

Editorial

Degrading Standard of Examinations in
Many Higher Institutions

Emmanuel Oluwatobi SALAWU

ABSTRACT [ENGLISH/ANGLAIS]

Nil

Keywords: Examinations, standard, universities, specificity, sensitivity, PPV, NPV, pretesting, blinding, marking scheme, rubrics

RÉSUMÉ [FRANÇAIS/FRENCH]

Nil

Mots-clés: Examens, standard, les universités, la spécificité, la sensibilité, VPP, VPN, pré-test, l'aveuglement, le barème de correction, des rubriques

Affiliations:

Departments of Physiology Faculty of Basic Medical Sciences, College of Medicine, Ladoke Akintola University of Technology, Ogbomosho, NIGERIA; and
School of Public Health, SRM University, Kattankulathur, Tamil Nadu, INDIA

* Email Address for Correspondence/ Adresse de courriel pour la correspondance: seocatholic@gmail.com

Accepted/Accepté: February, 2011

Full Citation: SALAWU Emmanuel Oluwatobi, Degrading Standard of Examinations in Many Higher Institutions. World Journal of Life Sciences and Medical Research 2011;1:1-3

THE PROBLEM OF POOR STANDARD OF
EXAMINATIONS IN MANY INSTITUTIONS

The poor standard of examinations in many institutions of higher learning always breaks the heart of those who know the value of education and who are keen to identify “good” students. But the problem of poor qualities of examinations seems not to call the attention of the stake holders especially in many higher institutions in majority of the developing countries, and, most importantly, in many private higher institutions. Perhaps, they all forgot the exact/main purpose of examinations, and/or the essential characteristics of good examinations – High Specificity, High Sensitivity, High Accuracy, High Precision, High Positive Predictive Value, High Negative Predictive Value, Low False Positive Rate, and Low False Negative Rate. Ideally, an examination can only be said to be of high standard and quality only if it at least fulfils these essential eight characteristics. But for the past few years, it has been really very hard to find (even) just a few higher institutions that really do put these essentially important eight characteristics of a standard examination into consideration.

The important questions that would come to everybody’s mind would be: “Of what importance is an examination that is of poor standard? Would it not be

better not to conduct examinations at all than to end up making wrong decisions based on the outcomes (results) of poor examinations? Would it not be an unacceptable Type I Error if it happens that an examination falsely claims that someone is worthy of a position, when, in the actual fact, the person is not worthy of it? ... Can the academic world at large be ready for the negative consequences of unreliable, inaccurate, unspecific, insensitive examinations?” The answers to these questions make clear the extreme need for quality and standardized examinations.

CHARACTERISTICS OF A GOOD EXAMINATION

To swiftly explain the characteristics of a good examination enumerated above, using a highly simplified example in the following 2 X 2 contingency table (Table 1) would help a lot. Let us assume that 100 candidates/students sat for an examination, and their marks are distributed as follows, such that $a + b + c + d = 100$ (see Table 1).

Specificity, Sensitivity, False Positive Rate (FPR),
and False Negative Rate (FNR)

Specificity is a measure of the proportion of those who (in the actual fact) do not “deserve” to pass who were

correctly identified by examination and thus did not pass the examination. In this example, specificity would be $d/(b + d)$. Since it would be highly unacceptable that a candidate who do not deserve to pass an examination eventually passes (i.e. Type I Error), then a very high specificity is always an extremely important characteristic of any good examination. This invariably means that quality examinations would have very low FPR: $b/(a + b)$. For example, Xi [1] made it implicit in his article in (TOEFL iBT Research Report) a publication by Education Testing Services (ETS).

TABLE 1

Table 1 is a contingency table showing the result of an examination against the actual fact*

| | | Actual fact** | | |
|--------------------------------|--------------|-----------------|------------------------|---------------------|
| | | Deserve to Pass | Do Not Deserve to Pass | Total |
| Outcome /result of examination | Passed | a | b | a + b |
| | Did Not Pass | c | d | c + d |
| | Total | a + c | b + d | a + b + c + d = 100 |

* **a** and **d** (in bold fonts) are correct decisions, while **c** and **d** are errors (Type I Error and Type II Error, respectively)

** "actual fact" refers to what would be obtained if there was an extremely standardized examination that would at all times and in all situations make correct decision.

Whereas sensitivity is a measure of the proportion of those who (in the actual fact) deserve to pass who were correctly identified by examination and thus passed the examination. In this example, sensitivity would be $a/(a + c)$. Of course, high sensitivity is a highly importance characteristic of a good examination and is of high importance in placement tests, so that a candidate who deserves to be offered a position based on his/her ability to succeed in an examination would not be missed out eventually, and thus preventing Type II Error. This also means that even FNR ($c/[c + d]$) has to be very low for an examination to be said to be of high quality. Therefore, both specificity and sensitivity must be high, while both FPR and FNR would be preferred low for an examination to be said to be quality and standard.

Positive Predictive Value (PPV), Negative Predictive Value (NPV), Accuracy, and High Precision

Here we say that PPV measures the proportion of those who passed the examination, who in the actual fact deserve to pass – that is $a/(a + b)$. A high PPV will directly reduce the chances of committing a Type I Error, and will thus prevent a candidate who does not deserve to pass an examination from passing. In a similar way, we say that NPV measures the proportion of those who did not passed the examination, who in the actual fact do not deserve to pass – that is $d/(c + d)$. Therefore, a high NPV will directly reduce the chances of committing a Type II Error, and will thus prevent a candidate who deserves to pass an examination from not passing.

So many other highly important measures by which the quality and standard of examinations can be assessed have been adequately documented by Livingston and Zieky [2], and interested readers are strongly encouraged to refer to their publication.

THE POSSIBLE SOLUTIONS

Even though the problems of poor quality examinations seems to be huge, and the characteristics/requirements of a good examination seems to be highly difficult to meet, still the problem has got solutions and the many requirements of good examinations can in the actual fact be highly satisfactorily met provided that the necessary procedures are diligently followed and all necessary precautions are taken right from the planning to the conduction of the examination, and even down to the marking of the answers/awarding of grades. In this note, higher institutions may want to consider incorporating the following steps/tools in their examination system.

Pretesting (Potential) Examination Questions

Like every other measurement tools, (potential) examination questions need to be tested (and necessary adjustments made) prior to using them for assessment. This is a must to do step; else the examination might be at the risk of not having virtually all of the essential characteristics. In fact, it is only through pretesting that the degree to which examination questions possess the above characteristics can be assessed. Readers who are interested in extensive details regarding Pretesting (Pilot

Testing) should consult the available literature [2, 3, 4, 5, etc].

Allocating Optimum Time

Examinations must be conducted with appropriate timings, allowing about the top 50% of the students to completely finish supplying all the required answers to all the examination questions. The basic idea here is that a tool that is not completely administered might end up losing considerable part of its desirable characteristics/features, or it would become the essential that the examination scores be completely based on a scale that would be a posteriori-ly determined by the relative performances of the candidates. The methods to be employed when such (relative-) performance-based examinations are necessary have being diligently documented by McKinley et al. [6].

Blinding the Markers/Examiners

Even though (potential) examination questions might have been pretested (pilot tested) and found to be good with respect to the essential characteristics, and examinations might have been conducted using the optimum timings, still the quality of an examination might go down the drain if markers/examiners are not blinded, since this may introduce bias and one would not be able to guarantee the validity and reliability of the resulting examination results.

Using Marking Schemes (Marking Guides and Rubrics)

Marking of examination answers should by no means be subjective, rather it should be objective. To achieve this, there must always be some schemes (marking schemes/guides) guiding the awarding of marks based on the accuracy, quality/validity, quantity, and clarity of answers supplied by the examinees. This making schemes must as well have been tested and reviewed as appropriate and must have been found to satisfy quite a number of essential conditions/characteristics. Furthermore, they may often have to come in various forms based on the nature of the questions for which they are made. For example, marking schemes for multiple choice questions can be as simple as a list containing question numbers against their respective correct answers/options; while marking schemes for qualitative/descriptive questions can be a little complex and may often be better made in form of rubrics.

Some of the most important advantages of the using the marking schemes would be that it increases the validity and reliability of examination results, and considerably reduces bias and many other types of error.

Vetting and Validating

Being humans we are liable to make mistakes and even sometimes we somehow (perhaps unintentionally or even intentionally) deviate from the laid down rules. For these and many similar reason there should always be some sort of vetting of marked answers scripts by some superior colleagues so as to ensure that the Marking Schemes (Marking Guides and Rubrics) have been carefully followed. In an event that it is found that the marker/examiner (perhaps the lecturer) had deviated grossly from the Marking Schemes sets of serious actions must always be taken to ensure all such anomalies are corrected and, more importantly, that such do not happen in the future. The actions might involve asking marker/examiner to carefully remark all the answer scripts and that bring it back for re-vetting. It also might be important to state that repeated deviation from the appropriate marking guide is a very big offence: big enough to cost such careless marker/examiner (perhaps the lecturer) his/her job.

CONCLUSION

This editorial concludes by charging the stake-holders (most especially, the examination committees of Universities, and other institutes that conduct examinations) to rise to the challenge of poor standard of examinations and its impending dangers. In fact, they cannot afford to overlook the need to ensure that the examinations they conduct possess the entire essential characteristic futures (starting from when the test questions are being developed to the awarding of marks), and are therefore of the appropriate standards. To achieve this, pretesting (of potential) examination questions, allocating optimum time for the examinations takers, blinding the markers/examiners, using marking appropriate schemes (marking guides and rubrics), as well as vetting and validating would have to be adequately considered always.

REFERENCES

- [1] Xi X. Investigating the Criterion-Related Validity of the TOEFL Speaking Scores for ITA Screening and Setting Standards for ITAs. TOEFL iBT Research Report 2008. ETS, Princeton, NJ.
- [2] Livingston SA, Zieky MJ. Passing scores: A manual for setting standards of performance on education and occupational tests. Princeton, New Jersey: Educational Testing Service 1982.
- [3] Pretest Guide. University of Cambridge ESOL Pretesting. 2008. Accessed from: http://www.cambridgeesol.org/assets/pdf/what-we-do/pretest_guide.pdf . Last Accessed: January 2011.
- [4] How ETS develops test questions. 2010. Available from: http://www.ets.org/s/understanding_testing/flash/how_ets_creates_test_questions.html . Last Accessed: January 2011.
- [5] Collins D. Pretesting survey instruments: An overview of cognitive methods. Quality of life research 2003;12(3):229-38.
- [6] McKinley DW, Boulet JR, Hambleton RK. A work-centered approach for setting passing scores on performance-based assessments Evaluation and the health professions 2005;28(3):349-69.

SOURCE OF SUPPORT/ACKNOWLEDGEMENT

Nil

CONFLICT OF INTEREST

Nil

How to Submit Manuscripts

Since we use very fast review system, and since we are dedicated to publishing submitted articles with few weeks of submission, then the easiest and most reliable way of submitting a manuscript for publication in any of the journals from the publisher Research, Reviews and Publications (also known as Research | Reviews | Publications) is by sending an electronic copy of the well formatted manuscript as an email attachment to rrpjournals@gmail.com .

Submissions are often acknowledged within 6 to 24 hours of submission and the review process normally starts within few hours later, except in the rear cases where we are unable to find the appropriate reviewer on time.

Manuscripts are hardly rejected without first sending them for review, except in the cases where the manuscripts are poorly formatted and the author(s) have not followed the instructions for manuscript preparation which is available on the page of Instruction for Authors in website and can be accessed through <http://www.rrpjournals.com/InstructionsForAuthors.html> .

Research | Reviews | Publications and its journals have so many unique features such as rapid and quality publication of excellent articles, bilingual publication, some of which are available at <http://www.rrpjournals.com/uniqueness.html> .