1995

8. Item Banking

Betty A. Bergstrom
Computer Adaptive Technologies, Inc.

Richard C. Gerson
Computer Adaptive Technologies, Inc.

Follow this and additional works at: http://digitalcommons.unl.edu/buroslicensure

Part of the Adult and Continuing Education and Teaching Commons, Educational Assessment, Evaluation, and Research Commons, and the Other Education Commons

http://digitalcommons.unl.edu/buroslicensure/13

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Licensure Testing: Purposes, Procedures, and Practices by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
ITEM BANKING

Betty A. Bergstrom
Richard C. Gershon

Computer Adaptive Technologies, Inc.

Item banks developed by licensure agencies range from a collection of items stored on index cards to highly sophisticated electronic databases. Regardless of the storage mechanism, most banks contain items that have been organized and referenced according to procedures established by the licensure agency. This chapter outlines useful practices for building and maintaining a computerized item bank. We address storage of item text, graphics, and statistical history. We deal with the creation of paper-and-pencil and computerized tests from an item bank and the use of Item Response Theory (IRT) to calibrate and equate item banks. New directions in item banking are also discussed.

Apparently coined in England during the mid-1960s, the term “item bank” was used to describe a group of test items that were “organized, classified and catalogued like books in a library” (Choppin, 1985). Subsequently, Bruce Choppin and others interested in item banking based on Item Response Theory (Hathaway, Houser, & Kingsbury, 1985) attempted to distinguish between “item banks” (a collection of items calibrated with an IRT measurement system and equated to a common scale) and “item pools” (collections of items grouped by content but not calibrated). This distinction has not been widely embraced and often today the terms “item bank” and “item pool” are used interchangeably.

Computerized item banking employs a computer software program to store collections of test items and their associated classifications and statistics. Computerization allows easy storage and retrieval of hundreds (for some organizations the number may be thousands or even tens of thousands) of items. A well-organized, well-maintained computerized item bank can facilitate and enhance the
construction of both paper-and-pencil and computerized tests. Items can be sorted and filtered to enable easy review by content experts and psychometric staff.

The basic plan for item bank construction includes writing content valid, grammatically correct items (see Chapters 5 & 6), categorizing items according to the content outline or "blueprint" that the testing agency utilizes, and entering the items into the computerized bank.

Once a valid item bank is created, the orchestrated efforts of content experts and psychometricians are required to maintain it. Content experts must review the item bank on a systematic schedule to ensure that (a) items are current and relevant to the field of practice; (b) duplicate and similar items are identified and flagged; and (c) content within the bank is representative of the test blueprint. Psychometricians must also review the bank to guarantee that (a) the range of item difficulty is appropriate; (b) misfitting items have been identified and flagged for rewrite; and (c) the pass/fail standard is current.

Licensure and certification agencies test a large range of candidates. Some agencies test less than 50 candidates per year whereas others test hundreds of thousands of candidates. Still, even agencies that test relatively few candidates usually have item banks of at least several hundred items. Although switching to computerized item banking involves the initial cost of developing or purchasing software and possible conversion costs for existing items, graphics, and statistics into the computerized bank, cost savings are realized in the long run by reducing professional and clerical time for item maintenance and test production. Another important benefit of computerization is reduction of error—the more data are manually manipulated, the greater the chance for mistakes. Thus, even very small testing agencies will benefit by computerizing their item banks.

The following sections of this chapter outline various computerized item banking components. Licensure agencies need to review their item banking needs (both current and future) to decide which components of computerized banking are applicable for them.

ITEM STORAGE

From original draft through ultimate "retirement," an item should be maintained in the computerized bank. The life cycle of an item typically includes development, review by content and bias panels, field test, rewrite, test administration, analysis, review/rewrite, and additional test administrations. Some of these steps may be repeated more than once. A computerized item bank should provide a means of storing, retrieving, and maintaining test items and related descriptive information (Schroeder, 1993). The descriptive data that licensure organizations store varies. The types of information that may be stored in an item bank are as follows:

Item Identifier

Each item must be assigned a unique identifier (ID) which may be a number, a character, or a combination of the two. Whenever changes are made to an item, a new item ID should be assigned. Many organizations add extensions to an
existing number to indicate that the item has been revised (i.e., 1004 becomes 1004a). If item statistics are being maintained over time, it is essential that the item ID be updated with each change. Once a new ID has been assigned, the old item and the new item can be compared to assess the impact of the change on item performance. Remember that changes as innocent as altering the orientation or order of the choices, or simply changing the font, have been shown to affect item difficulty (Gershon & Bergstrom, 1993).

Item Type

Items types include multiple-choice questions (MCQ), short answer, matching, essay, etc. Some licensure organizations develop tests that include a specified percentage of item types on each test. Including this field allows the test developer to sort by item type for test construction.

Classification Schemes

The item bank should store all relevant classifications for each item. Many licensure organizations store multiple content classifications with an item; for example, in some medical areas an item is classified by content (anatomy, physiology, etc.) and by type of patient (pediatric or adult). The National Council of State Boards of Nursing classifies items for the NCLEX-RN examination along two dimensions of content codes, nursing processes and client needs (Haynie & Way, 1994). Another common scheme is Bloom’s Taxonomy, in which the item is categorized by the cognitive processes required to answer the item (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). Additional types of classifications may include task, process performed, or instrumentation required. If items are classified by a nested content outline, the bank should be capable of storing the nested structure.

Licensure organizations develop content outlines and test blueprints according to job analyses and input by experts from the field. A test blueprint defines the scope of practice and the content areas essential for demonstrating competence. Adherence to the blueprint is crucial to confirming test validity. Care must therefore be taken to ensure that items in the bank are classified correctly. Storing items with classification data in a computerized bank allows an agency to easily sort the entire bank or a test to determine if the percentage of items by classification meets the blueprint specifications.

Computerization, however, does not assure test validity. Whether items are stored in a computerized bank or on index cards, validity requirements remain the same. Items must be reviewed by members of the profession to ensure that they are current and relevant to the field of practice and tests must be reviewed to confirm that they meet blueprint specifications.

References

The item bank should provide for storage of references. This information allows the licensure agency to cite a specific reference if the validity or accuracy of an item is questioned by a candidate or by an item review committee.
Author

Storing the name of the author of an item allows the licensure agency to contact the original source if the item needs additional references, clarification, or rewriting. A convenient item bank feature enables the production of a report to each item writer on the performance of their items after piloting has been completed. An even more helpful feature produces the text of all items that must be rewritten by a given author after failure on a test pilot or following review by a content review committee.

Item Status

All items should be coded with a status to indicate the current use of the item. For example, an untested item may be coded as “new,” whereas an item that appears on a test but is not counted toward the candidate’s score may be coded as “experimental.” Agencies may wish to code items as “secure” or “non-secure” to indicate whether the item can be used on a practice test or as an example item. An item that has been used for testing should probably never be deleted from a bank—rather it should be coded as “retired.” This ensures that archival records are kept intact and enables test developers to avoid rewriting the same poorly performing item.

Testing Dates

The item bank should store the dates that an item was used and when it is scheduled for next use.

Equivalent Items

Stored with an item should be a list of “equivalent” (“similar” or “overlapping” or “mutually exclusive”) item IDs. An equivalent item contains similar content information or cues the correct answer. Once a particular item is selected for a test, equivalent items can be flagged so that they do not appear on the same test. A related list should include items that are different enough to be included on the same test, but too similar to appear on the same page. This is especially important if the test is assembled by the computer. When tests are manually assembled, content experts check for overlapping items; however, if test construction is automated, the only way to prevent overlap is by careful coding of the items in the bank. In practice, it may be impossible to do this in sufficient detail, and thus computer-generated tests should always be carefully reviewed by content experts before administration (Stocking, Swanson, & Pearlman, 1993).

Comments

A field for comments about an item is especially useful: Content experts may wish to comment on the relevancy of an item to the current field of practice; psychometricians may want to note an unusual statistic for a particular group or test administration. Reasons for retiring an item from the active bank should also be included in the comments.
Cases

A case is a graphic or a common piece of text (such as a reading passage) that is referenced by multiple items. Cases should be stored separately from the item and referenced by a case ID number. When an item is reviewed, the case should also be available for review. Conversely, when a case is reviewed, all dependent items (items that refer to the case) should be available. For some cases, all related items must be used together, and possibly in a specified sequence. For other cases related items may be separated. The item bank should be able to store information regarding any required sequence of items which share a common case.

Distractors

The correct answer, the number of “distractors” (“response alternatives” or “alternative choices”), and the weights for each distractor (if used) should be stored.

Statistical History

The item bank should store appropriate statistics for each administration of the item on each test form. Statistics for any group analyses performed should also be stored. Essential fields include:

- Test name
- Test form
- Date of administration
- Sequence number of the item on the form
- Number of candidates included in the analysis
- Number of candidates answering the item correctly
- Number of candidates omitting the item
- Group included in the analysis (males, females, first timers, all, etc.)

Optional fields might include:

- Classical statistics (e.g., discrimination and difficulty indices)
- Item Response Theory Statistics (e.g., item difficulty, standard error)
- Statistics for each distractor (e.g., weights and proportion of candidates responding to each option)
- Statistics to indicate differential item functioning (e.g., Mantel-Haenszel statistics, or IRT based DIF analyses)

Psychometric professionals can review item performance over time, and compare items individually or within categories. Storage for statistics should be user defined and have the capability for future expansion.

User-Defined Fields

Ideally an item bank should contain some user-defined fields to allow the licensure agency to store additional information unique to their specific needs.

COMPUTERIZED ITEM BANKING

In its simplest form, a computerized item bank is a word processing document containing the item text, and perhaps some simple scheme for identifying the item
author and content codes. Basic statistics, manually keyed into the document, may be included. In reality, this is probably the most common type of computerized item bank in use today (Gullickson & Farland, 1990).

Agencies that store items and related information in a word processing document, but have added merge codes to enable easy creation of final text, answer keys, and content distributions, are using a slightly more advanced computerized bank. This type of bank may be adequate for many organizations, but it requires a great deal of manual processing time and fails to take advantage of the potential power that today’s computer can afford the test developer. Although word processing may appear to get the job done (i.e., the test gets produced), it is still not the best tool for the job.

Relational Databases

A true computerized item bank must include a database component. A simple database affords minimal opportunities to sort items by content schemes, item difficulty, test administration date, etc. In the most basic system, a single database contains a single record for each item in the bank. Typically this record will contain an item identifier, content classification, and status (new, used, retired, etc). In many cases this record will also contain the item text itself, or a position indicator (such as a file name) where the text for the item is maintained. This is where many computerized item banks currently stop, and if each item is to be used only once, this simple item bank may be sufficient.

At a higher level of computerization, a fully relational database system can be constructed to maintain all information associated with the life of a test item. Relational databases can exponentially increase the functionality of the item bank; for example, the statistics obtained from each administration of an item can be stored in a related History database containing one record per test administration per item. The History database is a “child” to the main (parent) Items database. Thus for each item in the bank, there can be multiple history “children.” This type of relationship is also sometimes referred to as a one-to-many type system. The History database would minimally contain the number of persons who viewed the item and how many persons answered the item correctly. A more complete database would also include classical and IRT statistics such as item difficulty and item discrimination indices.

Another example of a related database in a full-featured item bank is a test database—a list of all tests associated with the item bank, including those already administered and those currently under construction. A related database to the test database would include an administration database consisting of one record per test administration per test (another parent-child relationship where the test is the parent and the administrations of that test are the children). This database should include administration dates, number of persons taking the exam, and pass/fail rate.

The concept of the “parent-child relationship” is quite important. In a relational database system each database is connected to at least one of the other databases through one or more “key fields.” In the case of the relationship between the Items and History databases, the key field is the item identifier. To “look up”
the history for that item, all items from the History database with the same item identifier are selected. This idea is not limited to the key field, but can also include additional filters; for example, to look up all of the times that an item was used in the last 2 years, both the item identifier and the date would serve as the filter. For organizations with extremely large item banks, the “look up” criteria can become much more complicated. The test developer may wish to locate all items (a) within a specific content area; (b) that have been administered at least two times to over 500 people; and (c) with item statistics in an acceptable range.

Complicated filtering conditions with large databases can be accomplished through “Structured Query Language” or “SQL”—a language relatively universal to all computer systems (microcomputers and mainframes). SQL is used to combine information from multiple databases in order to retrieve specific information. This language is not a programming language, but can be used by end-users (including psychometricians and clerical staff) to specify or “query” information whenever they need it. Filter capabilities for selection purposes can be maximized by using SQL-type queries rather than setting an actual filter on an extremely large database.

**Maximizing Computer Efficiency**

Two factors of speed influence the efficiency of the computerized item bank: (a) the speed in which data can be moved from the hard disk to the program and ultimately to the screen; and (b) the speed in which data can be found. The first factor is dependent upon the quality of the computer hardware and operating system. For instance, operating on a network, or from within Microsoft Windows™ greatly slows data access compared to using a stand-alone computer operating under DOS. Increasing local random access memory (RAM), and the inclusion of a fast local hard disk will serve to greatly improve speed in this regard.

The second factor—the speed in which data can be found—is largely dependent upon the quality of the underlying software being used. For example, in the case of a moderately difficult query applied to a 5,000-item bank, a poorly constructed query system could take long minutes or even hours to find the items that are needed. The same query would take a fraction of a second if programmed correctly. Computer Adaptive Technologies, Inc. (Gershon, 1994) recently demonstrated this speed advantage to one of the major national testing organizations that was using a mainframe to load up to 30,000 person records to their system each week. The loading process was taking up to 6 hours. Improved software on a personal computer enabled them to download all records and sort them in just under one minute.

**Text Editing**

Another component that can make a computerized item bank more useful for the test developer is an integrated word processor. Text editing within database software is usually awkward; however, if the item bank software integrates a word processor with the relational database system, item editing and paper-and-pencil test production become infinitely easier. This type of integrated system should
provide the test developer with all of the database capabilities plus state-of-the-art word processing capabilities including access to thousands of fonts, spell checking, formatting options, styles, codes, etc. When the item is actually stored as a word processing file, the stem and the distractors can be stored together, and the item is edited as one contiguous piece of “What-You-See-Is-What-You-Get” (WYSIWYG) text.

An efficient item bank will communicate directly with the word processor using features like the Windows Dynamic Data Exchange (DDE), which enables a database to start up a popular word processor such as Microsoft Word or WordPerfect just once at the beginning of the session. Later editing calls to the word processor result in a simple transfer of the item or test text to the word processor using DDE. Thus, the word processor does not need to be restarted for each editing procedure. This is particularly important within a Windows operating environment, where it may take up to a minute each time the program is started.

The item bank should also be capable of transferring item text back into the item bank without exiting the word processing software. Some item banks use “dedicated” word processing programs; in such instances, the word processor used was written specifically for the item bank and therefore cannot fully conform to any particular industry standard word processing program. This type of system results in increased learning time for the users, and undoubtedly means that there are significantly fewer editing features available. Given the power of today’s computers, there is no reason to settle for poor performance in text editing. Minimal editing requirements should include access to spell and grammar checkers, multiple fonts, columns, subscripts and superscripts, bold and italics, equation editors, etc. Medical and legal licensing boards should also have easy access to available custom spell checkers.

The ability to edit on-line combined with the portability of computers allows an item bank to be edited at virtually any location. Some organizations already use portable computers to transport their item banks to remote sites, allowing test committees to participate in item writing sessions and draft test production.

**Integrating Graphics**

Whenever possible, graphics should be stored “on line” in order to facilitate easy layout and graphical editing. There are several excellent graphics editing programs available on the market today, which can be used to produce graphics from scratch as well as edit or enhance scanned images produced by other sources. A graphic should be stored in a format that can be used to produce camera-ready copy, and later to produce a screen image for use in computer-administered testing. Many images scanned from a paper image require at least some manual editing before they are suitable for screen display, but this is not usually the case when the printed image is originally prepared on the computer. The layout file that was previously used to produce a camera-ready paper image can usually produce a suitable screen image as well.

A well-constructed item bank will include the capacity to bank images as well as text. And when an image is used within the text of an item, object linking and
embossing (OLE) can be used to edit the graphic without leaving the item bank. OLE enables graphics created by other programs to be imported directly into a test item. “Live” OLE links can also be created; for example, when spreadsheets are used to produce graphics and numbers for multiple items, a simple edit in the original spreadsheet will result in all of the items in the item bank being automatically updated as well. When editing an item on screen, the graphic can be moved and resized without leaving the item banking software. OLE further allows the image to be edited by simply clicking on the image. This results in the original program which created the graphic to appear on the screen so that the desired change can be made. When editing has been completed, the item should reappear on the screen with the modified graphic appearing in place. This type of functionality can savedozens of hours of test construction and layout time on a single test!

Multiple Language Support

Some licensure agencies administer tests in more than one language. An item bank with integrated word processing software that supports foreign languages is essential for developing test forms in alternate languages. Identical items available in more than one language can be stored under the same item ID, particularly when the item is expected to perform similarly regardless of base language. Such is likely to be the case with mathematical and short-answer items, but usually not true for items with long text passages. When a test is created in English, a comparable test in the second language can be automatically generated.

Automated Item Writing

The item development process can be facilitated by integrating an item writing diskette with the item banking program. An item writing diskette allows content experts to write items at home directly onto a diskette. Item writers enter the item identifier, item text, correct answer, comments, and references. This type of software can be configured to present the item writer with any classification scheme created by the licensure organization. The contents of the diskette are then imported directly into the item bank saving manual entry time, eliminating typographical errors, and ensuring standardized formatting. Diskettes can be exchanged among item writers and team leaders to facilitate the item writing process.

Item Bank Capacity

Item banks are theoretically capable of storing an infinite number of items. Licensure agencies need an item bank that can store all of their items and allow considerable room for expansion. This is a function of both the limits of the software as well as the hardware. A minimal configuration will include at least twice as much hard disk space as would be required to store all existing items, plus all of the items likely to be created in the next 5 years. (Note: Twice as much space will be needed to perform database maintenance functions.)

Statistical Analysis

Some item banks have the ability to analyze test results as part of the original software; others interface with a statistical module. At the very least, an item
A good item banking program will provide the licensure agency with user-defined options for pool book production. A pool book is effectively a printed copy of the contents of the item bank (for obvious reasons a pool book is not usually referred to as a bank book). The printed pool book may include all of the items in the bank or a user-defined subset. In addition to the identifying number, item text, and correct answer, each pool book may optionally include item classification, statistical history, item author, reference, and comments about the item. Case text and graphics should be included with each item, and the software should allow layout options such as printing each item on a separate page. Item pool books can act as an archival “hard copy” or can be used by test committees to aid them in item writing and review.

Security

Item banks for high-stakes licensure examinations should have a user-definable, multilevel security system. In most agencies, different levels of personnel will need access to the item bank. For example, some clerical personnel will only need access to item text to enter new items, whereas psychometric staff require access to test definition and test layout as well.

Similarly, many organizations maintain multiple item banks for different examinations. In many cases, some of the persons working on one item bank have no need to have access to the other banks. In larger testing organizations, security will need to be cleared on the test level, such that once a test has been created, embedded items can be modified only by project managers. Typically, limited access to various parts of the item bank will be automatically maintained by a password system that identifies the user when the item banking software is executed. The software should then be responsible for limiting access as appropriate.

Security can be improved even further through a variety of means, including limiting access with the use of regular network security, external hardware keys, and embedded encryption (the process of scrambling or “encoding” text to make it impossible to read without a proper password). A hardware key scheme prohibits access to a system unless the key (a small box connected to the printer port of a user’s computer or to the network file server) is attached. Typically, the key must be present and a password given for access to be granted. In this way, even if the entire item bank is stolen, the software will refuse to reveal the item text unless both the correct password is entered and the hardware key is present.

USING AN ITEM BANK TO CREATE AND ARCHIVE TESTS

Once items have been entered in a computerized item bank, paper-and-pencil tests or computer-administered tests can be created. A computerized item bank can be used to automatically create camera-ready copy for paper-and-pencil tests or computerized tests, store all previously administered tests, easily create new test
forms, and store overall test statistics such as dates administered, number of candidates examined, reliability, etc.

**Automated Item Selection**

Items for inclusion on a test can be selected manually, randomly drawn by the computer from all existing items in the bank, or drawn by the computer from prespecified parameters. There are currently a number of highly sophisticated schemes for automated test construction (Armstrong, Jones, & Wu, 1992; Boekkooi-Timminga, 1990; Stocking et al., 1993). Although some licensure agencies may wish to pursue these advanced algorithms, most test developers are satisfied to use less complicated item-selection algorithms that choose items within prespecified parameters such as content, item type, and item difficulty.

Typically, a test will be prepared to include a specified number or percentage of items from various content domains. In the case of a computerized item bank, these conditions can be defined by creating a computerized test plan (“blueprint” or “template”). Each cell in the test plan describes how many items must be included to fill a particular condition (for example, there must be five items from content area 1, and seven items from content area 2). These are unique conditions because they refer to rules which apply only to a single cell in the test plan. There can also be parallel sets of conditions, such as a condition that 50% of the items on the test must be new, and 50% must have been contained on a previous test. There are also total conditions which apply to all items on the test, such as a rule that all items must have been approved by a specific committee, or that all items must fall within a specified difficulty range.

Once the test plan has been created, it should be accessible whenever creating a new test. The item bank should be able to use the test plan automatically to pull items from the bank which fill the test plan conditions. A good banking program will also be able to conduct an “audit” of items manually selected for inclusion in the bank, to ascertain whether or not all of the conditions in the test plan have been met. On-screen warning messages or a written report should inform the test developer if and where insufficient items are available to meet the plan. This procedure can be accomplished in microseconds on the computer, but would otherwise take hours or even days when completed by hand.

The item bank should have the capability to reorder automatically—or allow for easy manual reordering—existing tests to create new test forms. When an item is selected for test inclusion, any graphics, tables, or cases associated with it should automatically be included.

**Camera-Ready Copy for Printed Tests**

Computerized item banking software should provide the capability to edit a test created from within the bank using standard word processing features such as spell check and the ability to change font type and size. An advanced bank will also be able to reorder test items so that the blank space typically left on some pages is minimized. This automation component can save days of manual layout work, while ensuring accuracy of item keys. Test administration formats, such as
instructions, examples, and layout should be stored in separate electronic files for easy import into a test document. An additional essential feature is the ability to produce both paper-and-pencil and electronic answer keys.

Computerized Tests

A modern item bank will be able to produce tests for both computerized and paper-and-pencil administration. Most certification and licensure organizations have at least contemplated using computers to administer their tests. Although there are a myriad of reasons for and against this approach, the important consideration for the purpose of this chapter is that agencies would be well advised to purchase or create an item bank which has the capability for the creation of computer-administered tests.

Tests that contain only items with short text require almost no user intervention to be included on a computer-administered test, as long as the items are stored in an appropriate item bank. Items containing longer text passages are also simple to convert, although the choice of administration software and hardware may narrow. The differences in the two modes of administration are most apparent when it comes to graphics. As mentioned earlier, most graphics or visuals that have been prepared for paper-and-pencil test administration are not directly transferrable to computerized administration. The relatively limited resolution of any computer screen compared to a printed page may necessitate some editing of the paper-and-pencil graphic. If a bank is to be used to produce both paper-and-pencil and computer test forms simultaneously, the bank of visuals must be prepared to store both print and screen versions of each illustration.

A quality item bank will allow the production of a paper copy of the computerized exam as well as the computer-administered version. This can be used to produce parallel versions or a paper copy of the computerized test for proofreading purposes.

It is also wise to ensure that item banking software integrates well with test administration software. Banking software should be capable of easily producing output files compatible with test administration software. Test administration software should not only be compatible with the item bank, but should also be functional regardless of the test administration vendor. In the event that the licensure agency changes to a different administration vendor, compatible software ensures that a painful translation procedure, which could even result in the need to repilot computer administered test items or ultimately force renorming of the test, can be avoided.

Archiving Tests

A “test bank,” consisting of all the tests that have been created within the item bank, is an important part of a complete item banking program. The text of the test should be stored with all of the historical statistics for each time the test was administered, including dates of administration, number of candidates examined, and test reliability. Group analyses and DIF analyses should be included with the test’s statistical history, and a comments field should also be available to store comments relating to overall test performance.
IRT CALIBRATED ITEM BANKS

If a licensure agency tests a minimum of 100 to 200 candidates per year, calibrating and equating their item bank using an Item Response Theory (IRT) model will provide additional valuable statistical information. An IRT model compares the difficulty of the item with the ability of the candidate and estimates the probability that the candidate will correctly answer the item. The major advantage of IRT models over classical test theory is that classical item and test characteristics (or statistics) vary depending upon the group of candidates taking the test whereas IRT item and test characteristics do not. Classical indices of item difficulty, point-biserial correlation, and reliability may all change if candidates differ in ability distribution (Hathaway et al., 1985). In licensure testing this often proves to be true; for example, a spring candidate population may be more able than a winter candidate population. An IRT model, such as the Rasch model or the three parameter logistic model, allows for the calibration and equating of items onto a common scale and also allows for the identification of items that perform poorly. To calibrate items with the Rasch model, however, requires a candidate population of at least 100 to 200 candidates (Linacre, 1994), whereas the three parameter model requires 1,000 to 2,000 candidates to estimate item parameters (Green, Bock, Humphreys, Linn, & Reckase, 1984).

New items are equated to the bank scale by administering them on tests with previously calibrated items from the bank. This procedure is called common-item equating (Wright & Stone, 1979). Other methods such as common-person equating for linking IRT item parameters onto a common scale are discussed by C. David Vale (1986) as well as in Chapter 12 of this book. Before items are added to the calibrated bank, the fit of the items should be assessed to determine their suitability for inclusion.

Using an IRT model as a measurement system requires that the group of items be “unidimensional.” This means that all of the items in the bank are defining one dimension (e.g., the ability to practice law). The bank of calibrated test items is a set of coordinated questions that develop, define, and quantify a common theme and provide an operational definition of the dimension (Wright & Bell, 1984). Of course, unidimensionality is an abstract idea, always violated to some extent in real life. Many licensure tests comprise items from different content areas—indeed the validity of the test is assured by the inclusion of items that are representative of these different areas as specified by the test blueprint. However, in most cases, the rules underlying item response theory are quite robust, and the items for a licensure examination can be calibrated with an appropriate IRT model. Still, it is highly recommended that an IRT expert be consulted when making initial decisions regarding unidimensionality and the appropriateness of using existing items when creating an IRT-based item bank.

When an IRT measurement system is used, a measure of the precision of the item calibration and ability estimate is available for each item and each candidate. This makes it possible to calculate a priori a reliability estimate for any score on a test drawn from the calibrated bank. The size of the error of measurement will depend on which items are selected, how many items are selected, and the candidate’s raw score on that set of items (Hathaway et al., 1985).
There are a number of advantages to using an IRT model to calibrate and equate all items within a licensure test bank:

- Easy preparation of parallel test forms
- Comparison of individual candidate performance over time (for candidates who repeat the test)
- Comparison of group performance over time (to evaluate overall candidate proficiency or proficiency by school, program, or specific content area)
- Usage of the item bank for computerized adaptive testing

Creating an Item Bank for Computerized Adaptive Testing

Computerized adaptive testing is a form of test administration in which each candidate takes an individualized test administered on a computer. Candidate competence is continually assessed on-line, and the difficulty of each item administered is targeted to the current ability estimate of the candidate. This mode of testing typically requires an IRT calibrated item bank.

The ability to order all of the items on the same scale is essential for computerized adaptive testing. Because all items are on the same scale in an IRT calibrated item bank, the particular items that are administered to a given candidate are irrelevant. Each individualized adaptive test created from the calibrated bank is automatically equated to every other test that has been or might be drawn from the bank (Wright & Bell, 1984; Masters & Evans, 1986).

When all items in the bank are calibrated to the same scale, a pass/fail point, (criterion-referenced standard) can be established for the entire item bank. Thus, all candidates are measured against the same criterion-referenced standard regardless of the group of candidates with whom they are examined, the particular set of items they are administered, or when they take the test. This makes it possible to determine a candidate’s pass/fail status with respect to the basic dimension that the items define.

To use an item bank for computerized adaptive testing, the bank must meet additional constraints. Following are some observations for maintaining item banks for computerized adaptive testing suggested by Mary Lunz, Ph.D., Director of Testing for the Board of Registry, American Society of Clinical Pathologists (Lunz & Deville, 1994).

Proportional Distribution

Items in the bank should be distributed proportionally to the test blueprint. For example, if 10% of the adaptive test will be drawn from a specific content area, then approximately 10% of the items in the bank should cover that content area. Most adaptive test algorithms allow for content balancing so that the items administered to each candidate follow content percentage specifications. Adherence to the test blueprint ensures that test validity, as defined by a job analysis and content experts, is maintained. When some content areas in the bank have fewer than the blueprint-specified percentage of items, the existing items will be over sampled.

Range of Difficulty

The range of difficulty of items in the bank should reflect the range of ability of the candidate population. Calibrated item difficulties should be adequately
distributed within each content area as well as across the entire item bank. Because each candidate is being administered an individualized test in which the difficulty of the items presented varies according to the estimated ability of the candidate, the range of items available for selection by the computer must adequately cover the distribution of candidate ability. When the range of calibrated item difficulty is adequate, the bank can provide appropriately targeted examinations, thus increasing measurement precision and therefore, increasing the amount of information gained about the candidate (Bergstrom & Stahl, 1992).

**Current Relevancy**

Items in the bank must be carefully screened for current relevancy to the field of practice. Because any item in the bank may be selected by the computerized algorithm for administration, outdated items must be removed from the active bank.

**Security**

For security purposes, the more high-quality items in the bank the better. Large numbers of items limit the number of candidates who are exposed to any one item (Stahl & Lunz, 1993).

**Long Term Maintenance**

Item bank maintenance is especially crucial when item banks are used for adaptive testing. The estimated ability of the candidate is calculated from the bank parameter values for the items. Thus, the item bank must be continually monitored for both parameter drift and relevance of all items to the current field of practice. When a bank is used to create paper-and-pencil tests, items that appear on a particular test form are checked by content experts to ensure that they are “good” items. Because this is not the case with adaptive tests, agencies administering computerized adaptive tests must have scheduled, systematic reviews of all items in the bank.

**NEW DIRECTIONS IN ITEM BANKING**

Computerized testing, multimedia, and integration will be the major themes in the item bank of tomorrow. Computerized testing is just beginning to take the world by storm. There are at least three national networks owned by computer administration vendors, and dozens more that belong to individual corporate and certification organizations. In many cases, vendors and government agencies have set up multiple testing centers which blanket a given state, allowing easy and constant access for all examinees.

Fixed length and adaptive computer tests are currently administered to hundreds of thousands of individuals and the numbers are expected to grow exponentially into the 21st century. Some of these examinees are taking tests on antiquated main frame or hand-held computers with limited display or memory capacity. The future will allow all candidates to take their tests on computers with color monitors, full-size keyboards, and answering devices like mice or touch screen panels. But the real revolution will occur when tests are readily capable of displaying situations that are more real-life oriented using technology such as video clips and actual
audible conversations. Advanced technology is already available to provide a computer “reader” or to automatically extend testing time limits in order to satisfy the Americans with Disabilities Act. In the future, computers will also be used to record verbal answers to questions, or even to “video record” responses to test items.

All of the above scenarios will require item banks to have new features including the ability to store sound and multi-media clips along with item text and to interface with test administration modules that include these multimedia components.

CONCLUSION

Most licensure agencies can probably streamline their test production through the use of computerized item banking. Agencies should carefully review their needs, taking into account the number of items in their bank and their procedures for item review and test production.

Options for acquiring software for computerized item banking include:

- Developing customized software. This option has the advantage of providing the agency with the precisely unique specifications they require. Unfortunately, it is also usually associated with high development costs.
- Purchasing off-the-shelf software. Sufficient for most testing organizations, off-the-shelf item banking software varies greatly in price. Purchasers should keep in mind that their item banking needs will increase as additional items are written and tests administered. Therefore, the greater the flexibility in the program and the greater the speed and capacity of the software, the longer the item bank will fulfill their requirements.
- Customizing off-the-shelf software. Some software developers are willing to customize their software. This may provide a good solution for organizations with unique requirements and result in significant savings in cost and aggravation over “from scratch” software development.

Prior to purchasing item banking software, careful consideration should be given to present and future item banking needs. Testing agencies should request working demonstration copies of the software products and compare features including storage capabilities, speed, and ease of item text editing and test production. It would be prudent to involve those personnel who will actually have to use the software—including psychometricians, content specialists and clerical staff—in the decision about which product to purchase. From item creation to test administration—use of computerized item banking can capitalize on advanced technology to streamline production procedures and construct psychometrically sound tests.

REFERENCES


