

Automation of Item Generation and Reporting Of Test Results: Experiences of a University in Nigeria.

Uwakwe Okereke Igbokwe

Faculty of Education

Abia State University, Uturu, Nigeria.

Uwakwe.igbokwe@yahoo.com

Abstract

This paper discusses the introduction of the automated test item generation of multiple-choice questions in the evaluation of students' academic achievement as well as the automation of the scoring and reporting of the outcomes. It traces the introduction to delays experienced in the release of results of examinations taken; sometimes unreliable questions set by examiners and biased scoring of scripts amongst other problems. The paper discusses the procedure adopted by the institution in the generation of items for the various courses that were amenable to the multiple-choice questions test format. Among these were the setting up of examinations committee and an examination centre; the training of examiners in all the faculties of the university; the development of question banks; the development of computer programmes for the automatic generation of items and the variants thereof; the development of programmes for scoring and reporting of results. The objective of early release of results was achieved as well as improved reliability of the examination question. A number of problems, however, beset the system. Among these were the reluctance of examiners to continue to enrich the banks and the sheer number of courses and scripts to be handled. There was also continued resistance to the change from the generally accepted and more commonly used essay question format to the multiple-choice questions format and the automated generation of items and scoring and reporting of test results. There was also the very potent problem of the regulatory institution's opposition. The paper concludes by expressing regrets that the match towards the use of computer-based tests has been truncated.

Keywords: Automation, Assessment, Item generation, Reporting of test results.

Assessment usually is done at the end of the teaching-learning process. Indeed, the process cannot be said to have been completed until assessment has been done. Assessment can be done for one of several reasons or purposes. Amongst these are to determine

- i. The effectiveness of a teaching strategy or procedure adopted;
- ii. The achievement of set objectives stated prior to teaching/learning;
- iii. If a set of learners can proceed to next learning content or if they can be allowed to move to the next class or level of learning;
- iv. If they can be certified to have successfully completed a course of study and adjudged competent (Igbokwe, 1994; Kaplan & Saccuzzo, 1997; Gregory, 2004; Nwanna, 2007).

Assessment has been defined as “the process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs” (Wikipedia, 2013:1). Assessment sometimes focuses on the individual learner or the learning community (which includes a class, workshop or other organized group of learners as in institutions of higher learning).

Assessments in institutions of higher learning in Nigeria have been observed to be fraught with many faults. Assessment here is taken to be the end of semester or session examinations given to students of these institutions to determine grades obtained by them in the courses they had registered for the semester or session. It also includes the semester or session examinations administered by lecturers to obtain scores which can be used as continuous assessment scores. The faults observed include:

- i. Lack of integrity of the test items or questions set by the teachers. This arises from the fact that the items or questions are not well thought out, written in a hurry and do not represent a fair sample of what had been taught/learned or is supposed to have been taught/learned.
- ii. A high percentage of the lecturers outside the Faculty of Education have not taken courses in the methodology of teaching their various courses. This, in effect, means that they also do not understand the proper procedures to follow in the assessment of learning of their students. Thus, apart from the problems associated with the development of test items or questions, they also have problems with scoring and grading of the students scripts. It has been observed that they do not use marking guides (schemes) and therefore, their grading are subjective.

Individuals as well as organizations and institutions are always seeking for new ways of increasing the productivity of their workforce. In this regard, they will willingly, unreservedly, embrace any change that is positive. Thus, with the introduction of automation in assessment the institution was willing to adopt it. Automation in this paper is understood to mean the ways institutions adopt to mechanize their assessment processes and in so doing eliminate wasteful practices as well as increase the release of results. It is also seen as a way of generally improving assessment practices. It is this desire to improve on its practices that led the institution into changing its mode of end of semester assessments and to try to automate the process.

Prior to this, assessment was done using the essay type test except in the practical orientated courses. The problems besetting this system of assessment were weighty enough to force the Senate of the University to seek to adopt some other system of assessment. Igbokwe and Anyaehie (2010) in discussing the system listed the following as the problems identified by the Senate:

- i. Use of limited sampling frame for the setting of questions. This meant that most lecturers' set-questions which did not cover the content of the courses taught/learned.
- ii. Choice of questions to answer. All lecturers except some of those in measurement and evaluation, offered students options of questions to choose from. This meant that students were not being assessed on the same bases of questions answered. Thus, students were doing as many assessments as there were options.
- iii. Easy predictability of examination questions by students. This arises from the frequent repetition of questions and leads given by lecturers either in the course of teaching or during revision.
- iv. Lecturers did not develop marking schemes for the purposes of grading the students' script. This gave room to subjective scoring of the questions. Thus, students were sometimes victims of the impressions of the lecturers about them, or the mood of the lecturer at the time of grading or sheer bias.
- v. There was a lot of stress on the lecturers arising from the grading of scripts for large groups or classes of students. It also meant the reduction of time spent by staff on research and community service due to unending grading of students scripts.
- vi. Results of end of semester assessments were delayed. This resulted in students' inability to register failed courses for the following semester/year. It also meant that graduating students could not be eligible to take part in the mandatory one year National Service at the appropriate time when they should.

Procedure Adopted In the Automation of the Generation and Scoring Of Test Items

The decision of Senate to adopt the multiple-choice mode of assessment was well accepted by the University's academic community. An examinations committee worked for one session and was replaced by an Examinations Centre which had a full complement of staff and was headed by a Director. The Examinations Committee and later the Director of the Examinations Centre was expected among other things to:

- i. Receive all question papers, as well as the keys to these questions, from Deans and Provost at a specific time within the semester.
- ii. Undertake the selection of questions (items) and the production of the question papers.
- iii. Oversee the marking of the answer scripts.

- iv. Undertake the collation of examination and Continuous Assessment (CA) scores.
- v. Liaise with the Vice Chancellor and receive approval to release the results to the College and Faculties.

The enormity of the tasks outlined above associated with the handling of all examinations, which hitherto had been the exclusive preserve of the lecturers through their departments, meant that new ways had to be sought. First, as had been noted, the mode of assessment had to change from the essay to the multiple-choice mode.

From a survey carried out before the decision to change, it had been found out that not many lecturers were familiar with this mode. It was therefore, important to first train them in the development of test items. This was organized by the Examinations Committee in consultation with the Measurement and Evaluation Unit of the Department of Psychological Foundations. The training took place in eight faculties of the University. The zeal of lecturers to understand the system was palpable, so also was the pessimism.

With the completion of the training, it was then time to start the development of a question bank for all the courses adjudged to be amendable to the new mode. To achieve this, each lecturer or team of lecturers was expected to submit in the first instance a total of one hundred and fifty (150) items in each course being handled by him/her or it. The bank continued to be enriched with the submission of the same number of items in the subsequent corresponding semester. This meant that at the end of 2009/2010 session some courses had as many as seven hundred and fifty (750) items to their credit in the bank. Some other courses had as low as two hundred (200) items. The low holding (credit) resulted from the reluctance or inability of those concerned to submit the required number of question (items) for each of the semesters.

Automatic generation of the test items did not imply the abandonment of the human handling of the process. First appropriate computer programmes were acquired. These were programmes that ensured that appropriate formats in typing were followed and that the items were joggled in such a way that a chosen number of variants of question papers were produced. The number of variants was dependant on the number of items available. The more the number of items, the more the number of variants that could be generated. The typing of the items and the handling of the process was done by Computer Operators who were trained and supervised for sometime by the Programmer. This unit of the Examinations Centre also handled the mass production and packaging of the question papers. The questions are mass produced using isograph machines. This ensures the rapid reproduction of a question paper with a space of minutes.

At the end of each examination, the Optical Mark Recognition (OMR) sheets were returned to the centre. They were then sent to the Scanning Unit. Here the OMR sheets were read and transformed to excel format. The scanned outcome is sent to the scoring unit. This is made possible because both units are networked. Scoring is done with the keys to the various variants of the question paper. These keys were not released by the Production (Typing subunit) Unit to

the Scoring Unit until the examinations had been administered or taken and only when the Director or a duly authorized officer in the absence of the Director gives the directive. The computers had been programmed to score each variant and then produce a single result for each course. The CA scores are then manually added to the examination scores to obtain the final score and grade for each student. The examination scores are, however, automatically upgraded to a hundred percent (100%) when there are no CA scores.

Challenges

The effort to automate the examination system of the University faced a number of challenges. These were related to: hardware, software, staffing, cooperation of the academic staff, students' data, production environment and the number of scripts to be handled.

The first constraint to automation was the availability of the necessary hardware required for the process. The University had to find resources to buy hardware. Though it succeeded in buying quite a number of computers, isograph machines, scanners, printers and so on, the present stock is just enough. With further perfection of the system more will still be needed.

The initial problem of finding the adequate software to use was overcome by the development of the appropriate programmes. This enabled the computer operators to have the question automatically randomized immediately after vetting of the typed questions. Further improvement of the programmes led to the production of compact question papers. This reduced the volume of papers used for examinations. The problem of handling the continuous assessment scores still persist. These scores are still being manually entered. The development of appropriate software to handle the presentation of CA scores and adding these to the examination scores will further reduce the time for the release of results as well as ensure a more efficient handling.

The availability of adequately trained computer operators within the University was an initial problem of the change of automation. This has been overcome by the rapid retraining of the University staff and the engagement of temporary staff.

The cooperation of the academic staff can be said to be paramount in the success of the automation of the system. This cooperation was in the area of submitting of questions (items) in order to establish and continually enrich the questions bank (account) of each course. Though a majority of the academic staff had continued to be cooperative, a significant minority was observed to have been uncooperative. The cooperation expected was in the areas of:

- i. Early submission of structured questions (items) papers with the specified number of items.

- ii. Submission of the correct keys to the questions. Failure to submit this may lead to the inability of the centre to produce question papers or where produced to the mass failure of the testees
- iii. Careful vetting of vetting of question papers.
- iv. Proper collation of question papers in order to eliminate the wrong matching of test variants with OMR sheets. (The stage of automatic collation of the question papers has not been reached).
- v. Conscientious submission of all OMR sheets used by students.

Availability of the correct number of students registered for the different courses had remained a problem. These data are needed to produce the question papers and in order to avoid under or over production of question papers. The problem emanated from the inefficiency of the course registration process. Many students held on to their registration forms sometimes until the end of the session. Thus, even lecturers do not often know, especially with large classes, the number of students they are teaching. The problem however is being solved by the enforcement of online payment of fees and of registration of courses as well as the enforcement of the payment of fees for late registration. It is hoped that with this, the number of students involved in any one course will be available before the commencement of semester examinations.

Another problem was securing the production environment. The computer files as well as the results needed to be protected in order to prevent their being accessed by unauthorized persons. This was done through the introduction of codes and passwords for those who work with them. Limiting access to the rooms housing the equipment also help in the control. Security personnel were also posted to the building housing the centre.

One of the greatest challenges was the volume of scripts handled by the centre. The number of scripts to be scored and results reported was above one million (1m) during each of the semesters. Each of this had to be scanned before being scored. This took time and threatened to defeat the aim of early release of results. This problem was solved in part in by increasing the number of scanning machines (scantrons) as well as deploying more personnel to the scanning and scoring units. The staff was also encouraged to work longer hours beyond the normal work hours.

Conclusion

The automation of aspects of test preparation and reporting of results had achieved the hopes of Senate in deciding to change the mode of examination from the essay to the multiple-choice. Namely, to ensure that : results were released to students soon after taking the examination; the subjective assessment of students was eliminated; unscrupulous lecturers stopped the abuse of their position and authority through the victimization and/or harassment of testees; the unwholesome practice of ‘sorting’ of lecturers and those non teaching staff handling results was also eliminated.

The experience of the University in automation has been that the human element cannot be eliminated. People are used in typesetting of the questions as well as their production. People are also used in scoring and the reporting of the test results. It then is obvious that no matter how efficient automation may be, the contribution of people is still needed. This then implies that there is a need for continued human capacity development. This fact bears re-emphasizing.

As had been pointed out part of the challenges to automation in the University was the reluctance of some lecturers to submit questions on time. Attitude is an important issue. Change sometimes is difficult to accept. Resistance to change should, therefore, be expected and should not be allowed to truncate the introduction of good ideas and best practices. Resistance appears to be more and persistent when such changes are drastic and overwhelming. The resistance to the change in examination system was partly because it was inclusive and introduced at one fell swoop. On account of this altitude some lecturers unconsciously worked against the system. Opposition also came from the regulatory body of Universities. Its argument was that the University cannot solely use this mode in the assessment of her students. It was the hope of the University that it would arrive at a point where its examinations would be computer-based. This would have completely eliminated the incidences of leakages either through the course lecturers or the examinations centre as well as cheating during the administration of the tests. These oppositions have forced the University to reduce the number of courses where item generation and scoring are automated. They have also put a check to the progress towards computer-based testing of courses in the university.

Finally experience so far has shown that not all aspects of assessment can be automated. For example assessment in practical courses was seen not to be amenable to automation. There is, therefore, the need for research or perhaps continued research into how these aspects can become amendable.

References

Educational Assessment (2013) Wikipedia, the free encyclopedia. Retrieval 22-03-13
from http://en.wikipedia.org/wiki/Education_assessment.

Gregory, R.J. (2004) *Psychological Testing: History, principles and application (Fourth Edition)*. Boston: Pearson Education Group, Inc.

Igbokwe, U.O. (1992). An introduction to psychological testing. In Mkpa, M.A & C.C Ekennia (Eds), *Education in Nigeria: Issues, trends and practices*. (pp. 317-327). Owerri: Whyte and Whyte Publishers, Inc

Igbokwe U.O. & Anyaehie, E.O. (2010) The University Examinations Centre. In Mkpa M.A. (Ed), *My leadership*. (pp. 77-85). Uturu: ABSU press

Kaplan, R.M. & Soccuzzo, D.P. (1997). *Psychological Testing: Principles, Applications and Issues, Fourth Edition*. Pacific Grove: Brooks/Cole Publication

Nwanna, O.C. (2007). *Textbook on educational measurement and Evaluation*. Owerri: Boma Way Publishers