

Online Testing Research: Informing and Guiding Transitions to Computerized Assessment

A white paper from Pearson Educational Measurement

Walter (Denny) Way, Ph. D.

Vice President, Psychometric and Research Services

April 2006



wp0601

*Using testing and
assessment to
promote learning*

Online Testing Research: Informing and Guiding Transitions to Computerized Assessment

April 2006

Introduction

Pearson Educational Measurement has been a trusted partner in district, state and national assessments for more than 50 years. Our mission is to improve teaching and learning by providing the highest quality educational measurement solutions. We help states, school districts, teachers and parents use testing and assessment to promote learning and academic achievement. To accomplish this mission, we provide a comprehensive and integrated suite of services, [Pearson Educational Measurement Solutions \(PEMSolutions™\)](#), which is used to create, deliver, score, report and analyze summative, formative, English language, and alternative assessments, and provides many other custom online and paper-based testing and reporting services.

A critical component of PEMSolutions is high-quality, scientifically-based research in support of our assessment services. In these times of visible and high-stakes assessments, research plays an important role in supporting state testing programs. PEM's staff includes a growing number of researchers and psychometricians who provide research services in support of our clients. Perhaps the most active area of our research relates to the transitions many of our clients are making to introduce and expand online, computer-based administration as part of their assessment programs.

The purpose of this paper is to highlight online testing research initiatives at PEM that we believe will help inform and guide states as they transition to online testing. This research includes reports that have been published in our PEM research report series as well as papers that were presented at the American Educational Research Association (AERA) and National Council on Measurement in Education (NCME) conferences.

Online Testing Research to Inform Practice

Virtually every state is considering online testing for at least some portion of their K-12 assessment program. The advantages of online testing are numerous. However, online testing also presents psychometric challenges and a number of areas where research is needed. Researchers at PEM have been working actively in several of these areas, including comparability of paper and online assessments, scoring essay responses online, computerized adaptive testing research, research on universal design for computer-based testing, and the design of user-friendly score reports that can be accessed either online or by paper. In the sections below, we highlight a number of these projects and provide links to research papers that can be accessed from PEM's web site (www.pearsonedmeasurement.com).

Online Versus Paper Test Comparability

Researchers have been investigating the comparability of online and paper assessments for many years. The need to establish comparability for K-12 assessment programs exists because the infrastructure for implementing online assessments is uneven across schools and school districts in virtually every state. Because not every school in a state can implement online testing for all their students, the expectation is that at least some schools will have to test by paper for the foreseeable future. This being the case, professional testing standards require that states provide evidence about score comparability when assessments are delivered both online and by paper.

PEM has recently completed a research report that reviews the research addressing the comparability of computer-delivered tests and pencil-and-paper tests. The first part of the paper summarizes the state of online testing technology and the different methods used in the comparability studies. The second part discusses the results from the studies, specifically in K-12 testing. The last part discusses the potential of online assessments. The full research report, [*Recent Trends in Comparability Studies*](#), can be found on the PEM web site.

PEM has also undertaken a number of original studies investigating the comparability of online and paper assessments. This research involves a new methodology that provides scientifically valid statistical comparisons while imposing minimal constraints on the local schools and districts that participate in online testing. The method compares online and paper test performance in terms of the reported scale scores for a test and provides alternate score conversions for the online test that can be used for score reporting if significant differences in performance by testing mode are found. A [*study completed by PEM researchers for the Texas Assessment of Knowledge and Skills*](#) was presented at the 2006 NCME conference.

Evaluating online administrations of high-stakes assessments involves more than merely comparing scores with paper versions of the test. For example, it is also important to compare online and paper performance at the level of individual items, as items with certain presentation formats may be more likely to cause mode-of-administration differences than others. [*A study of item-level mode effects*](#) for items in four different content areas (reading/English language arts, mathematics, social studies, and science) and two grade levels (grade 8 and 11), was presented by PEM researchers at the 2006 NCME conference.

An additional aspect of introducing online testing is quantifying the reactions and attitudes of students and school administrators to testing by computer. Another [*paper presented by PEM researchers at the 2006 NCME conference details survey development for online tests*](#), what can be learned from such surveys, and how that information is an essential component in comprehensive comparability research as a testing program is transitioned to computer.

Scoring of Essay Responses Online

As states transition to online testing, one potential enhancement is the use of computer algorithms that can provide valid and reliable scoring of student writing. PEM's sister company [*Pearson Knowledge Technologies \(PKT\)*](#) provides industry-leading automated text analysis technologies and products, including the Intelligent Essay Assessor (IEA), which automatically assesses and critiques electronically submitted essays, providing assessment and instructional feedback useful in almost every subject area. PEM researchers have worked with colleagues from PKT to conduct research with IEA using data from state assessment programs. Results from [*one recent PEM study provided positive evidence for the use of IEA scores as measures of writing achievement*](#).

A paper presented at the 2006 NCME conference compared online and paper essay responses on the Texas Assessment of Knowledge and Skills English language arts test administered at grade 11. PEM researchers utilized automated scores from IEA as part of this research. They found that IEA (and its underlying Knowledge Analysis Technologies engine) produced reliable and valid scores when it was trained using essays entered online and then applied to a second set of essays entered online. However, when IEA was trained using essays that were written by hand and then applied to essays entered online, agreement rates with human scoring decreased. In addition, it was found that the essays entered online differed from the handwritten essays in terms of various characteristics, such as number of words, number of sentences, and the percentage of words that were mis-spelled or unidentified by the automated scoring engine. [Results from the study also suggested that scorers may have been more stringent when rating the online essays compared with the paper essays.](#)

Computerized Adaptive Testing (CAT)

Several PEM research papers and reports address topics related to computerized adaptive testing. One report, titled [Practical Questions in Introducing Computerized Adaptive Testing for K-12 Assessments](#), describes some of the successful practices that have been used in operational high-stakes CAT programs, as well as the challenges these programs typically face. This PEM research report provides advice for state departments of education in considering the use of CAT as they move to transition testing programs to online delivery in the future.

Another [PEM research report provides an in-depth look at a technical topic in CAT research](#). Specifically, the study compares exposure control procedures when an adaptive test is based on a polytomous IRT model. The 2005 study evaluated four methods and recommended using two of them (the Kingsbury-Zara and the modified-within-.10-logits methods) due to their simplicity and effectiveness.

PEM researchers also presented a [paper at the 2006 NCME conference that examines the design and evaluation of a state assessment for limited English proficient students using CAT](#).

PEM researchers and psychometricians have experience with CAT both in operational and research settings, and we continue to investigate new CAT algorithms as part of our research agenda. However, we also caution our clients that CAT is not a panacea for online testing and that a state should consider carefully a number of related issues before deciding whether or not CAT makes sense for a particular assessment application.

Universal Design for Computer-Based Testing (UD-CBT)

In 2005, PEM initiated a project with the Center for Applied Special Technologies (CAST) to extend the concepts of Universal Design to a computer-based testing (UD-CBT) framework. Applying UD-CBT principles relies, in part, on the flexibility of digital media. Just as transferring a text book to a digital format does not immediately make it more accessible, transferring a traditional paper-and-pencil test to a computer does not necessarily make it more accessible. It is the flexibility of the digital format to incorporate multiple representations (text, video, audio), its ability to transform within and across media, and its ability to incorporate tools such as highlighters and linked dictionaries that makes it more accessible. By combining the potential of digital media and technology in CBT, a more accessible assessment can be created by incorporating multiple options for accessing and responding to assessment items. UD-CBT extends the principles of universal design, provides focus for ongoing accommodations research, and facilitates the creation of a systematic, research-based foundation for designing a new generation of computer-based testing tools. Two papers resulting from this joint project were

presented by PEM and CAST in a symposium organized by PEM at the 2006 NCME conference. The first of these papers is [*Constructing Innovative Computer-Administered Tasks and Items According to Universal Design: Establishing Guidelines for Test Developers*](#). A companion paper titled [*Constructing Innovative Computer-Administered Tasks and Items According to Universal Design: Illustrative Examples with Pilot Data*](#) is also available.

User-Centered Score Reports

A recent PEM research focus has been on understanding how score reports are used by stakeholders in the field, and what aspects of score report information are most valued by users. This research has been extremely valuable in the development of [*Perspective™*](#), PEM's customized reporting service that offers both paper-based and electronic reporting solutions. The research spanned nine months and consisted of face-to-face focus group meetings as well as one-on-one phone interviews and usability testing with more than 250 teachers, principals, district test coordinators, state educators and parents. The results of the research included a variety of needs and preferred uses that were identified by the various stakeholder groups, not only for paper-based reports but also for reporting information that would be accessed through a companion web site. A paper based on this research, titled [*Developing Score Reports for Statewide Assessments that are Valued and Used: Feedback from K-12 Stakeholders*](#), was presented at the 2006 AERA conference.

Research in Progress

PEM researchers and content experts are currently engaged in a number of important projects related to online testing. There are three projects in particular that we believe will be of interest to our clients and the measurement field in general:

- *Techniques to overcome bias in human scoring of typed-constructed responses.* The literature has suggested a tendency for raters to score typed essays more stringently than essays written by hand. PEM essay scoring experts have designed explicit training strategies to reduce or eliminate this source of potential rater bias in scoring typed essays, and are working on a research report that documents these strategies.
- *Development and implementation of innovative item type in science assessment.* PEM content experts and researchers are involved in several projects piloting science assessments that include innovative items and tasks. The content development process, technical aspects of preparing items for administration, and scoring issues involved in developing innovative items are among the topics to be addressed in a research report that is currently being planned
- *Usability Considerations for Online Testing.* Usability is the combination of fitness for purpose, ease of use, and ease of learning that makes a product effective. For an online testing system, usability is assured through intuitive design. Intuitive design prevents the introduction of construct irrelevant variance within online testing by assuring that computer experience is not necessary for success. At PEM, current and alternate online testing interface designs are being evaluated by researchers with expertise in usability research. As a by-product of their efforts, a research report is planned that will outline the role of usability research in an online testing system, particularly in the development of new online item types and non-traditional assessments.

Through these and other research projects, we believe that PEM will continue to offer value to our clients as they transition their testing programs to online testing and begin to think about how to design their assessments to take full advantage of new technologies.

Summary

PEMSolutions provides an integrated suite of assessment services to meet the needs of our clients. PEM realizes the importance of scientifically-based research in supporting our clients as they transition their traditional testing programs online and consider how the computer can change the assessments they offer. As part of PEMSolutions, Test and Measurement Services includes psychometricians, content experts, and researchers who have strong academic backgrounds as well as valuable applied experience. PEM subscribes to a “research practitioner model”. The premise of this model is that experience with operational client work informs research, and that quality research supports operational practice and adds value for clients. The research papers and reports summarized in this document are evidence of this philosophy in action, and illustrate some of the ways through which PEM can be trusted to lead our clients into the new world of online testing.

*Walter (Denny) Way, Ph.D.
Vice President of Psychometric and Research Services
Pearson Educational Measurement*

About PEMSolutions™

PEMSolutions (Pearson Educational Measurement Solutions) is a comprehensive suite of custom assessment services to meet the specific needs of states and large school districts. PEMSolutions provides the ability to create quality assessment instruments through Test & Measurement Services, deliver tests both on paper and online with its Assessment Network, score a full spectrum of test items with the Scoring Network, report test results to educators and parents quickly through Reporting Services, and analyze results through Research Services.

PEMSolutions provides summative, formative, English language, and alternative assessments. Schools benefit from flexibility in type or mode of testing and are able to test later while getting results back earlier. All PEMSolutions programs are enhanced by certified program management, world-class customer service and a highly responsive account team.

About Pearson Educational Measurement

Pearson Educational Measurement is the largest comprehensive provider of educational assessment products, services and solutions. As a pioneer in educational measurement, PEM has been a trusted partner in district, state and national assessments for more than 50 years. PEM helps educators and parents use testing and assessment to promote learning and academic achievement.

PEM's full-service offerings for K-12 and other assessment organizations include PEMSolutions™ (Pearson Educational Measurement Solutions) for custom assessments, both online and on paper; PASeries™ (Progress Assessment Series) for formative assessments; Perspective™ for performance reporting; EDWARD™ for assessment-based education data management and reporting, as well as other essential educational assessment products and services.

Pearson Educational Measurement operates as a business of Pearson Education, the world's largest education company. More information is available at www.pearsonedmeasurement.com.

Copyright © 2006 Pearson Education, Inc. or its affiliate(s). All rights reserved.