The Use of ICT for Assessment and Evaluation

Dr. Muhammad Naqeebul Khalil Shaheen
Assistant Professor
Department of Education, University of Wah

Dr. Nazir Haider Shah
Assistant Professor
Department of Education, Mohiud Din Islamic University, AJ&K

Hajira Naqeeb
Federal Government Education Institute (Cantt & Garrison)

Abstract

Information and communication technology (ICT) has become an integral part of the education system round the globe. E-assessment and evaluation have gained the status of the most reliable emerging trends in information and communication technology to get a brief assessment of complex competencies. The major purpose of this study was to review the use of ICT for assessment and evaluation in the educational settings of Pakistan in terms of shifting from paper based assessment to e-assessment, opportunities and challenges for this shift and future inculcation of assessment and evaluation in the educational settings. The study concluded that there shifting from paper based assessment to e-assessment requires four generations i.e. e-assessment, adaptive tests, transforming assessment and automated scoring. The former two are practiced more frequently in the educational setting, while, there is a need to practice the latter two. Furthermore, a number of tools and settings undertaken to assess learners' skills are not widely used in the field of education and still passing the experimental phase. The study recommends policy action for developing curriculum-fit ICT settings, knowledge exchange through teachers' networking and providing guidance and encouragement for the use of ICT for Assessment and Evaluation.

Key Words: *ICT, Assessment, Evaluation, E-Assessment and Evaluation, E-learning.*

1. Introduction

The importance of information and communication technology (ICT) in the field of education has proved its worth round the globe (Kozma & Anderson, 2002; Pelgrum, 2001; Hennessy, Ruthven, & Brindley, 2005; Goodison, 2003; Kangro & Kangro, 2004). The outcome of this realization resulted into an increased number of computers and networking in educational institutes (Pelgrum, 2001). In addition, a number of researchers have claimed that the place it has gained today will remain on an increase in the future (Becker & Ravitz, 2001). In spite of its wide spread use, its introduction and subsequent integration in classroom activities is not an easy task to accomplish (Cooper, 1998), therefore, the predicted number of computers to be used in the field of education is in the progress (Kozma & Anderson, 2002).

The future promising trends varies from an increase of digital technology based settings in the process of teaching and learning to the emergence of virtual settings that ensure adaptable learning situations. These trends also strengthen the idea of integrating educational software in order to provide quick and minute response and make the teachers able to bring the process of learning to the level which suits most the learners (Becker & Ravitz, 2001).

The learning environment supplemented with technology is mostly integrated in higher studies are likely to be introduced at school level. Here it is pertinent to note that these are already in use at some schools as a source to create learning conducive situations which depend on higher level of thinking, problem solving ability and collaboration strategies (Kozma &Anderson, 2002). Both the teachers and the students are given a chance for assessment of performance, understanding of mistakes and learning from these mistakes (Pelgrum, 2001).

Immersive environments and multiplayer games produce space for learning conducive situations which depend on higher level of thinking, problem solving ability and collaboration strategies and consequently develop the aforementioned expertise/skills (Goodison, 2003). These learning settings ensure the reliable situations; enhancement of collaborative environment, guidance and discussion. Particularly computer simulations and virtual laboratories make the learners able to enhance and use the expertise and information gained in a real life like environment and provision of response well in time (Goodison, 2003). Free online resources, for instance, simulations, virtual laboratories and games are in access today and foster competencies in science. In spite of their successful consultation by a number of teachers and students, they are used limitedly (Hennessy, Ruthven, & Brindley, 2005). Games-approach based settings are commonly witnessed outside the curriculum range, and leave no directions regarding the measurement of learners' learning outcomes (Pelgrum, 2001).

Generally, Learning Analytics has gained the status of the most reliable emerging trends in technology to get a brief assessment of complex competencies. It covers the explanation of a broad range of data obtained and collected on the part of the students in electronic settings for assessment of progress, and finish the educational strategies to bring it to the level of individuals in more effective way (Becker & Ravitz, 2001). This article is designed to have a look at the use of ICT for assessment and evaluation in the educational settings of Pakistan.

1.1. Objectives of Study

Major purpose of this study was to review the use of ICT for assessment and evaluation in the educational settings of Pakistan and its possible future perspectives. For this the researchers aimed to focus on:-

1. Shifting from paper based assessment to e-assessment
2. Opportunities and challenges for this shift
3. Future inculcation of assessment and evaluation in the educational settings

2. Methodology

The study was qualitative in nature and the researchers reviewed existing literature in the field of ICT for assessment and evaluation. For this purpose, the researchers reviewed research articles, conference papers, books, periodicals and internet resources.

2.1 Conclusions from the Review of the Studies

2.1.1. Shifting from Paper Based Assessment to E-Assessment

During past few decades, an increased use of technologies to strengthen and form the process of assessment is observed. By the end of 80s, Bunderson, Inouye and Olsen (1989) produced a significant research study predicting the possible progress by using four generations of computerized educational measurement (Martin, 2008).

The first generation involves the administration of conventional tests using computers.

The second generation which is also known as ‘Computerized adaptive testing’ is based on examinees’ responses. It included sequential items based on the responses of the participants. The third generation is also known as ‘Continuous measurement’. In involves using calibrated measures embedded in a curriculum to continuously and unobtrusively estimate dynamic changes in the student’s achievement trajectory and profile as a learner. While, the fourth generation also known as ‘Intelligent measurement’ involves producing intelligent scoring, interpretation of individual profiles, and advice to learners and teachers, by means of knowledge bases and inference procedures (Bunderson, Inouye and Olsen, 1989).

It is interesting to note that the above mentioned predictions are considerable. The initial generations (two) of e-Assessment or Computer-Based-Assessment (CBA), are more clearly pointed to as Computer-Based-Testing (CBT), have become influential and are used in administration of tests especially large in scale. The point to ponder here is how to make a transition from the earlier two generations to the later two. The tremendous change in data-mining and the subsequent trials and progress in the first two generations with intelligent electronic tutor systems, the later two i.e. 'Continuous measurement' and 'Intelligent measurement' could be possible technologically in coming five years (Johnson, Smith, Willis, Levine, & Haywood, 2011). To put in other words, a long time period will be there to bridge the gap between the era based on computerized testing and the era of 'learning' based on emerging technologies. While the earlier two generations of Computer-Based-Assessment stand on the idea of testing along with the inclusion of computers to upgrade the working of testing processes, on the other hand, the later two focus the idea of tests and strongly include a kind of assessment into learning that is holistic and personalized.

Hence, the sublime idea working behind generations 3 and 4 becomes evident. It unfolds the reality that transformation of explicit testing into obsolete one will be done by computers. Close observation, monitoring and guidance will be made possible in electronic settings which they are supposed to use during learning activities, therefore, diagnostic, formative and summative assessment get embedded in the process of learning. Consequently, (generation 4), assessment will be interwoven in the process of teaching and learning. Whereas the learning system has the capacity to make possible the provision of quick and reliable response and suggestion for both the teacher and students regarding strategies set forth for future and mainly addressing the learners' personal learning requirements and priorities, as expressed during his/her past and present learning tasks.

The inclusion of ICT can improve the effectiveness and viability of these assessment strategies.

2.1.1.1 First Generation: e-Assessment

The terminologies such as e-Assessment, computerized testing, Computer-Aided /Assisted Assessment (CAA), Computer-Based Assessment (CBA) and computer-administered testing of assessment are frequently used as the substitute of one another to reflect a process of

electronic assessment where expression of assessment activities and record of responses are carried out by ICT (JISC, 2006). It is observed that the above mentioned terminologies are often used to point a computer-based multiple-choice or test responses obtained in short answers. The concept of Computer-Based Testing (CBT) expresses this assessment paradigm more appropriately when compared with many other integrated paradigms, for instance, e-Portfolio assessment. However, it is pertinent to note that the later period witnessed the integration of a varied format of questions and answers and propagated as "generation re-invention" testing, which points out the fact that administration of CBT can be done in more new directions. Tests of the first and second generation, i.e. conventional tests based on computer and tests based on computer adaptation, respectively, have proved their worth in traditional assessment procedures during the course of time (Martin, 2008). Currently, an expansion and enrichment of tools used for assessment have made to express the underlying spheres more clearly; to integrate more reliable tasks; and last but not least the permission to carry assessment of constructs that is usually not easy to assess and also came into consideration as a fruit of the information age (Pellegrino, 2010). It is acknowledged fact that accurate measurement of all the test paradigms relies on the standard of the items include in it. The procedure to select these items, such as, Item Response Theory or mathematical programming has a pivotal role in the process of assessment (El-Alfy & Abdel-Aal, 2008).

Tests based on computer benefit the user in a number of ways when compared with tests based on papers. The first includes the distribution of paperless test and subsequent collection of data, a better standard to administer the test, examination of level of students' motivation, getting responses to write and speak and scorable using machine, and provision of quality tools for those who attempt it (e.g., calculators and dictionaries). These also provide a chance for questions interactive in type (Bridgeman, 2009).

Moreover, it is witnessed that tests based on computer played a positive impact in motivating, concentrating and performing tasks assigned to the students; make the teacher able to access highly standardized materials; they are further free due to scoring obtained automatically just to ponder the outcomes and their explanations; and the most latest e-Assessment applications are also echoed which made the provision of detailed reports that reflect the positive and negative aspects, consequently favoring assessment formative in nature (Ripley, 2009).

2.1.1.2 Second Generation: Adaptive Tests

After the administration of each item, the skill level of examinee is calculated in adaptive type of testing. Thus the administration of every next item is pre-selected, to some extent on this updated estimate, hence, on the basis of prior performance, the examinees are grouped for answering easy or difficult questions (Bennett, 2010; Bridgeman, 2009). The salient feature of the testing based on computer is its efficient mode of conduction (lesser items and lesser testing time) and precise measurement as compare to the paper and pencil (Martin, 2008).

CAT is the most popular kind of algorithm testing. It is a test in which the algorithm's designing ensures the provision of an exact point of estimate of learners' abilities or gains (Thompson & Weiss, 2009). Another name of the list includes Computerized Classification Testing (CCT), also recognized as sequential testing in which the algorithm designing focuses on the classification of the learners. For instance, they are grouped either as successful/failure or categorized into levels according to the achievements in the field of education such as primary/intermediate/advanced (Thompson & Weiss, 2009). The major drawback in CATs is the variation of items exposure. Another aspect addresses the test security where the inclusion of high frequency for the most discriminating items stands as a risk (Martin, 2008).

For instance, in USA, The Measures of Academic Progress (MAP), is a test series based on adaption of computer for assessing reading comprehension, mathematical skills, language competency, and science at primary and secondary level schools (Bridgeman, 2009; Csapó, et al., 2012). Similar instance can be seen in the Oregon Assessment of Knowledge and Skills (OAKS) that is a test adaptable at grade 3 to 12 in assessment of reading comprehension, mathematical skill, understanding science and social studies (Bennett, 2010; Csapó, et al., 2012). Admission procedure to higher studies requires a number of adoptable tests as pertinent part, for instance, the GRE-CAT, different graduate level programs of masters and doctorate level, the GMAT for management courses of graduation level, and the test of TOEFL designed for foreign learners.

2.1.1.3 Generation Reinvention: Transforming Assessment

Bennet, in 2010, segregated between and among the three generations of Computer-Based Assessment (substituting, innovating and transforming), widely considering the usual phases adoptable in technology (Meijer, 2008). Roughly stated the earlier two were coinciding

with generation 1 and 2 asserted by Bunderson, Inouye and Olsen (1989). On the other hand, the third category – Generation R (Reinvention) – points towards the progress in (1st and 2nd generation) CBT that prepared the ground for the launch of later two generation (3rd and 4th) of assessment based on computer.

Likewise Ripley, in 2009, negotiates the "transformational" paradigm to CBA. He describes test makers journey in order to create educational change. He argued that the developers may reconsider the methods of testing/assessing students that may involve creativity, novelty, working in teams, problem solving skills and better communication through web. In this regard, computers are excellent monitoring devices for the teachers during assessment of the tests e.g. ETS i-skill test. Lent (2008) argued that developing these tests are actual challenges for the teachers who usually rely only on trivial assessment approaches.

2.1.1.4 Fourth Generation: Automated Scoring

Automated scoring programs using variety of algorithms for analysis of automated language have been introduced. The sublime end to these programs is to improve the efficiency of computerized assessment, particularly when it has to handle a large number of students. Automated Language Analysis is a process where individuals' feedback to composition writing (Let say essay type of questions/typing) are assessed and awarded marks using designed software and the words used and phrases given can be awarded marks using process of matching key terms (Csapó, et al., 2012).

Automated scoring is an amazing introduction of the technology in the field of education that claims the reduction in time and cost when it comes to the question about assessment of higher levels of skills, for instance, written expression, but it needs the validation regarding acceptance by those who are going to use it (Weigle, 2010).

The removal of validity threats of automated scoring is done by isolating elements i.e. feature extraction (able to score), feature evaluation (analyzing those elements), and feature accumulation (scoring generation) (Csapó, et al., 2012). For instance, scoring programs designed for essay writing, have the capacity to consider the aspects which are easy to compute i.e. sentence structure, complexity of words, average length of words etc. and then compile them using a method that forecasts more appropriate scores which closely resembles to those awarded by human evaluators in conditional environment, and last but not least removing the aspects that

a machine is unable to grasp (Ben-Simon & Bennett, 2007). Currently, an extension has been witnessed in which automatic evaluation is done by some of the e-assessment systems that examine free-text students' responses to generate automatic learners' modes of concepts (Pérez-Marín & Pascual-Nieto, 2010).

The evaluation and testation of tasks readable by computer, and other formal writings can be done automatically (Amelung, Krieger, & Rösner, 2011). It is found difficult to use Automatic assessment of free-text responses. Though automated scoring is already in use to score length of responses in essay, for instance, by Intellimetric (Vantage), the Intelligent Essay, e-rater (ETS), Project Essay Grande (PEG) (Measurement, Inc.) and Assessor (Pearson Knowledge Technologies) (Bennett, 2010). An Electronic Feedback Software (Microsoft Office) is another introduction of assessment tool in which an email consisting of reports processed by Micro Soft Word is generated and sent to the learners in order to provide more clear feedback and to allow for an easy detection of plagiarism.

2.1.1.4.1 Reliability and Validity of Automatic Scoring

The scores awarded by human evaluator closely resemble the electronic scoring of essays. Thus, it ensures as much close resemblance in results as lies between the human judges, and in some of the cases it has found greater (Bridgeman, 2009; Bennett, 2010). Research studies also reflected that automatic grading exhibits a higher level of reliability when used to assess open-ended questions (Wang, Chang, & Li, 2008). Another fact found is the higher inter-rater reliability when scored using automatic scoring programs which lacks in case of human raters (Wang, et al., 2008). However, these programs primarily focus the length of text and use of mechanics rather than the content and organization of the ideas (Ben-Simon & Bennett, 2007). Therefore, the findings of a research study done in UK on 11-year-olds who produced six hundred essays established the fact that there exists a close relationship between the marks awarded by human raters and those awarded by machine. Albeit, there existed some of the factors abstract in nature like interest relation to the topic and past experiences, whereas little differences were found in the perspective of mechanics, for instance, paragraph demarcation (Hutchison, 2007).

Spoken expressions are also possible for computer to capture and score. Automated scoring for preset responses, for example, answering a simple question in one or two lines, builds

a strong correlation with human raters, whereas the same goal is not achieved yet in case of detailed responses (Bridgeman, 2009).

Automatic scoring has proved its worth over human raters in case of short length responses of a sentence (Butcher & Jordan, 2010). Moreover, programs are designed which not only provide detailed description of wrong and incomplete answers but also provides the examinee a chance to repeat the same task from the available feedback (Jordan & Mitchell, 2009). Hence, in this way the conventional areas of concerns regarding the value of automated scoring of the material for formative assessment are minimized to some extent (Gipps, 2005).

2.2 Opportunities and Challenges

Development and research in educational technology requires to focus the most important techniques for precise assessment based on competence, for instance, Learning Analytics, virtual and cooperative software for education to regulate self-learning (Bridgeman, 2009). Implementing the existing solutions offered by technology should concentrate on increasing the spectrum, usage, variation, and curricula-adjustment on the inclusion of valid and multifaceted assessment tasks including options for individual as well as collective assessment (Bennett, 2010).

Pedagogical strategies which include ICT to assess students' competences should: include individual and group assessment to increase self-regulated learning; choose assessment designs that promote conducting tests and arouse alternate solutions; and create learning situations that permit learners to show themselves across a range of media and communication forums (Hutchison, 2007).

To sum up, a number of tools and settings designed to assess learners' skills are not widely used in the field of education and still passing the experimental phase. Particularly learning analytics and formative assessment, which are seen as the pivotal to assess the learners' abilities, are still in the way to reach the entire destination. As the rapid progress in the field of technology resulted into introduction of more recent formats to assess learners' abilities, so to take over is not evident.

3. Recommendations for the Future Inculcation of Assessment and Evaluation

More policy support and guidance is required for parents, teachers and students to motivate the take-up of accessible tools and programs in schools, colleges and universities. Particularly, the following options may be considered:

Curriculum-fit ICT settings: Policy action is required to support the development, use and accessibility of ICT settings and tools that consider curricular demands and are better suited to use by teachers as part of their routine teaching strategies.

Knowledge exchange through teachers' networking: Many teachers are not conscious of the possibilities that ICT offer to improve assessment of teaching and learning process. Teacher networks can facilitate knowledge exchange and learning and can contribute to upgrading and mainstreaming existing successful practices.

Development and research: Development and research may focus the innovative learning and assessment settings, for instance, educational multiplayer games and simulations, and consider how learning analytics can meaningfully be used to foster formative assessment.

***Providing guidance and encouragement*:** A critical and open discussion between and among educators, researchers and policy makers is suggested on the advantages and drawbacks of ICT-enhanced assessment strategies, in order to identify viable strategies that allow the precise assessment of all Key Competences for Lifelong Learning.

References

Amelung, M., Krieger, K., & Rösner, D. (2011). E-assessment as a service. *IEEE Transactions on Learning Technologies, 4*(2), 162-174.

Becker, J. H. &Ravitz, J. L. (2001). *Computer use by teachers: Are Cuban's predictions correct?* Paper presented at the 2001 annual meeting of the American Educational Research Association, Seattle, Washington.

Bennett, R. E. (2010). Technology for Large-Scale Assessment. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 48-55). Oxford: Elsevier.

Ben-Simon, A., & Bennett, R. E. (2007). Toward a more substantively meaningful automated essay scoring. *Journal of of Technology, Learning and Assessment, 6*(1).

Bridgeman, B. (2009). Experiences from Large-Scale Computer-Based Testing in the USA. In F. Scheuermann& J. Björnsson (Eds.), *The Transition to Computer-Based Assessment*. Luxembourg: Office for Official Publications of the European Communities.

Bunderson, V. C., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (Third ed., pp. 367-407). New York: Macmillan.

Butcher, P. G., & Jordan, S. E. (2010). A comparison of human and computer marking of short free-text student responses. *Computers and Education, 55*(2), 489-499.

Cooper, J. R. (1998). A multidimensional approach to the adoption of innovation. *Management Decision, 36*(8), 493–502.

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012). Technological Issues for Computer-Based Assessment. In P. Griffin, B. McGaw& E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 143-230). Dordrecht, Heidelberg, London, New York: Springer.

El-Alfy, E. S. M., & Abdel-Aal, R. E. (2008). Construction and analysis of educational tests using abductive machine learning. *Computers and Education, 51*(1), 1-16.

Gipps, C. V. (2005). What is the role for ICT-based assessment in universities? *Studies in Higher Education, 30*(2), 171-180.

Goodison, T. (2003). Integrating ICT in the classroom: a case study of two contrasting lessons. *British Journal of Educational Technology, 34*(5), 549–566.

Hennessy, S., Ruthven, K. & Brindley, S. (2005). Teacher perspectives on integrating ICT into subject teaching: commitment, constraints, caution, and change. *Journal of Curriculum Studies, 37*(2), 155–192.

Hutchison, D. (2007). An evaluation of computerised essay marking for national curriculum assessment in the UK for 11-year-olds. *British Journal of Educational Technology, 38*(6), 977-989.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.

Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology, 40*(2), 371-385.

Kangro, A. &Kangro, I. (2004). Integration of ICT in Teacher Education and Different School Subjects in Latvia. *Educational Media International, 41*(1), 31–37.

Kozma, R. & Anderson R. E. (2002). Qualitative case studies of innovative pedagogical practices using ICT. *Journal of Computer Assisted Learning, 18,* 387–394.

Lent, G. V. (2008). Important Considerations in e-Assessment: An Educational Measurement Perspective on Identifying Items for a European Research Agenda. In F. Scheuermann& A. G. Pereira (Eds.), *Towards a Research Agenda on Computer-Based Assessment. Challenges and needs for European Educational Measurement*. Luxembourg: Office for Official Publications of the European Communities.

Martin, R. (2008). New possibilities and challenges for assessment through the use of technology. In F. Scheuermann& A. G. Pereira (Eds.), *Towards a Research Agenda on*

Computer-Based Assessment. Luxembourg: Office for Official Publications of the European Communities.

Meijer, R. (2008). Stimulating innovative item use in assessment. In F. Scheuermann& A. G. Pereira (Eds.), *Towards a Research Agenda on Computer-Based Assessment. Challenges and needs for European Educational Measurement*. Luxembourg: Office for Official Publications of the European Communities.

Olsen, J. B., Bunderson, V. C., & Inouye, D. K. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (Third ed., pp. 367-407). New York: Macmillan.

Pelgrum, W. (2001). Obstacles to the integration of ICT in education: Results from a worldwide educational assessment. *Computers & Education, 37,* 163–178.

Pellegrino, J. W. (2010). Technology and Learning - Assessment. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 42-47). Oxford: Elsevier.

Pérez-Marín, D., & Pascual-Nieto, I. (2010). Showing automatically generated students' conceptual models to students and teachers. *International Journal of Artificial Intelligence in Education, 20*(1), 47-72.

Ripley, M. (2009). Transformational Computer-based Testing. In F. Scheuermann& J. Björnsson (Eds.), *The Transition to Computer-Based Assessment*. Luxembourg: Office for Official Publications of the European Communities.

Thompson, N. A., & Weiss, D. J. (2009). Computerized and Adaptive Testing in Educational Assessment In F. Scheuermann& J. Björnsson (Eds.), *The Transition to Computer- Based Assessment*. Luxembourg: Office for Official Publications of the European Communities.

Wang, H. C., Chang, C. Y., & Li, T. Y. (2008). Assessing creative problem-solving with automated text grading. *Computers and Education, 51*(4), 1450-1466.

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*(3), 335-353.