

Article Title	Date	Author
Best Practices in Computer-Based Test Development	8.21.08	Nikki Eatchel

Whether an exam is computer based or offered on paper, the challenges relating to its development, including security, psychometric editing or review and legal defensibility, remain the same. In addition to these “age old” issues, computer or Internet based exams, since they are delivered more widely, lend an added layer of risk, especially with regard to exam security.

To address ongoing and new testing challenges organizations are well-served by following a standard set of processes for test item development and psychometric editing. For example, many organizations use multiple item writers to develop exam content. While this is common practice, it may lead to variations in test item style, format or difficulty. A style guide with templates and item development standards and rules can go a long way in improving item consistency, format and variety. In addition, content development training can ensure that writers have the tools to develop credible, defensible items and item templates that can be used to create different variations of the same question, thereby increasing the size of the item bank in a shorter amount of time.

Statistical evaluation of test items in the field is critical to obtain feedback on specific item performance, exposure levels, answer patterns and cognitive levels. This intelligence allows the revision of item development processes and feedback for specific test item developers – helping determine what’s effective and how the items fare in the field. This also enables the organization to make decisions on item retention, modification and assignment. Measurements surrounding usage and exposure are helpful in determining when it is necessary to refresh item banks or author new items entirely.

Any organization developing or administering tests should be conscious of the psychometric editing process that includes the evaluation of item difficulty levels and takes things such as grammar, sensitivity and style into account. Psychometrics also provide for the review of test item form and function, such as parallel options, sufficient information to answer the question and answer length.

Objectivity in item and test development is extremely important, and as such, psychometric editing is best performed by test development professionals, not subject matter experts or item writers. Individuals trained in the complexity of psychometric editing evaluate items in a different, more critical light than subject-matter experts or item writers. That does not mean that subject matter experts are not integral to the process; on the contrary. It is important to have review and approval of the final, edited item by subject-matter experts in the appropriate field.

Items developed for any form of testing must be legally defensible to ensure the organization is protected in the event of a legal challenge. Implementing a standard process for item development and psychometric review, as discussed above, can maximize the defensibility of an organization.

Evaluation of legal defensibility includes a critical review of the exam both from a content and psychometric perspective to ensure that the exam was developed according to the *Standard for Educational and Psychological Testing*. The courts defer to the *Standards*

when evaluating the credibility of the exam in question. Legal defensibility can be accomplished via several methodologies. The most important aspect of the development process is to follow and document standardized methodologies and include appropriate test development personnel in the process. There are many different steps in the test development process and different methodologies that can be used for each step. For example, when determining the cutscore for an exam, processes such as the Modified Angoff or the Bookmark Method can be used to determine the appropriate standard for passing. Each of the methods uses a different technique to determine the bar that a candidate must reach in order to receive a passing status.

Developing large test item banks from which test content is routinely refreshed also mitigates the risk of item overexposure for testing companies. Taking the lead from the large test developers and administrators, organizations administering computer-based tests or Internet based tests will want to consider using expanded item banks and scheduled test item refreshment to ensure that candidates do not see the same items or designs, generally decreasing the likelihood of candidates sharing information or recognizing previously used questions during a re-test situation.

In many high-stakes testing programs, test administrators collect and examine forensic data in order to measure how often testing candidates are exposed to particular test items, the average time candidates spend on items and how candidates responses to items change over time and exposure. This ensures the ongoing adjustment of the item development process and content to ensure credibility, legality and security.

The multitude of factors to consider during test development lend credibility and integrity to the exam itself. Organizations able to thoughtfully consider the design and implementation of their testing programs proactively fare better than organizations that don't do the due diligence and publish exams in a hurry. A proactive approach that accounts for item development and editing resources as well as security and IT parameters serves the organization better over the long-haul, as it increases test validity, candidate fairness and offers a higher level of protection against legal challenges.

Many of the "next generation" item banking solutions perform all of these functions and more – taking the critical components of test development and facilitating their application. Using one, powerful "next generation" item banking tool can transform the entire test development life cycle. Item banking tools can streamline the way items are developed, change the way testing programs are maintained, evolve the quality control process, upgrade item security, gather and integrate multiple aspects of items, improve the quality of item pools and facilitate overall management of testing programs. Creating a sound, valid and effective test has never been easier.

Nikki Shepherd Eatchel is the vice president of test development at Prometric. Ms. Eatchel joined Prometric in 1998 and has more than 12 years of experience in the field of test development. Ms. Eatchel frequently presents at national conferences and also serves as a psychometrician on a number of national committees.