What needs to be done for Successful E-Assessment Implementations?

Elif Tanrikulu

University of Bremen, Germany
elifcik@informatik.uni-bremen.de

Abstract. Assessment is one of the most important parts of the education system. The effectiveness of teaching and learning needs to be assessed so that all parties involved in this system can be improved. When "Assessment" is powered with technology and named as "E-Assessment", generating the data to make the necessary improvements in education gets easier. However, as is the case in other fields, technology brings risks and challenges along with its indispensable benefits. This paper provides insights about success factors that are extracted from the literature and validated through a Delphi study. With the participation of eleven experts in two iterative rounds, in addition to the success factors, practicable solutions to achieve success are also collected. The decision makers who plan the implementation and administration processes of E-Assessment can consider the proposed key success factors and the practicable solutions when allocating resources.

Keywords. E-Assessment; Delphi method; success factors.

1 Background

The use of technology in educational assessment is quantitatively increasing. In the three decades since the first generation of computer-based tests took place in the United States (US) [1], many countries have delivered their formative and summative examinations in electronic form (e.g. see [2] for the US; [3] for the UK; [4] for the Netherlands and [5] for Australia). One of the major international surveys, the 'Programme for International Student Assessment" (PISA), intends to move completely to electronic mode by 2015 although "there are still a lot of open questions to be discussed" [6, p. 7].

E-Assessment can be defined as digitizing any assessment-related activity [7]. E-Assessment can make the whole assessment process more efficient and provides numerous possibilities, which cannot be achieved with traditional paper-pencil assessments. For example, test scores can be received quickly (sometimes almost immediately) [8], richer assessment data for decision making can be collected [9] and consistency in marking can be ensured [8].

Considering the benefits and challenges of E-Assessment and also some of the open questions in the field, the study presented in this paper addresses the following questions:

- What are the most important organizational and technical success factors of E-Assessment?
- What is the order of perceived importance of these factors according to experts' opinions?
- What are the practical solutions for achieving the proposed success factors?

This empirical study will validate the relevance of identified success factors and explore new and context-specific ones. Thus, decision makers who plan the implementation and administration processes of E-Assessment can utilize the identified and ranked success factors in this study while allocating resources for their projects.

2 Methodology

2.1 Data Acquisition: The Delphi Method

The Delphi method has been applied in a wide variety of fields [10] as a decision making tool to deal with complex problems [11]. It is an approved technique in information systems research [12, 13] and has been used broadly "to identify and rank key issues for management action" [14, p. 763] which is one of the objectives of this study. In spite of the different applications of the method, there are three common features of Delphi procedures identified by Dalkey (1969): anonymity; statistical group response' and iteration with controlled feedback [15].

Participants are kept anonymous from each other so they are not affected by dominant respondents [10]. Because participants are not anonymous to the researcher, the answers can be tracked in case follow-ups are needed [10]. Individual contributions are aggregated and the group's judgment is assessed through statistical procedures in order to ensure the reliability of the results [16]. The iterative process is built by basing each round on responses to the preceding round. Feedback is given to the group at each round with a summary of the results and respondents have a chance to revise their answers by considering the group's views on the subject [10].

Delphi administration: The Delphi process usually has four phases which can be summarized as exploration of the subject, understanding the view of the group, resolving significant disagreements and final evaluation [11].

For the first phase of the study, to explore the subject, an initial set of success factors was identified through a literature review. Journal, conference and workshop papers, technical reports and books were collected using the search term "E-Assessment" and other relevant terms (e.g. "computer-based test", "online assessment", "technology-based assessment") on online databases. The criteria for the document selection were that these documents had some relevance to the implementation of E-Assessment with the focus of organizational and/or technical issues.

In the second phase, views of experts were collected through an online questionnaire. Participants were asked to rate the importance of values of success factors based on a 5-point Likert scale ranging from "Unimportant" to "Very important". Respondents were also asked to add key components for each success factor.

The second iteration was based on the results of the first one and used to resolve disagreements and improve the reliability of the results.

In the final phase, the results of the two rounds of questionnaires were analyzed and documented.

Identification of the experts: The participant selection criteria for this study were that the candidates would either have relevant publications or have been actively involved in a number of relevant projects.

Eleven out of eighty expert candidates initially contacted participated in the study. Delphi studies, differently from traditional surveys, do not depend on representative sample size since they require contributions of qualified experts who have deep knowledge on the subject [10]. For this reason, eleven was considered reasonable and invitations to the second round were only sent to those experts who took part in the first round.

2.2 Data Analysis

"There are many different views on what are the 'proper', 'appropriate', 'best' and/or 'useful' procedures for accomplishing the various specific aspects of Delphi" [11, p. 3]. It has been suggested that consensus can be determined by examining the aggregate of judgment and the convergence of opinion [17]. In this study, the evolution of consensus between two rounds had been observed as a central tendency and convergence of opinions by checking the movements in mean, median, standard deviation (SD) and percentage response rates [16, 18]. The mean, as a measure of central tendency, represents the group opinion. The standard deviation, as a measure of spread, represents the disagreement level [18]. High SD values mean high disagreement. Percentages for each level of the importance scale are used to assess convergence of responses.

Qualitative coding [19, p.81] (using both deductive [19] and inductive [20] approaches) was used to analyze the free-text responses in order to expand the list of success factors and their key components by deriving new elements from the dataset.

3 Initial Set of Success Factors

In this section, the initial set of success factors which formed the base of the first round questionnaire is explained. Short descriptions and some of the practical solutions resulting from the literature review are presented for each success factor. Organizational factors tend to deal with managerial issues such as policy, planning, structure and human resources. However, technical factors tend to deal with specification, acquisition, development, use and protection of hardware and software tools.

3.1 Organizational Success Factors

Human Resources: Specifying roles and responsibilities of personnel is a 'must' for the success of computer-based assessment systems [21]. Cooperation among assessment, learning and technology staff was indicated as an important driver in successful E-Assessment implementation [22]. Training has also been pointed out as one of the most important tasks [8, 22].

Institutional Support: Effectiveness is dependent on institutional support [23]. Funding is inevitably required [24].

Administration Support: Technical support should be given throughout the assessment process [25, 26]. A “Service Center” can help to solve technical problems quickly and prevent further issues, possibly via telephone [27, 28]. Documentation such as administration manuals and quick reference sheets can be disseminated to help out with common technical issues [27].

Risk Management: Managing risks has paramount importance [21]. Pilot tests can be used to identify and analyze the risks [21, 27] and emergency plans can be formulated [22]. Having spare computers and paper-based options is suggested [21, 25].

Continual Improvement: The effectiveness of E-Assessment projects is dependent upon “a long range plan for sustainability” [23, p. 16]. Stakeholders’ and examinees’ needs should be considered to make improvements in the E-Assessment system. Service center call logs can be analyzed to get the recommendations for changes and improvements; also post administration surveys can be conducted to directly ask for suggestions [29].

3.2 Technical Success Factors

Assessment System Features: The features of E-Assessment software tools affect all phases of the assessment process [8, 30]. The assessment software platform should support the design of a variety of question types and be an open source platform which is available for free use [30]. Collected data must not be lost in case of any system failure [2, 31, 32].

Testing Room: “An obvious but nevertheless important need for computer-based tests is the test administration sites” [32, p. 2]. It should be quiet and comfortable [5, 22, 33]. Proper lighting is needed to prevent glare on the computer screens [22, 24].

Connectivity: Adequate bandwidth is one of the crucial factors [22, 28, 29]. A wired internet connection is recommended for reliability [31, 33]. In order to deal with low bandwidth, local caching software is used [28, 33] or the data are downloaded on to PCs [34].

Security: Ensuring security is critical [30, 35]. Assessment data including personal registration information, questions and scores need to be secured throughout the ad-

ministration [4, 27, 31, 35]. Secure authentication [4, 34, 35] and protection against viruses and hacking attempts [26, 34] are of utmost importance.

Interoperability: "Interoperability is about portability" [22, p. 18]. E-Assessment materials must be portable between different platforms and learning management systems [8, 26] so assessment resources can be shared and reused [22].

Accessibility: Accessibility has to be taken care of [9, 27, 35]. Necessary accommodation should be provided including extra time [9], Braille versions [9], alternative hardware [32] and software adjustments [27]. The design of the E-Assessment project must comply with legal requirements [26].

Usability: Ease of use of the E-Assessment system is another crucial factor [32, 34]. An intuitive interface design is needed for non-technical personnel in item authoring [30]. Also, examinees should be able to concentrate on the test items instead of spending time understanding how to navigate or how to indicate a response [32].

4 Findings

4.1 Participants' Profile

The E-Assessment project directory (see http://www.dur.ac.uk/smart.centre1/jiscdirectory/page_06.htm), which was used to recruit some of the candidates, mostly consisted of projects in the United Kingdom (UK). Therefore, not very surprisingly, almost all of the participants (10 out of 11) are academics who work in universities in the UK. Only one participant works in a research centre in Greece and all participants except one practitioner have several publications in the field. The expert panel is very homogeneous and the results of the study largely reflect a UK perspective. In their publications, they summarize their work in relevant projects or effective and innovative practices of E-Assessment in other projects through case studies. Some of the publications are literature reviews which are the proof of participating experts' theoretical knowledge in the field. One of the participants conducted an adapted version of a Delphi study in 2006 as a forecasting method to probe visions about the future of E-Assessment [21]. A few other participants have been involved in some projects of the Joint Information Systems Committee (JISC) (e.g. E-Assessment glossary [36], Case studies on effective practice with E-Assessment [37], Case studies on advanced E-Assessment techniques [38], and A guide to technology-enhanced assessment and feedback [3]).

4.2 Ranking

According to the results of the second round, the five most important success factors are *Security*, *Assessment System Features*, *Accessibility*, *Institutional Support* and *Connectivity*. *Institutional Support* is the only organizational factor among the top 5 most important success factors, and in contrast to the findings of Conole and War-

burton (2011), the experts in this study found technical factors outweigh organizational issues [8].

The rankings of the success factors for both rounds are presented in Table 1. For all of the success factors, the SD values were either the same or decreased in the second round compared to the first. This confirms the evolution of consensus between successive rounds.

There are slight decreases in the mean ratings of almost half of the success factors. Nevertheless, on average, respondents perceived the proposed success factors important for implementing E-Assessment projects. When the respondents were asked to comment on the ranking based on the first questionnaire results, they reflected that the ranking was generally fair.

Based on participants' comments on the definitions of success factors, *Administration Support* was split into two factors as *Administrative Support* and *Technical Support*.

Table 1. Importance of ranking of success factors listed by second round ranking order.

Success Factor	Round One (N=11)				Round Two (N=11)			
	Rank	**Mean**	**Median**	**SD**	**Rank**	**Mean**	**Median**	**SD**
Security	#1	3.73	4	0.65	#1	3.73	4	0.65
Assessment System Features	#3	3.46	3	0.52	#2	3.64	4	0.50
Accessibility	#7	3.18	3	0.98	#3	3.18	3	0.98
Institutional Support	#7	3.18	3	0.98	#4	2.91	3	0.54
Connectivity	#10	2.64	3	1.50	#5	2.82	3	1.08
Technical Support	N.R.				#6	2.73	3	0.65
Continual Improvement	#6	3.18	4	1.25	#7	2.73	3	1.01
Risk Management	#9	2.91	3	1.38	#8	2.46	3	1.13
Administrative Support	N.R.				#9	2.36	2	0.92
Human Resources	#4	3.27	4	1.19	#10	2.27	2	0.90
Interoperability	#11	2.09	2	0.94	#10	2.27	2	0.90
Testing Room	#8	3.09	4	1.22	#11	2.00	2	1.10
Usability	#2	3.46	4	0.82	N.R.			
Administration Support	#5	3.18	4	1.08	N.R.			

Notes: SD: Standard deviation; N.R.: Not rated.

The items are ordered by mean first, then by median and then by SD in the case of ties.

0=Unimportant, 1=Of Little Importance, 2=Moderately Imp., 3=Important, 4=Very Imp.

5 Discussion

In the following section, the details of the results, key components and practical solutions, are presented in Tables 2 to 6, showing the results of the literature review and the Delphi study. The list of items in these tables can be used as a checklist for improving the success of E-Assessment implementations.

5.1 Top Five Success Factors

Security: *Security* was ranked as the most important success factor in both rounds. There is neither convergence nor divergence in the distribution of responses because the percentages for each importance level value remained the same. A low SD (0.65) confirms the high consensus on this factor. One of the respondents pointed out that *Security* is essential for the entire process to be stable and another noted that *Security* "leads to confidence in the system". It was also highlighted that the importance given to *Security* depends on whether the test is summative or formative. For formative low stakes assessments where providing feedback is the main goal, some security issues remain important but they are not as vitally important as in summative tests.

Table 2. Key Components and Practical Solutions for *Security.*

Key Components		Practical Solutions
Secure storage & transmission of data	• Security of questions, answers & data**	• Loading examinations on the server in the last minute • Isolated network of server (questions) • Protocols to remove questions from the PCs • Protected answer file directories on server • Software features to prevent: print, copy, send, download • Fragmentation • Time protected files • Password protection • Proxy server • Encryption**
Secure authentication		• Pre-registration of computers • On-site photo of examinees • Fingerprint • Retinal scan • User name and password • Attestation statement • Valid photo ID
Prevent cheating		• Privacy screens • Large item banks and random questions • Video and audio taping • Isolated network of student machines • Cardboard carrels • Video and audio surveillance equipment • Deleting browse history
Protect against viruses & hacking		• Firewalls • Virtual private networks • Encryption

Note: Items with ** were mentioned by experts and in the literature; items with * were mentioned only by experts; and the rest of the items were mentioned only in the literature.

Assessment System Features: In the first round, respondents were asked to rank *Usability* as a separate success factor from *Assessment System Features* but the answers to open-ended questions suggested a merge between these two factors. In the second round, *Usability* was presented as a key component of *Assessment System Features* and ranked as the second most important success factor. This might be caused by the merging, since *Usability* was also ranked high in the first round. The low SD values of both rounds show that there is high consensus on the importance of ratings for this factor. The importance of usability features for authors was emphasized by the respondents. They noted that scheduling, easy marking and reporting features would make life easier for authors. A wide variety of question types and flexibility were also mentioned as important components of E-Assessment systems by five participants.

Table 3. Key Components and Practical Solutions for *Assessment System Features.*

Key Components		**Practical Solutions**
Usability**	For authors**	• Item templates • Style sheets • Easy marking* • Scheduling features* • Multimedia usage* • Reporting features*
	For examinees**	• Practice tests and items • Help options • Video tutorials • Instruction screens**
	General**	• Software evaluations • Usability heuristics • GUI design standards**
Flexibility**		• Cross-platform compatibility • Various question types**
Adaptability**		• Modular design • Open coding**
Robustness**		• Using minimum hardware resource
	Preventing data loss**	• Recovery from system failures • Back up assessment data

Accessibility: *Accessibility* shows a rather divergent response pattern. The responses range between "Very Important" and "Of Little Importance". A slightly high SD (0.98) is also a sign of low consensus. Nevertheless, it was ranked in the third position due to its high mean value.

Three respondents stated that they expected to see *Accessibility* in a higher order when they were given the results of the first round. They also added that *Accessibility* is often neglected. Four respondents highlighted the importance of legal requirements in the UK because institutions have to adhere to the Special Educational Needs and Disability Act while preparing teaching and learning materials [39].

Table 4. Key Components and Practical Solutions for *Accessibility*.

Key Components	Practical Solutions
Hardware & Software**	• Alternative input devices • Larger monitors • Braille versions** • Adjustments on the interface**
Allowances**	• Extended testing time • Individual rooms* • Amanuensis support*
Legal requirements**	

Institutional Support: Mean values show that *Institutional Support* became less important to the respondents. The percentages of responses for "Very important" and "Important" varied appreciably. Also, a convergence occurred between two rounds. The decrease in SD showed a movement towards consensus.

In the first round, *Institutional Support* was defined as "the top management's support on the implementation". In the second round, when only the collected key components were presented instead of the short definition, one of the respondents commented that "Now I understand that this category includes funding, I think that it is more important than I may have first thought". It is assumed that this was not the case for the other respondents because we would expect an increase in ratings instead of the slight decrease that occurred in the second round.

Table 5. Key Components and Practical Solutions for *Institutional Support*.

Key Components	Practical Solutions
Funding**	• Hardware /infrastructure*
Developing awareness*	• Institutional strategy* • Institutional policy*
Coordination*	• Planning • Organization*
Analysis & monitoring*	

Connectivity: A modest fall in SD values confirms the convergence and thus the evolution of consensus. In the second round, more respondents found *Connectivity* important and this change resulted with a slight increase in mean rating. In the second round, respondents were presented with the key components of *Connectivity* (fast response and robustness). Based on their comments in the second round, we can assume that the poor agreement occurred because some of the respondents did not think about these components in the first round.

Table 6. Key Components and Practical Solutions for *Connectivity*.

Key Components	Practical Solutions
Fast response**	• Copying data onto PCs • Local caching software/server**
Robustness**	• Wired connection • LAN

6 Limitations

This study has involved a small sample size and the participant profile is very homogenous (10 out of 11 are from the UK). Therefore the results should be treated with caution. In some Delphi studies participants are recruited from certain countries and they are grouped in panels (e.g. Hong Kong, Finland and US panels in [13]). The results from each panel are compared and interpretation can be considered accordingly. We can assume that this Delphi study consists of a UK panel and if it is followed by a larger study, the results can be compared with other panels.

This research focused on only organizational and technical success factors but with success factors from other dimensions (e.g. pedagogic/psychometric), a bigger picture could be seen.

7 Conclusions

This study empirically identified and validated the organizational and technical key factors affecting the success of E-Assessment projects. The Delphi method was used to analyze and rank these factors through two rounds of online questionnaires. The data were collected from a panel of experts who had been involved in a number of E-Assessment projects and/or had relevant publications. Eleven experts participated in both rounds. The top five factors among twelve were *Security, Assessment System Features, Accessibility, Institutional Support* and *Connectivity*. The expert panel identified a number of key components and practical solutions which were not recognized in the literature review. These findings can assist decision makers when implementing E-Assessment projects and allocating resources.

Acknowledgments. The author would like to thank all the participants for their generous contribution of views and time.

References

1. Bennett, R.E., Goodman, M., Hessinger, J., et al.: Using multimedia in large-scale computer-based testing programs. Computers in Human Behavior, 15 (3), 283–294, (1999)

2. Bridgeman, B.: Experiences from Large-Scale Computer-Based Testing in the USA. In Scheuermann, F., Björnsson, J. (eds.) Reykjavik, Iceland, 39–44 (2009)
3. JISC: Effective Assessment in a Digital Age: A guide to technology-enhanced assessment and feedback. Joint Information Systems Committee (2010)
4. Martinot, M.J.: Examinations in Dutch secondary education - Experiences with CitoTester as a platform for Computer-based testing. In Scheuermann, F., Pereira, A. (eds.) Ispra, Italy, 49–52, (2008)
5. Fluck, A., Pullen, D., Harper, C.: Case study of a computer based examination system. 25 (4), 509–523, (2009)
6. Scheuermann, F., Björnsson, J. (eds.): The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing. Joint Research Centre, Italy (2009)
7. NC: Transitioning to Online Assessment in North Carolina. The North Carolina State Board of Education, NC (2010)
8. Conole, G., Warburton, B: A review of computer ‐ assisted assessment. ALT-J, 13 (1), 17–31, (2011)
9. Ripley, M.: Transformational Computer-based Testing. In Scheuermann, F., Björnsson, J. (eds.) Reykjavik, Iceland (2009)
10. Okoli, C., Pawlowski, S.D.: The Delphi method as a research tool: An example, design considerations and applications. Information & Management, 42 (1), 15–29, (2004)
11. Linstone, H.A., Turoff, M., Helmer, O.: The Delphi method : Techniques and applications. Addison-Wesley (1975)
12. Brancheau, J.C., Janz, B.D., Wetherbe, J.C.: Key issues in information systems management: 1994–95 SIM Delphi results. MIS Quarterly, 20 (2), 225–242, (1996)
13. Schmidt, R., Lyytinen, K., Keil, M., Cule, P.: Identifying software project risks: An international Delphi study. Journal of Management Information Systems, 17 (4), 5–36, (2001)
14. Schmidt, R.: Managing Delphi Surveys Using Nonparametric Statistical Techniques. Decision Sciences, 28 (3), 763-774, (1997)
15. Dalkey, N.: An Experimental Study of Group Opinion: Delphi Method. Futures, 408–426, (1969)
16. Holey, E.A., Feeley, J.L., Dixon, J., Whittaker, V.J.: An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. BMC Medical Research Methodology, 7 (52), (2007)
17. Dajani, J.S., Sincoff, M.Z., Talley, W.K.: Stability and Agreement Criteria for the Termination of Delphi Studies. Technological Forecasting and Social Change, 13, 83–90, (1979)
18. Greatorex, J., Dexter, T.: An accessible analytical approach for investigating what happens between the rounds of a Delphi study. Journal of Advanced Nursing, 32 (4), 1016–1024, (2000)
19. Lewins, A., Silver, C.: Using Software in Qualitative Research. A Step-by-Step Guide. Sage Publications, London (2007)
20. Kondracki, N.L., Wellman, N.S., Amundson, D.: Content Analysis: Review of Methods and Their Applications in Nutrition Education. Department of Health and Human Services, Washington DC (2002)
21. Zakrzewski, S., Steven, C.: A Model for Computer-based Assessment: The catherine wheel principle. Assessment & Evaluation in Higher Education, 25 (2), 201–215, (2011)
22. Pearson Assessment: Considerations for Next-Generation Assessments: A Roadmap to 2014 (2010)

23. Buzzetto-More, N.A., Alade, A.J.: Best Practices in E-Assessment. Journal of Information Technology Education, 5, 25–269, (2006)
24. Sandene, B., Horkay, N., Bennett, R.E., et al.: Online Assessment in Mathematics and Writing: Reports from the NAEP Technology-Based Assessment Project. U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office (2005)
25. Brown, S., Bull, J., Race, P.: Computer-Assisted Assessment in Higher Education. Kogan Page, London (1999)
26. NQC: E-Assessment guidelines for the VET sector. National Quality Council (2011)
27. ITC: International Guidelines on Computer-Based and Internet Delivered Testing. The Council of the International Test Commission (2005)
28. Schaffhauser, D.: High-Stakes Online Testing: Coming Soon! T.H.E. Journal, 38 (6), 28–40, (2011)
29. Mass.: Final Report of the Spring 2011 MEPA Online Test Administration. Massachusetts Department of Elementary and Secondary Education (2011)
30. Ryssevik, J.: Large-scale Computer-based Testing of Foreign Language Competences across Europe. In Scheuermann, F., Björnsson, J. (eds) Reykjavik, Iceland, 57-62 (2009)
31. Haldane, S.: Delivery Platforms for National and International Computer-based Surveys: History, issues and current status. In Scheuermann, F., Björnsson, J. (eds.) Reykjavik, Iceland, 63–67 (2009)
32. Parshall, C.G., Spray, J.A., Kalohn, J.C., Davey, T.: Practical considerations in computer-based testing. Springer-Verlag, New York, NY (2002)
33. Levin, D., Fletcher, G., Chau, Y.: Technology Requirements for Large-Scale Computer-based and Online Assessment: Current Status and Issues. State Educational Technology Directors Association (2011)
34. Bartram, D.: Available Guidelines and Standards for Psychometric Tests and Test Users. In Scheuermann, F., Pereira, A. (eds.) Ispra, Italy, 37–48 (2008)
35. Van Lent, G.: Risks and Benefits of CBT versus PBT in High-Stakes Testing Introducing key concerns and decision making aspects for educational authorities. In Scheuermann, F., Björnsson, J. (eds.) Reykjavik, Iceland, 83–91 (2009)
36. JISC: e-Assessment glossary (Extended). Technical report. Joint Information Systems Committee e-Administration (2006)
37. JISC: Effective practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. Technical report. Joint Information Systems Committee (2007)
38. Ripley, M.: Review of advanced e-Assessment techniques (RAeAT) final report. Technical report. Joint Information Systems Committee (2009)
39. Sim, G., Holifield, P., Brown, M.: Implementation of computer assisted assessment: lessons from the literature. ALT-J, 12(3), 215–229, (2011)