



RESEARCH ARTICLE

ACCURATE MEASUREMENT AND EVALUATION OF STUDENTS' PERFORMANCE AND NATION BUILDING

*Dr. Monday, U Agbonkpolo

Benson Idahosa University, P.M.B. 1100, Benin City, Nigeria

ARTICLE INFO

Article History:

Received 19th April, 2017
Received in revised form
21st May, 2017
Accepted 14th June, 2017
Published online 31st July, 2017

Key words:

Nation building, Quality assurance,
Measurement and evaluation,
Segmented marking method.

ABSTRACT

The paper is a theoretical study on the problem of inconsistency and prospect in measurement and evaluation of students' performance in the foundation lying of nation building. The paper extolled the role of accurate measurement and evaluation of students' performance in quality (or safety) assurance in nation building. Clarifying the role of measurement and evaluation in nation building was not farfetched since the concept of 'building' is synonymous with construction whose key element is measurement. The task of 'construction' is the responsibility of skilled workers who are trained through the process of education. The inputs and outputs in the training system go through measurement and evaluation process. The outcome of the process forms the basis for making accurate decisions and policies for training the manpower which is strategic to nation building. And above all, measurement and evaluation report forms the basis of selecting prospective workers in different capacities. The marking of essay test is a significant aspect of measurement and evaluation in education and the focus of this study. The study thus periscopes the search for marking method in literature, that would yield improved measurement and evaluation of students' performance in essay tests. Giving all the merits and demerits of the different methods, the conclusion of this paper is that the segmented method of marking essay test with the requirement of pre marking training session for examiners and legible examinees' handwriting has the best comparative advantage. This would improve the accuracy of students' grade, which will in turn, make the decisions in education and recruitment exercises in the various establishments enhance the effectiveness in the building of the nation's institutions.

Copyright©2017, Dr. Monday, U Agbonkpolo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Dr. Monday, U Agbonkpolo. 2017. "Accurate measurement and evaluation of students' performance and nation building", *International Journal of Current Research*, 9, (07), 55258-55261.

INTRODUCTION

No building can be constructed without carrying out accurate measurement in line with the aim and specifications of the building. Having concern for safety, even the building materials are produced through scientific processes which are predicated on measurement and evaluation. Needless to say that quality assurance in building materials production is safety assurance because any building that cannot withstand environmental hazards will collapse and the casualty may be fatal. Thus safety assurance in building depends essentially on accurate measurement and evaluation. So, it is incontestable that planning and execution of projects without accurate measurement and evaluation data are the major causes of the failure of many projects in our nation building. In this scientific age every wise society identifies primarily the type of nationhood it wishes to build and then measures and

evaluates the required inputs in relation to the expected outcome and sets in motion the relevant machineries. Thereafter it evaluates the outcome at regular fixed intervals, to determine the effectiveness of the inputs, processes and products. Furthermore, the policy statements regarding inputs, methods and strategies are reviewed on each occasion. These policies could be reinforced by amendment or replaced as the case may be. The whole system is a continuous cyclical process. Nation building is the macrocosm of home building. It involves the employment of skills to harness and organize inputs. Thus nation building begins with the training of skilled men and women for the onerous task of making and laying the nation's building blocks. Education is the process of creating these personnel. It is indeed the building of the nation's foundation. The quality or safety assurance of this educational process is ensured by accurate measurement and evaluation of inputs, processes and the products of the system in relation to the national goals. Educational inputs consist of learners' entry behavior, personnel qualification and experience, curriculum offerings, quality of facilities and materials. The 'process' encompasses the methods, strategies and procedures while the

*Corresponding author: Dr. Monday, U Agbonkpolo,
Benson Idahosa University, P.M.B. 1100, Benin City, Nigeria

products refer to the graduates, who have been equipped with theoretical and practical skills to develop and transform all necessary inputs into a 'national home' for all Nigerians. Building a 'national home' is the common aim of every sector, ministry, department and worker. So the Nigerian graduate, irrespective of his or her specialization must be initiated into the 'national home' orientation with global outlook on the basis of international comparative advantage. Thus education is the foundation of the Nation's building. Effective education rest squarely on accurate measurement and evaluation. Therefore evaluation is an essential and continuous aspect of teaching and learning. It is the systematic process of identifying the decision options; selecting appropriate information collecting and analysing same for the purpose of creating data base needed by decision makers in selecting among alternatives (Okpala, Onocha & Oyedeji, 1993). Evaluation induces motivation for learning, which in turn produces good study habits. Evaluation enables the teacher to determine the extent to which behavioural objectives have been achieved. (Okpala, Onocha & Oyedeji, 1993). Similarly, accurate evaluation information on students' learning, helps teachers arrive at the correct solution to students' learning problem. It is useful in counseling services in schools in the areas of students' education programme, vocational interest, and study habits. It helps principals and the teachers take accurate decisions on students in respect of selection and promotion. It is useful for educational policy formulation, parental decisions on children's and wards' education, and also for decision making in the area of employment of labour (Mamta, 2004; Okpala, Onocha & Oyedeji, 1993). Marking of essay test is a major aspect of measurement and evaluation. This is the focus of this paper. Scoring of essay test is significantly affected by the problem of inconsistency. The following section throws more light on this problem.

Problem of Inconsistency in the Scoring of Essay Test

Having regards for the crucial role of accurate evaluation, a considerable effort is being made to ensure objectivity in scoring test. This is most pertinent in the scoring of essay tests. In external examinations, for instance, several steps are taken to create the enabling conditions for objectivity in scoring essay test. Despite the painstaking steps, it is a common fact that unacceptable variation between the original scores and vetted scores still exist. A sample of vetted sheets, for instance, taken by the researcher from the examination department of Edo State Ministry of Education, which had 90 students' scores showed that the variations between original scores and vetted scores ranged between 0 and 15. There is substantial evidence in literature of unreliable marking of the essay test in secondary schools and higher education. What are the sources of this inconsistency?

Sources of Inconsistency in the Scoring of Essay Test

Two major sources of inconsistency in marking are the restricted or the non-detailed marking scheme and some question formats that necessitate markers' opinion during marking thereby giving room to influences from subjective factors such as mood, fatigue, behaviour, handwriting, contrast effect, halo effect, among others. (Hout, 1990). As a result of this problem of inconsistency, some examination bodies as earlier mentioned resort to various screening exercises to ensure that only experienced markers participate in marking of scripts. This measure often results in the weeding of significant

number of potential markers leaving the enormous task to few hands which in turn creates the problem of increasing pressure on the enlisted markers. This is a situation, which may create the problem of fatigue, leading to loss of concentration, which potentially could increase marking error. Fatigue is the physical and mental stress from marking for several hours without rest or break to eat, which begins to take its toll on marker's concentration. This problem has attracted substantial research efforts with the aim of finding solution to the problem.

The Search for Methods of Reducing Inconsistency in the Scoring of Essay Test

Many measures for reducing these influences have been suggested and tested without significant result. For instance, Wolf, (1995), Lave and Wenger, (1998) recommended the use of exemplar material. Exemplars are examples of students' work which are used for scoring examination scripts instead of assessment criteria. However the works of Baird, Greatorex and Bell (2002, 2003) did not support the claim that the use of exemplar material alone could increase the marking reliability of essay test. The use of double marking method has also been recommended by some educationists like Smith, B., Sinclair, H., Simpson, J., Van Teijlingen, E., Bond, C., & Tylor, R. (2002);. Double marking was actually adopted by some examination bodies such as the GCE and CSE examination boards in England in the late 1970s. Though this method yields better reliability than the use of a single marker, the cost implication has made it difficult to use the method. (Meadows & Billington, 2005). The US and UK examination bodies have embraced e- marking as an alternative to the conventional marking method. However a couple of studies revealed small and inconsistent differences between the reliability of e - marking and the conventional method (Fowles, 2002, Raikes, 2002, Sturman & Kispal, 2003). Another measure was the adoption of the automated marking, that is, using the computer, to assess the mechanical features of candidates' responses (Cohen, Ben-Simon & Hovav, 2003) and in the marking of the short answer tasks format in science (Sukkerieh, Pulman & Raikes, 2003; Fowles, 2005). Although automated marking is reliable, the validity of this type of marking cannot be guaranteed because of the use of computer keys that will attract undue marks (Ridgeway & Mmusker, 2004). So it is uncertain whether the use of computer alone will be accepted in the foreseeable future (Meadows & Billington, 2005). Thus Lamprinou (2004) suggested the combination of human marker and a computer, and whenever there is a significant difference in the scores awarded a second human marker would do a second blind marking (having only candidates' numbers on scripts). This suggestion amounts to double marking and its cost implication.

Some experts also believe that the holistic and the impression marking methods give more room for subjective interpretation, so they recommended instead the analytic marking in which every facet of an essay question is duly allocated marks in the marking scheme (Huot, 1990, Vaghan, 1991;). Research findings, have confirmed the superiority of the analytic marking to the holistic and impression marking methods (Hout, 1990; Vaghan, 1991), but studies have shown that in analytic marking the marking reliability decreases with the increasing complexity of the essay (Delap, 1993 & Ucles, 2000). Many strategies have been tested by researchers yet the desired increase in marking reliability has not been achieved.

The search for a solution is still ongoing. One such expected solution is the use of segmented marking method.

The Hope of a More Reliable Making with the Segmented Scoring Method

Some educationists like Okpala, Onocha and Oyedeji (1993) identified the use of the whole script marking method, as the problem in marking reliability of essay test. Whole script marking is the marking of all the responses in one script before going to another script. This method of marking suffers from the problem of halo effect. Halo effect is the bias, in the marking of a particular item due to a previous impression the marker had of the script owner. This factor tends to make measurement error higher among the scores of upper achievers and that of the lower achievers. Therefore, these experts recommended the use of individual question marking method. Some researchers called this method segmentation (Bakker & Van Lent, 2003; Meadows & Billington, 2005). Segmented marking refers to the marking of one item across the scripts before engaging in the marking of another item. Segmentation or part marking affords the marker the opportunity to compare item response of each candidate to other candidates' responses to the same item there by reducing the level of inconsistency arising from the non - detailed marking scheme, halo effect and the effect of poor concentration likely to plague the marking of scripts by inexperienced markers. Nevertheless the objectivity of this method is not without a threat, as non-cognitive elements such as handwriting, and contrast effect may have higher compelling influence on students' scores because of the direct comparative scoring. Contrast effect is the tendency to underrate or overrate an essay response of an average quality when preceded by a series of responses of excellent quality or responses of poor quality respectively. In addition, if this method is more laborious and time consuming as some examiners claimed (Meadows and Billington, 2005) the effect of fatigue may offset a substantial portion if not all the gains that would have accrued from the method. The foregoing rationalization can only be resolved by empirical evidence and since this method of marking is being used by some examiners; it is pertinent to have empirical evidence and justification for its use.

The Search for Empirical Evidence of the Effectiveness of Segmented Scoring Relative to the Whole Script Scoring of Essay Test

Ofqual (2014) says: "There is currently limited empirical evidence available to enable a robust comparison of the relative merits of whole – script marking and item – level marking". This position is supported by experts such as Bakker & Vant Lent, 2003, Fowles, 2005, Meadows & Billington, 2005. Bakker and Van Lent (2003) expressed the lack of evidence on the relative effectiveness of segmented marking method in literature. They said that as e-marking becomes common there will be increased opportunities for empirical study of the belief that segmentation can 'add to the objectivity of the marking. Fowles (2005) discovered the persistence of this gap. This situation was reported by Meadows and Billington (2005), as follows: "Although part versus whole script marking is a topic that might be expected to have received research attention, Fowles (2005) found little reference to this aspect of marking". With respect to the e – marking version of this mode of marking – the digital separation of students' scripts into the different items and each

item marked by different markers online. The two methods were shown in the work by Ucles (2002) to be similarly consistent in mathematics on screen marking. He also showed that, two examiners in English Literature were similarly consistent in the method of SMM. Ucles (2002) showed in his study that the examiners used in his study, generally were least consistent in the SMM method on screen marking in English Literature. Fowles, (2005) reported the study carried out by the Assessment and Qualifications Alliance (AQA) in United States, in which markers who used the SMM method on CMI+ e- marking had 98.4% level of agreement, although this agreement was largely in short response questions. In a recent research carried out by Agbonkpolo (2014). He used 48 NECO examiners in a conference marking.

The researcher used the standard error of measurement (SEM) and the average mark change (AMC) to assess the scores awarded by the NECO examiners. The average marking time and regression analysis were also used. The examiners were randomized into two equivalent groups in terms of numbers, qualification and experience. One group used the SMM while the other group used the WSMM. The findings showed that both groups had similar levels of SEM (random errors) in the scores they awarded. Marking experience, tend to reduce SEM in SMM though the effect was not significant. However as marking experience increases, SEM reduces in SMM and increases in WSMM, making the former have less SEM than the latter among markers with more years of marking experience. The findings also showed that SMM had slightly higher AMC (random and systematic errors) than WSMM unlike the parity in their SEM. This shows that the systematic errors were slightly more in the SMM than in the WSMM, probably due to the fact that if a script was over - rated or under - rated with the SMM, this error was exported to subsequent scripts, because of the direct comparative approach. This error was due to poor examinees' handwriting although the effect was not significant. The study shows that as examinees' handwriting clarity increases, AMC becomes smaller in SMM than in WSMM. Finding also shows that SMM had greater AMT than WSMM though the difference was not significant. This was due to the fact that they were not used to SMM. Earlier, the work of Shohamy, Gordon & Kramer (1992) has shown that training matches experience in marking reliability. The implication of this therefore is that if pre marking training session is given to less experienced markers, to put them in the same pedestal with the more experienced markers, among whom the SMM was shown to be more reliable than the WSMM; then the hope of a better marking reliability with the use of SMM would become a reality. This could be boosted by warning examinees against illegible handwriting to minimize systematic error.

Conclusion

The foregoing has shown that if examination bodies adopt the use of SSM, with the requirement of pre-marking training session and legible examinees' handwriting, students' results will become more reliable. Thus students, teachers, educational administrators, parents and wards, entrepreneurs and the society will benefit as follows:

Students will gain better understanding as teachers' decisions with respect to instructional objectives, selection of contents and learning experiences will enhance teaching and learning when the feedback from evaluation becomes more objective.

The school teachers' job will be made easier because students' motivation for learning will be boosted when their grades are a true reflection of their efforts. Educational administrators will have less worry when there is more reliable marking of essay test. They will be more fulfilled, when evaluation based decisions in areas like selection, promotion, certification, and granting of scholarships, bring more success to the education system. Examination bodies will have improved public confidence if the results they award become more reliable. With enhanced objectivity of the reports from schools; Parents and Guardians will make better and rewarding decisions on their children's and wards' education. Finally, the operations of private and public organizations and the society at large will be enhanced as recruitment decisions bring increase in the level of productivity and success because of the accuracy of the evaluation reports upon which such decisions are based.

REFERENCES

- Agbonkpolo, M. U. 2015. Effectiveness of Segmentation and Whole script Marking in the Reliability of Scores in Essay Tests. (Unpublished PhD Dissertation)
- Bakker, S. & Van Lent, L. G. 2003. *National testing on - line: How far can we go?* Paper presented at the IAEA Conference, Manchester. Retrieved 5 January 2005 from <http://www.aqa.org.uk/support/iaea/papers.html>.
- Cohen, Y., Ben-Simon, A. & Hovav, M. 2003. *The effect of specific language features on the complexity of systems for automated essay scoring.* Paper presented at the 29th Annual Conference of the International Association for Educational Assessment, Manchester, October 2003.
- Delap, M. R. 1993a. *Marking reliability study in business studies (665)* AEB Research Report RAC/609.
- Fowles, D. 2005. *Literature review on effects on assessment of e-marking.* AQA Research Report.
- Huot, B. 1990. Reliability, validity and holistic scoring: What we know and what we need to know. *College Composition & Communication*, 41, 201-213.
- Laming, D. 1990. The reliability of a certain university examination compared with the precision of absolute judgments. *Quarterly Journal of Experimental Psychology*, 42, 239-254.
- Lamprianou, J. 2004. *Marking quality assurance procedures: identifying good practice internationally.* Report prepared for the National Assessment Agency.
- Lave, J. & Wenger, E. 1991. *Situated learning legitimate peripheral participation.* Cambridge: Cambridge University Press.
- Mamta, A. 2004. Curriculum reform in Schools: the importance of evaluation. *Journal in Curriculum Studies*, 36, (3), 361- 376
- Meadows, M & Billington, L. 2005. *A review of the literature on marking reliability.* Retrieved from W:\ Michelle\ Michelle Meadows\Quality of Marking\Lit Review\Review Sections\ A Review of the Literature on Marking Reliability.doc
- Ofqual, 2014. *Quality of marking, review of literature on item – level marking research.* Retrieved from www.ofqual.gov.uk
- Okpala, P. N., Onocha, C. O., & Oyediji, O. A. 1993. *Measurement and evaluation in education.* Jattu – Uzairue: Stirling – Horden Publishers
- Raikes, N. 2002. *On screen marking of scanned paper scripts.* Cambridge: University of Cambridge Local Examinations Syndicate (UCLES).
- Shohamy, E., Gordon, C., & Kramer, R. 1992. The effect of raters' background and training on the reliability of direct writing tests. *Modern Language Journal*, 76, (10) ,27-33.
- Smith, B., Sinclair, H., Simpson, J., van Teijlingen, E., Bond, C., & Tyalor, R. 2002. What is the role of double-marking? Evidence from an undergraduate medical course. *Education for Primary Care*, 1, 497-503.
- Sturman, L. & Kispal, A. 2003. *To e or not to e? A comparison of electronic marking and paper-based marking.* Paper presented at the 29th International Association of Educational Assessment Conference, Manchester, UK, October 2003.
- Sukkarieh, J. Z., Pulman, S.G. & Raikes, N. 2003. *Auto-marking: using computational linguistics to score short, free text responses.* Paper presented at the 9th Annual
- Vaughan, C. 1991. Holistic assessment: What goes on in the rater's mind? In L. H.- Lyons (Ed.), *Assessing second language writing in academic contexts.* Norwood, N.J.: Ablex Publishing Corporation.
