

Article Title	Date	Author
Legal Defensibility of Online Tests	6.18.07	Nikki Shepherd Eatchel

Organizations developing and administering online tests must understand that test item development is about more than simply writing questions. To protect both candidates and the testing organization, online test items must be able to stand up to legal challenges. A legally defensible exam is one that is designed to provide all candidates with an equal opportunity to display their knowledge and does not discriminate against a group based on race, color, national origin, sex, religion or protected disability.

But how can organizations ensure that their online tests are legally defensible? Integrating a number of best practices in item development into the testing program can help ensure that test items accurately and fairly measure knowledge, skills and abilities for all candidates in the pool.

First and foremost, following a standard process for test item development is essential to the legal defensibility of online test items. This process should include standardized training for all item writers as well as a rigorous item review and approval process. The review process includes evaluation of things such as sensitivity, style, correctness, cognitive level, and overall structure of all test items. Part of the standardized review process is conducted by subject-matter experts and part is conducted by psychometricians and test development professionals. Since the focus of psychometric and test development review is objectivity, it is best performed by test development professionals, not solely by subject matter experts or item writers. Individuals trained in the complexity of psychometric evaluation of items in a different, more critical light than subject-matter experts or item writers and are more effective in ensuring objectivity of test items.

Second, organizations can improve the legal defensibility of online tests with ongoing data collection and analysis of test results. Using metrics to evaluate the performance of online test items in the field and to determine test response patterns can ensure that no one group is responding differently to questions across the board.

Analysis of item response data (the way in which candidates answer an item) can identify if terminology or descriptions within an item are inappropriate for certain portions of the candidate population. Surprisingly, the modification of only a few words in an exam item can have a substantial effect on candidate performance. For example, when developing items for a particular licensing exam, it was found that a particular item writer like to use the word “pilfer” instead of “steal.” As the word “pilfer” did not have any significance to the licensing area being tested, and the word “steal” is a more common word known across various ethnic and socioeconomic groups, changing that one word created a more valid item for the entire testing population. Post-administration analysis confirmed that the change created more equitable item performance.

Lastly, a written record of the test program’s ongoing development and evaluation can serve as protection in the case of legal challenge. A test development process is only valid if it is documented. Documentation should include the job analysis from which item writing assignments are based and the qualifications of subject-matter experts and test development staff involved in the item development and review cycles. Item and exam

level data should also be stored after evaluation along with standard setting information and associated scoring methodologies.

When conducting a review of exams that have been challenged in a court of law, it quickly becomes obvious that it is not always the “best” exams that are successful against a legal challenge. The exams that withstand legal challenge are typically those exams that followed standardized, industry-accepted test development practices and documented their steps along the way. Without process and documentation, even exams that perform well can fail to impress the courts.

Organizations that develop and administer online tests must always design their exam content and scoring with potentially adverse legal scenarios in mind. The addition of these best practices into an existing testing process will go a long way towards lessening the probability of legal challenges and addressing litigation if necessary:

- Standardization of test item development
- Rigorous item review and approval processes
- Ongoing data collection and analysis of exam and item results
- Documentation of test development processes and ongoing evaluation

The knowledge that organizations are attempting to measure is objective, and by implementing the above best practices, organizations can help ensure that their exams are objective as well.