# The Importance of Validity and Reliability on Computer Based Assessments.

**Bhaugeerutty Vinod Sharma\***

*PhD Student, Open University of Mauritius, Mauritius.*

***Corresponding Author:*** *Bhaugeerutty Vinod, Sharma PhD Student, Open University of Mauritius, Mauritius. E-mail: vsharma471@yahoo.co.uk*

**ABSTRACT**

*As expectations have increased for student around the world, more and more schools are being asked to design new means of assessments based as part of their educator evaluation process. Thus, it is necessary to include technological devices in educational evaluations of students. So, accepting technology in assessment is different with accepting it in learning environment especially in a Mauritian context where the only criterion for going upper level is failing or passing on final exams. With the introduction of tablets in schools, further investigations are needed to ensure the validity and fairness of testing administration mode in comparison with traditional one. These terms, validity and reliability, can be very complex and difficult for many educators to understand.*

**Keywords:** *ICT - Information Communication Technology, CAIE - Cambridge Assessment International Education.*

## INTRODUCTION

Assessment is regarded as a key part in the educational system as it requires a methodical approach of gathering evidence in assessing student's learning. However, the approach has changed since I started teaching. Initially assessment was considered as a method to give a student a grade by measuring the knowledge acquired by the student, but nowadays assessment is a strategy to promote a better acquisition of concepts and skills as well as to encourage a more effective teaching and learning process**.** So educators can evaluate whether a student has learned what they are expected to learn by securing valid and reliable information through various assessment methods. According to Gagné (1985), assessment method chosen would depend on the three learning domain, namely cognitive, affective or psychomotor domains.

## PURPOSE OF THIS PAPER

The purpose of this document is to reflect on the importance of computer based assessments upon the key assessment principles of validity and reliability. This paper will first explain what is computer based assessments, before considering the significance of having a valid and reliable assessment. At the end, a conclusion will be draw as to these impacts.

## WHAT IS COMPUTER BASED ASSESSMENTS

In their literature review of "E-assessment" Ridgway, Mccusker and Pead define computer based assessment as "processes involving the implementation of ICT for the recording, transmission, presentation and processing of assessment material" (Ridgway, Mccusker and Pead, 2007, p.41). But according to Brown, Race and Bull (1999), the computer based assessment is a facilitator of formative assessment. The computer based assessment has the ability to provide feedback for each individual learner, with appropriate tools for examining data, can provide feedback on each learner's strengths and weaknesses in relation to their replies to assessment items. This is definition could not be covered within the scope of this assignment as it is broad. Hence, this document will highlight impacts of computer based assessments in improving validity and reliability. For example, computer based assessments allow learners to take the assessments at anytime and anywhere. Having an automated marking element of computer based assessment has the possibility to offer timely feedback to enable learner to engage in self-assessment. However, feedback from computer based assessment by itself is unlikely to develop the self-assessment skills of learners, as Black and William point out:

- For formative assessment to be productive, learners should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve. (Black and William 1998, p. 10)

## VALIDITY AND RELIABILITY

Luoma (2004) highlighted that validity and reliability are the two features for a quality computer based assessment. Validity refers to how well an assessment measures what it is supposed to measure, while reliability is the degree to which an assessment tool produces stable and consistent results.

### Types of Validity

I am aware of the different types of validity but I appreciate that there is much more to understand, it is very useful to look at computer based assessments with a focus on validity, (Cambridge Assessment, 2009). As part of this Cambridge Assessment Course A101: Introducing the Principles of Assessment, classifies different aspects of validity. The three main types of validity that are most appropriate to classroom assessments are face validity, content validity and construct related validity.

Face validity refers to the extent to which the assessment 'looks right' (Cambridge Assessment Network, 2019). Would a reasonable look at the assessment and say 'yes, that seems fair'? It is vital in certifying that the assessment contains some important features like fairness, practicality and manageability. These allow the assessment-takers to take the assessment under optimum conditions, so as to endure them to try their best on an assessment. Thus a computer based assessment that appears to be other than what it appeals to be evaluating without face validity may persuade students from proceeding with the assessment. Therefore, establishing whether an assessment occupies face validity does not require the opinion of expert.

In regards to face validity, content validity requires support from expert as it is essential to determine how closely the content and coverage of the computer based assessment matches the content and coverage of the syllabus/specification it is designed to assess (Cambridge Assessment Network, 2019). The expert should look into whether the assessment content is representative of the skills that are supposed to be measured. This involves looking into the steadiness between the syllabus content, the assessment objective and the assessment contents. If the assessment contents cover the assessment objectives, which in turn are representative of the syllabus, it could be said that the assessment possesses content validity.

The use of computer based in an assessment has an impact on construct related items.

Construct-related is the extent to which the assessment measures what the specification or syllabus says it is intended to measure, example, whether it assesses the abilities which are the focus of the course (Cambridge Assessment Network, 2019). As the use of computer to deliver an assessment may introduce measurement of assessment taker characteristics that are irrelevant to the construct measured. Examples of such construct irrelevant variance include computer proficiency, computer platform and interface effects, speediness and assessment anxiety. For example, a computer based assessment paper is absolutely not a valid tool for measuring mathematics skills. On the other hand, an algebraic assessment is to some extent a valid tool for mathematics skills because the capability to solve problems is a feature of a person's mathematics proficiencies. Yet, the algebraic assessment is not highly valid because mathematics skills are not restricted to the aptitude of writing solve algebraic questions only. Hence, to make the assessment more valid, other criteria of mathematics skills must be included in the assessment. If the assessment is valid and reliable, a student who demonstrates good mathematics skills on that particular assessment should also do equally well on other mathematics assessment of similar content and objective. In other words, students do not just need algebraic or computer based skills to solve mathematics problems on a computer based assessment. To summarise, the decision on what to include in an assessment will depend on what the content of the syllabus is, as well as what the assessment objectives are. It is of great significance for educators to understand that the extent of assessment validity depends on the assessment's coverage of the different objectives, which, eventually, depends upon the syllabus.

Validity is often connected with the principle of reliability, and it is that principle that will be looked at next.

### Types of Reliability

Reliability is referred to as "the consistency of the assessment, for example, "if an assessment

is reliable then whether the assessments were repeated on another day in different places or using different markers, the same outcomes are measured in the same way" (Cambridge Assessment, 2009, p.14).

According to Black (1999), there are three types of reliability that are most relevant to computer based assessments, namely: internal consistency, inter-scorer and intra-scorer reliability.

The internal consistency is concerned with the uniformity of the objectives among different items of the assessments. For example, in an ICT assessment which is supposed to evaluate students' ability to solve an algorithm. In this case, the question of internal consistency refers to the answer to the question: are the different items measuring he same competencies (for example, students' ability to define the variables or the programming constructs), or are the different items measuring something else entirely or others besides the stated objective?

The inter-scorer reliability means the consistency between the grades allocated by different educators. The main issue about inter-scorer reliability could arise when the same quality of answers is given different grades by different educators. On the other hand, it refers to grades given by the same educator on different occasions. An example of intra-scorer reliability at stake is when an educator gets tired of marking and starts to give lower marks as time goes on. So consistent grading is essential in order to ensure the reliability of assessment.

Scorer reliability can be enhanced by a marking scheme or a scoring rubric that is created in advance and used to support educators in scoring answer scripts.

As we can see that there are several factors that affects reliability but the four most important ones are the length of the assessment, types of questions and quality of questions, thirdly on administrative conditions and finally the assessor reliability. I would focus on assessors because they have a major role to play in maintaining consistency especially with marking. Thus the computer based assessments can support the abolition of assessor bias, through automated marking which means that the marking would be consistent regardless the place and time the assessment is being carried out where the assessor would be the computer.

Another way to eliminate bias in a computer based assessment is to have a robust mark scheme for them where it is possible and also

ensure they are taking through standardization where the qualification requirements are explained in details so they can have good understanding of the expectations. There should be ongoing training for the assessor to keep them abreast of the qualification requirements, observations and performance feedback of their practice.

The use of computer based assessment to automate and standardize assessments can positively impact both validity and reliability. Given the interrelated nature of these principles it is reasonable to expect that an impact on an assessment's validity and reliability would also impact its fairness.

## FINDINGS

How valid and reliable is the outcomes from every session?

To achieve a certain degree of validity and reliability, the assessment and evaluation process has to be looked at in its totality, and the factors that may affect validity and reliability need to be identified. Typical activities in the computer based assessments and evaluation process are as follows:

- Deciding on the computer based assessment's objectives;
- Designing and developing a computer based;
- Evaluating the computer based assessment;
- Administrating the assessment.

At each stage, something could be carried out to enhance the validity and reliability of an assessment.

The review of the issues and literature on computer based assessment validity and reliability indicate that the potential benefits and threats are identified. However, it is also clear that more research needs to be done before we can conclude that well-constructed computer based assessment promote validity and reliability. With technological advancement, computer based assessment has facilitated the building of bridges between psychometrics and cognitive psychology. The effects of incorporating information regarding cognitive components and assessor taker response time into scoring and construct validation needs to be addressed by the assessors.

## CONCLUSION

When integrating computer based assessments in teaching and learning, we need to evaluate the positive impact of reliability and validity. This can only be possible to analyse the extent to which

human error and bias are eliminated, thus ensuring that the computer based assessments are set, administered and automated marking by providing instant feedback. The assessment validity and reliability may be accomplished from the steps taken throughout the design and administration stages. When using computer based assessment it is vital that the assessment is generated randomly and need to be in line with the syllabus. For increased efficiency, educators may decide to work in teams to design and develop computer based assessments that can carry out adjustments to ensure that learners are treated fairly. Lastly, although following the recommended measures previously discussed does not provide a guarantee for a perfect and valid assessment, it can certainly help educators from getting it totally wrong.

## REFERENCES

[1] Black, T.R., Doing Quantitative Research in the Social Science: An Integrated Approach to Research Design Measurement and Statistics. London: Sage (1999)

[2] Cambridge Assessment, "The Cambridge Approach to Assessment", 2009. Accessed 22nd April 2019. https://www.Cambridgeassessm ent. org.uk/Images/cambridge-approach-to-assess ment.pdf#page=12

[3] Cambridge Assessment Network, "3.2.4 Different aspects of validity", 2019. Accessed 22nd April 2019. https://vle.canetwork.org.uk/ mod/ page/ view.php?id=6160

[4] Gagné, R.M., The Conditions of Learning and Theory of Instruction (4th edn). NewYork: Holt, Rinehart and Winston (1985).

[5] Green, A. (2014) Exploring Language Assessment and Testing. NY. Routledge.

[6] Luoma, S. (2004) Assessing Speaking. Cambridge: Cambridge University Press.

[7] McMillan, James H. (2004). Classroom assessment: principles and practice for effective instruction. Boston: Pearson/Allyn and Bacon,

[8] Messick, S. (1989). Validity. In: R.L. Linn (Ed.) Educational measurement. 3rd ed., 13–103. NewYork: American Council on Education. Online Library https://onlinelibrary. Wiley. com/ doi/ full/10.1046/j.13652923.2003.01594.x

[9] Ridgway, J.&McCusker, S. (2003). Using computers to assess new educational goals. Assessment in Education, 10, 3, 309–328.

[10] Ridgway, J., Mccusker, S., Pead, D. Literature Review of E-assessment. A NESTA Future Lab Research report- report 10. 2004. Accessed 29th May 2019. https://www.Researchgate.net /publication/30050842_Literature_review_of_e-assessment/download

[11] S. Downing (2003), Validity: On the Meaningful Interpretation of Assessment Data, University of Illinois at Chicago, Wiley

[12] Why use computer-based assessment in education?https://www.cambridgeassessment.o rg.uk/Images/472553-why-use-computer-based assessment-in-education-a-literature-review.pdf