B.S. and Mills, C.N. (1998). Handbook on setting standard on performance assessment. Washington, D.C: Council of Chief State School Officers.
- Hansche, L.N. (1998). *Handbook for the Development of Performance Standards: Meeting the Requirements of Title I*. U.S. Department of Education and the Council of Chief State School Officers.
- Livingston, A. and Zieky, M.J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service,
- Plake, B. S. (1997). Criteria for evaluating the quality of a judgemental standard

setting procedure: What information should be reported? Unpublished paper.

Roach, A. T. and Elliot, S. N. (2004). Alignment analysis and standard setting procedures for alternative assessment. Wisconsin Centre for Education Working Paper.

Girand, G., Impara, J. C. and Plake, B. S. (2000). A qualitative examination of teachers' conception of the just competent examinee in Angoff (1971) workshop. A paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, L. A., April 24-28, 2000).

Kastrinos, W. and Livingston, S. A. (1979). The Development of a Proficiency Examination for Dental Auxiliaries (Princeton, N. J. Educational Testing Service), P. 64.

Appendix 1

Adaptation of Hambleton (1998) Description of Examinees Performance Abilities (Knowledge and Skills)

Basic: The basic examinee perform all items requiring recall of information, do simple calculations, recognise changes in events have ability to manipulate equipment, read and complete simple information on tables, graphs. Can observe relationships between two objects and make description of changes that are visible etc.

Proficient: Candidates can recall, predict how things will occur, comprehend, explain phenomena, organise data, and interpret charts, graphs; recognize and explain changes of events, analyse effects. He is able to comprehend relationship between and among objects, things and phenomenon. Can also manipulate equipment, complete, read, locate and interpret complex data tables & graphs, and have ability to describe phenomena.

Advance: An advance examinee has ability to make analysis, explain implication and changes, identify interactions and relationships among phenomena. Can apply events to new situations. Can interpret data and infact perform all activities that require higher order cognitive skills that involve synthesis and evaluation; is able to calculate and predict effect of one phenomenon on another and test hypothesis.