

INCITS-V36/08-0011

United States National Body DRAFT comments on ISO/IEC JTC1 SC36 N1528 New Work Item (NP) on e-Assessment – The Use of Technology for the Delivery and marking of Fair, Reliable and Valid Assessment

Purpose

This document provides DRAFT USNB comments on ISO/IEC JTC1 SC36 N1528 New Work Item (NP) on e-Assessment – The Use of Technology for the Delivery and marking of Fair, Reliable and Valid Assessment.

Defining e-assessment and e-portfolio

E-assessment is defined as the end-to-end electronic assessment processes where Information Communications Technology is used for the presentation of assessment activity and the recording of examinee responses. This includes the end-to-end assessment process from the perspectives of learners, certification or licensure candidates, testing administration center personnel, test development personnel, and certification and learning sponsoring organizations. E-assessment can be administered in a variety of locations such as fixed or stationary testing centers, portable or flexible testing centers, over intranet and Internets and in corporate and learning center computer-networked training centers and laboratories.

An e-portfolio is defined as an electronically-based portfolio, i.e. an electronic file storage and information management system which is modeled on the working method used for paper portfolios, but which takes advantage of the capabilities of Information Communications Technology, notably allowing learners or examinees to store digital artifacts and streamlining the process of review and moderation for learners, certification and accreditation bodies, trainers, tutors, moderators and verifiers.

Stakes Associated with e-assessments

Sponsor organizations for e-assessment should specify the stakes or consequences that are associated with the particular e-assessment being administered. E-assessment consequences can be high stakes when used for certification, licensure, higher education selection, professional enhancement assessments. Medium stakes e-assessments can be associated with end-of-course assessments and grading/marking. Low stakes e-assessments can be used for learning progress assessments, formative assessments, etc.

Technical Considerations for Assessment Validity, Reliability, Fairness and Security

E-assessments must provide professionally appropriate and documented evidence to allow independent individuals and organizations to determine whether or not the e-assessment provides sufficient documented information regarding the validity, reliability,

fairness and security for the e-assessment. The amount and type of evidence and documentation provided for the validity, reliability and fairness should be congruent with professionally accepted standards and guidelines regarding the technical adequacy, appropriateness and use of the examination. The selection of technical standards and guidelines will necessarily depend on the stakes that are associated with the e-assessment as noted above.

Technical testing standards and guidelines that may be referenced in this evaluation include: *Standards for Educational and Psychological Testing* 1999 (AERA, APA, NCME), *International Guidelines for Computer-Based and Internet-Delivered Testing* 2005 (ITC), *Computer Based Testing Guidelines* 2002 (ATP), International Standard ANSI/ISO/IEC 17024 *Conformity Assessment—General Requirements for Bodies, Operating Certification of Persons*, ANSI-PCAC-GI-502 *Guidance on Psychometric Requirements for Accreditation, Principles for the Validation and Use of Personnel Selection Procedures*, Fourth Edition 2003 (SIOP) and *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices).

Integration of e-assessments with other International Assessment, Certification and Accreditation Standards Efforts

There is a crucial need for integration, communication and coordination of the e-assessments with other International Assessment, Certification Standards Efforts. Relevant International Organizations that are involved in the assessment standards efforts include American National Standards Institute, International Organization for Standardization Technical Committee 230 Psychological Assessment, International Electrotechnical Committee, International Accreditation Forum, Technical Assessment Group, International Test Commission, Association of Test Publishers, National Commission for Certifying Agencies (NCCA) and Buros Institute for Assessment Consultation and Outreach.

Establish Common International Glossary and Vocabulary Lexicon for e-assessments

There is a need to establish a common international vocabulary and glossary lexicon for e-assessment terminology. Examples of such international vocabulary differences include: scoring vs. marking; proctor vs. invigilator, centres vs. test delivery systems, awarding bodies vs. certification or accrediting bodies, practice e-test session vs. practice tutorials, “jump” navigation, e-portfolios, etc.

Specify Security Procedures for Tests, Items, Scores and Report Results

Since e-assessments are registered, scheduled, delivered, scored and reported via electronic means there is a crucial need to maintain professionally appropriate security for all elements and systems involved in the development, registration, scheduling, delivery, scoring, reporting, and storage of test results from the e-assessments. The type of security required depends on the stakes associated with the e-assessment.

However, issues of copyright and intellectual property must always be maintained according to the copyright and intellectual property provisions. An e-assessment should maintain security of the tests, items, item banks, test scores, examinee biographical information, and test score reports throughout the required lifecycle of the assessment.

Technical and Psychometric Documentation and Information on the e-assessment

Sponsors of e-assessments should maintain technical and psychometric information concerning the development, administration, analysis, and score meaning of the e-assessment. Information supporting the technical and psychometric appropriateness of the e-assessment can include content domain definitions, job task or role definition analyses, test blueprint specifications, item/ task development and revision, item/task banking, scoring algorithm verification, e-assessment tryouts and revision, item and test analyses, item response theory calibrations, test form assembly, test form comparability and equating, ongoing test and item/task analyses, and technical report preparation. The stakes that are associated with the e-assessment will help to determine the breadth and depth of the technical and psychometric documentation that are needed for the e-assessment.

Selection of the Formats and Types of Items/Tasks Administered

Sponsors of the e-assessment should carefully consider the most appropriate selection and appropriate mix of display and response formats for the items and tasks administered in the e-assessment. The sponsors should consider the types of thinking skills and content display tasks that are required to successfully measure the examinee's knowledge, skills and abilities required for the e-assessment. Each item/task display format has its strengths and weaknesses and professional judgment is needed to determine the item and task types and mix of types that should appear on each e-assessment.

Determining the performance standard, scoring and scaling and equating that should be used for the e-assessment

Sponsors of e-assessments should determine relevant performance standards, scoring procedures, scaling and equating procedures that are needed for the e-assessment. Procedures and personnel involved in establishing any performance standards should be carefully documented and performed using professionally accepted procedures and analyses. Scoring procedures should be specified and validated to determine that the e-assessment is performing as expected. Scoring procedures should be verified especially when innovative test item formats such as constructed responses, concurrent applications, and simulations are used within the e-assessment. Score scaling methods

should be specified and documented to allow for consistent transformation between raw and scale score measures. Where alternative test forms are used technical information should be provided to allow for independent determination of the comparability of the scores obtained by different individuals from the different test forms. Information on the test equating process used, equated score scales, and technical equations used to ensure that scores from the two or more different forms are technically accurate and appropriate.

Easy to Read, Use and Interpret Test Report Results to Examinees and Sponsors

E-assessments should provide easy to read, use, and interpret test score reports for examinees and sponsors. Formative and summative evaluations are needed to determine that the test score and test decision reports from the e-assessment are person friendly and that the test scores or test decisions are clearly specified and information needed for interpreting the scores are provided to examinees and test sponsors.

Overall Comments

We recommend a study period to ensure that the new work proposal is aligned with ongoing standards activities, has a scope that is clearly defined in a more comprehensive context, and is using and referencing established terminology.

Title

The use of the term “e-assessment” and “marking” are not universally understood. The terms need further defining or other words used to more precisely define the scope of the proposed standard.

Scope

The current scope is unclear, ambiguous and is not comprehensive. The relationship between qualitative and quantitative assessments and how “fair”, “valid” and “reliable” are defined to reach an appropriate decision regarding a “qualification/certificate” needs further explanation. Missing elements that should be included in the scope include:

1. Performance assessment beyond essay,
2. Use of computer generated simulations,
3. Simulated work activities,
4. Observation of the person or team being assessed,
5. Lab performance judgments.

The scope should also address whether the standards will:

1. Specify the security requirements for the development and delivery of these alternative assessment types,
2. Specify constraints on the use of Internet technology as opposed to private networked technology for development and delivery,
3. Specify constraints derived from the assessment domain on the use of some these proposed technology,
4. Specify technology and infrastructure required to verify authenticity of the individual taking the assessment.

We would agree that terminology will be crucial to this project, and expect that sources of existing terminology be identified in the new work proposal.

Purpose and justification

Although the purpose and justification seems reasonable, it does not take into consideration a wide variety of standard and conformity assessment activities within ISO or IAF (International Accreditation Forum) that is or has occurred. A more comprehensive analysis of existing standards efforts is needed for the justification of a new work proposal.

Programme of Work

Is there a justification for a three part standard? The issues listed in the comments on scope suggest that an explanation is needed here of what is expected as the scopes of the three parts.

Relevant documents to be considered

Given the wide variety of standard and conformity assessment activities within ISO or IAF, we expected to see a significant list of reference documents here, but found none.

Co-operation and liaison

Considering the variety of standard and conformity assessment activities that are ongoing, we expected to see more specifics about the liaison.

A.2 Regulatory Context

In the United States, both the federal and state governments are increasingly relying on standards that can be used in conformity assessment processes to demonstrate that the assessments are an accurate measurement of actual practice in various occupations or professions. This is important in the US because assessments have to be “legally defensible” in a court of law.

B Related Work

This area needs a great deal of more research. Activities such as ISO PC 230, ISO TC 232, ISO/IEC 17024, IAF GD 24:2004, and APA, AERA and NCME Standards for Educational and Psychological Testing or the SIOP Principles for Validation and Use of Personnel Selection need to be review for relevance and how this new standard “fits” within the total concept of assessing people. A related organization would be the International Accreditation Forum to look at the conformity assessment aspect of this proposal.

Mature Technology

There are differences of opinions about the maturity of this technology. The tools for test development and test delivery are evolving because of the new alternative types of assessment.

Conformity Assessments

We would certainly agree that this standard would be used in conformity assessment activities such as certification and accreditation.

Summary

Because of the lack of coordination with other activities and the lack of clarity of the scope, a study period to further explore the proper scope, depth and use of this proposed standard should be implemented.

After this study period, the United States will reconsider our current vote.