

Item Development and Review

Question: *What is the best procedure for supplementing our item banks with new items? Can we use other item types besides multiple choice?*

Answer: To maintain a healthy item bank, new items should be constantly added on a periodic basis. The best way to top do this is by conducting an item writing workshop. The workshop should be directed to subject matter experts (SMEs) that are interested in learning how to write well-constructed items. The workshop should include an overview of how the test specifications were developed, and training on the “do’ and don’ts” of writing items. Each SME should be assigned a content area in which he or she is knowledgeable. All SME’s should be reminded that every item must be directly linked to the test specifications. After a few items are written it is useful to have them reviewed and critiqued by the entire group, using the newly acquired skills obtained from the training. The remainder of the session can be alternated between item writing and review.

Once new items are written, they should be reviewed by additional SMEs. Typically, these SME’s serve on an Examination Committee (EC) and are responsible for ensuring each item is properly linked to the test specifications, is written clearly and succinctly, has only one correct or best answer, and has plausible distractors. Additionally, each item should be reviewed for possible bias. It is best not to include race, age, or gender data in an item unless it is necessary to determine the best answer.

At this point, all items approved by the EC are ready for inclusion on an examination. Depending on the speed of scoring desired, newly written items can be included as pre-test items or potentially scored items, pending further review. If the examination program is mature, and the item bank is healthy, and instant reporting of candidates’ pass-fail status is desired, new items should be included as un-scored pre-test items. As a pre-test item, it has no point values (does not influence a candidate’s score), but is included on an examination until a sufficient number of candidates have responded to it so that stable item statistics may be generated. A sufficient number may be 100-150, or 500-1000 candidates, depending on the measurement model that is employed.

Sometimes it is necessary to include new un-tested items as scored items on an examination. This may be for a newly developed exam program, or one that is still trying to develop a healthy item pool. In this instance, candidates’ receive no score reports for several weeks after the exam is administered. This allows for a preliminary scoring review to verify each new item’s performance. In this model, some items may be, by necessity, multi-keyed (have more than one correct answer), or be given “all credit”, if fatally flawed. Doing this preliminary review prior to final scoring assures fairness to all candidates prior to determining their pass-fail status. As the exam program matures, it can be transitioned into immediate score reporting, if desired.

Most licensing examinations use multiple-choice items. However other item types may be used. It is important to remember that alternative item types are not as well-behaved, psychometrically speaking. For example, case studies may be used where several multiple-choice items are asked about the same situation. These item types can be very job-related, but care should be taken to such that each item in the group stands alone. That is, the answer to one item cannot be linked to correctly answering other items. Other item formats (e.g., short answer essay, practical exams, oral exams), require even greater skill in their development and scoring to ensure reliability and validity. As a general guideline, their use should only be included in a licensing exam if the Board, or its testing agency, has the requisite expertise and resources to construct, administer, and score them in a reliable manner.