

Concept Paper

On

An Item Bank Approach to Testing

Prepared by

Dr. Paul Squires



20 Community Place
Morristown, NJ 07960
973-631-1607 fax 973-631-8020
www.appliedskills.com

An Item Bank Approach to Testing

An item bank¹ is an important new approach to testing. Some use the term item bank to refer to any collection of test items. But, properly designed, an item bank is much more. An item bank is a catalogue of test items that are classified according to the content they measure and their difficulty to answer correctly (e.g., easy, medium and hard). Item banks are closely associated with computer administered testing and a set of rules for selecting items. The rules include requirements about the item's content and difficulty level. Using a computer and an item bank, each test taker is administered a set of items randomly selected from the item bank while at the same time the items meet the requirements of the rules. A computer can draw items from an item bank based upon the set of rules. In this way, each test taker receives a different set of items and each set of items covers the same topics and is of equivalent difficulty.

Items in an item bank are organized in a database according to topic and difficulty. For example, a computer technician test might include four topics – operating systems, PC hardware, application software, and network protocols. For each of these topics, items are written at three levels of difficulty – easy, medium and hard. The combination of topic and difficulty level forms a group. Each group has two, three or more times as many items than are needed for any one-test administration. An item bank may contain, for example, six topics and 3 levels of difficulty for a total of 18 groups. If each group contains 20 items in each group, the item bank would contain a total of 360 items (6 X 3 X 20). A single test administration includes a small set items chosen randomly from the item bank's groups but subject to a rule, for example, three items from each group. In this manner each test administration presents to each test taker a different 54-item test of equivalent content and difficulty.

Competency models

The item bank topics derive from a properly conducted job analysis or competency model. A properly developed competency model identifies and provides evidence for the skills required for successful job performance. In the development of a competency model, subject matter experts link skills to the important work functions and job tasks of

¹ Test developers use the word "item" to mean a test question.

An Item Bank Approach to Testing

the jobs studied. Good choices for topics in an item bank are the skills contained in the competency model. Competency models form the basis for the validity of the skills that serve as topics in an item bank. These topics also serve as a blueprint for subject matter experts who then write items for each of these topics. If the test developer follows this approach, test takers can feel confident that the item bank and the set of items randomly selected for each person's test are related to the important skills needed to perform critical job tasks.

Specificity of Topics

One challenge faced when forming topics is to determine the specificity of the topics. For example, suppose a competency model for system administrator jobs lists skill in using operating systems, such as Unix, Mac OS, Linux, and Windows XP, among the important skill requirements. When developing topics for a system administrator item bank, the developer is faced with the question of making Operating Systems a topic or making each operating system a separate topic or making each function an operating system performs a separate topic. The level of detail or specificity that is selected to be the topics is an important consideration in developing an item bank. What level of specificity is best? The answer is that it depends upon several considerations - the depth of operating system knowledge required, the purpose of the test, the type and extent of feedback that will be provided to test takers, and the expected length of the test. More topics mean that the item bank is larger and the test is longer. But more topics provide the opportunity to give the test taker more detailed feedback. If a test is used as an employment test and a single overall score to make a hire/no-hire decision is needed, then fewer topics are required. But, more topics means more items must be written. If a test is to be used to diagnose training needs or certify mastery of all critical job tasks, then more topics are required. As a rule of thumb, if feedback is intended for each topic, the topic score should be based upon at least 6 – 10 test items. This means that for an item bank with 6 topics, each test should contain 36 – 60 items. In this example, if there are three levels of difficulty there would be 18 groups (6 topics X 3 levels of difficulty = 18 groups). The item bank for this testing should contain about 100 – 180 items.

Quality

The quality of the item bank is determined, not only by the topics chosen and their linkage to the competency model, but also by, (1) how well the items meet the standards of good item writing and (2) how well the items represent the topics and levels of mastery of the topic.

Items that are well written minimize the effect of guessing, avoid ambiguity about the meaning of the item, are easy to read, avoid tricking the test taker, and measure what the item writer intended to measure and nothing else. The correct answer to a multiple-choice question with four answer choices can be guessed correctly 25% of the time (1 out of 4). Guessing undermines the quality of the test. Anything that increases a test taker's ability to guess the answer correctly should be avoided. Many of the common flaws in test items increase the test taker's ability to guess the answer correctly. On the other hand, ambiguously written items prevent anyone from answering the item correctly – even the most skilled test takers - other than by chance. Ambiguity undermines the quality of the test. Finally, good items measure only what they were intended to measure. Items should not require a high level of reading ability, knowledge of special terminology or acronyms, or sorting through a lot of detailed information when these are not the purpose of the test items. When flaws in the items occur, the test taker may answer the item incorrectly for the wrong reason.

Items should be written for each topic in a manner that comprehensively and accurately measures all aspects of the topic. The items written for a topic should be, in a sense, a representative sample of all possible items that could be written for a topic. The competency model provides direction to the item writers about the aspects of a topic. When the topics are derived from the competency model, the competency model will provide the information needed to write a comprehensive and accurate sample of items for a topic. The goal of writing a set of items for a topic assumes that a test taker's score on one random set of items is pretty much the same as the test taker's score on another random set of items. As it turns out, experience indicates that this is a reasonable

An Item Bank Approach to Testing

assumption as long as the items written by the item writers are proper, accurate and comprehensive for that topic.

Difficulty level

An item bank is comprised of groups of items, the groups having been formed by topics and difficulty levels. In the example above, three difficulty levels were used. However the more difficulty levels the better. More difficulty levels increase the certainty that any two tests will be equally difficult. The equivalence of difficulty for two tests is more important for high stakes tests than low stakes tests. High stakes tests are those tests that are used for hiring, promotion, opportunities for training and other important outcomes. For high stakes tests, five or more difficulty levels are recommended and special item analysis techniques should be applied to ensure fairness and accuracy.

Another aspect of writing items of different difficulty levels is to write them in a manner that represents how a person moves from being a novice to being a master for the skill represented by the topic. An item writer who is truly an expert in the topic being measured should be able to describe the initial skills and knowledge, common mistakes and points of confusion typically experienced by a novice and, for a master performer, the level of skill and knowledge, special insights, and heuristics possessed. The test items should be written in a manner that represents all points along this continuum. This strategy is particularly valuable in aligning tests and test feedback with training programs. Item bank topics, difficulty levels, training topics and learning objectives should align closely. This approach ensures that the diagnostic feedback accurately directs the learner to training content requiring more attention.

The item bank approach supports computer adaptive testing. An advantage of computer adaptive testing is shorter tests - test takers answer fewer items without a loss of fairness or accuracy. Here's how it works. A test taker is first given a test item of average difficulty. If the test taker answers the item correctly, the computer selects a more difficult item. If the test taker answers the more difficult item correctly, the computer picks an even more difficult item. This process continues until the test taker begins

An Item Bank Approach to Testing

answering items incorrectly. At this point the computer has discovered the test taker's approximate skill level. The computer then administers more items in the difficulty range that the test taker first began to fail. The same process is followed when the test taker fails to answer the first item correctly. The computer administers easier items until the test taker begins to answer them correctly. Additional items are administered in the difficulty range that the test taker began to answer correctly. This process is followed for each test topic. In this way the test taker is administered an accurate and fair test using the fewest number of items possible.

Experience indicates that this approach works well provided the item bank is well designed and maintained. If a topic contains too many difficult or too many easy items, some test takers will not answer any item correctly or some will answer all of them correctly. These conditions undermine the ability of computer adaptive testing to measure the test taker accurately.

Item Bank Maintenance

Items in an item bank must be reviewed periodically by subject matter experts to ensure their on-going relevance. This is necessary because, over time, test items become obsolete. Other test items, after being evaluated statistically, turn out to work poorly, despite the test developer's best efforts to write only the best test items. Obsolete items and items that don't work statistically, must be replaced. In traditional testing approaches with a static set of items, the test must be replaced and an additional validation study must be performed. However, with an item bank approach to testing, updates and revisions can be done easily without disrupting the testing process. New, "trial" items can be written and added to the item bank. Typically, these new items are not initially included in the scoring. After the new items are administered for a period of time, statistical analyses can be performed to ensure they work properly. If the new items "measure up" under statistical scrutiny, they are included in the scoring. If the new items are not working properly, they can be corrected or eliminated. With item banks, trying out new test items and collecting data to determine their quality and difficulty level can be accomplished unobtrusively. In this way the item bank can be replenished and

An Item Bank Approach to Testing

updated as necessary. The on-going evaluation and updating of an item bank is a critical and necessary process to ensure the quality of the item bank and integrity of the testing program. This work is typically done by experts with special knowledge of psychometrics. Psychometrics is the field of study that deals with the specialized concepts and mathematical methods used to establish the reliability and validity of a test.

Summary

The use of item banks to support testing programs has many advantages over traditional approaches. Item banks are a very flexible approach to testing because items and topics can be added or eliminated with relative ease and without the need to conduct a complete validation study (as is necessary when publishing a new form of a test). Item banks support waiver tests, practice tests, employment tests, certification tests, and mastery tests all from a single item bank. Item banks support multiple and frequent retests. This is particularly valuable when testing is associated with training and waiver tests, practice tests and mastery tests. And, an item bank approach is a more secure form of testing because for each test taker and on each test taking occasion a different set of items is presented.

Item banks are an important new development in the design and administration of testing programs. The advancement of computers and the Internet make item banks an excellent option. Item banks involve the development of test items and their assignment to groups based upon the topic the item measures and their difficulty. The topics are derived from a competency model and the difficulty level is associated with the skill levels and progression that a person experiences as a person moves from being a novice to a master. A computer algorithm selects a set of test items from the item bank for each test taker. The algorithm ensures that the same topics and difficulty levels are included in each test. In this way, each person receives a different but equivalent test. Item banks have been successfully applied to diagnostic tests, waiver tests, practice tests, high stakes tests, mastery and certification tests. Item banks are easier to maintain and update compared to traditional forms of testing.

Written: February 19, 2003

Revised: April 15, 2003

Concept paper prepared by Applied Skills & Knowledge, LLC