Automated Item Banking and Test Development Model used at the SSAC.

Tural Mustafayev
The State Student Admission Commission of the Azerbaijan Republic
Item Bank Department

Item Banking

For many years tests have been administered for various reasons, such as educational testing or assessing particular skills. Results of these tests are used, for instance, as a factor in determining whether an examinee should be admitted to study at an educational institution. Such tests are typically administered to a large number of examinees. Due to some concerns, including administration costs and security, automation of the entire testing process becomes an issue of great importance. Automated item banking and test development make a significant contribution to this process.

Test constructed from item bank is a very useful tool in educational and psychological measurement. An item bank is a large collection of test items all measuring the same domain of knowledge, stored in a computer storage together with their parameters. Known item parameters allow the test constructor to have strict control of them and make it possible to search for items according to a number of criteria.

The goal of the testing is to construct a test efficiently for the purpose of measuring a skill, ability, etc. Therefore, each test is constructed to conform to a test specification, which defines the rules and/or constraints for selecting the items. In constructing a test, items are selected from item bank, so that the combination of selected items satisfies the test specification.

A test is typically divided into sections of questions. The test specification generally defines the number of items to be presented in the test, the number of test sections, the number of items in each section, and the time set for taking the test. The test specification also defines criteria for item selection. These are based on item characteristics, which include, for example, item content , e.g., mathematical questions relating to arithmetic, algebra, or geometry and parameters of items, difficulty of the selected items as an example.

The approach to design of an automated item banking and test development system can be envisioned as an integration of three phases. Phase I involves the development of an item bank that contains each item's content and parameters, and has full editing capabilities. The parameters include, but are not limited to, item statistics, such as item difficulty, and statistics of the test where the item appeared. Phase II provides the capability to retrieve items from the item bank in order to construct tests with prespecified characteristics. Item parameters are used for this purpose. Finally, in Phase III, the actual test booklets are published with specific format.

In order to provide these three phases SSAC has developed the in-house Automated Item Banking and Test Development system. This system consists of two parts. The main part of the system is the item bank consisting of items stored in computer storage as MS Word files. This item bank is subject-based, i.e. items are grouped by subjects. This bank can also be considered as a set of independent subject-based item banks. The organizing structures of the bank have been designed and developed by discipline committees of teachers. Paying particular attention to what is now being taught throughout the Republic, they have developed comprehensive course outlines. Test items have been collected from a large number of teachers. The second part of the system is the software suite designed to sort and register items, to get statistical report about the state of item bank, to retrieve items from the bank in order to construct tests with prespecified characteristics, to provide item editing capabilities, to compile and format test booklets, to get a list of answers to a test (key) in different formats, and to add usage information and statistics of the test where the item appeared.

Since 1994 SSAC (State Students Admission Commission of the Azerbaijan Republic) has been developing item bank for administering entrance examinations to Bachelor's level of higher schools. Today SSAC is running the item bank consisting of over 100 thousand items in 17 subjects and this bank is in the process of continuing development. Besides, starting from 2005, SSAC has been developing item bank intended for use at entrance examinations to Master's level of higher schools.

Classification of item parameters and item format

Each item in SSAC's item bank has 5 parameters, namely a subject and the topic within that subject, degree of difficulty, cognitive level called "relevancy", indicator of previous usage called "usage marker", and indicator of similarity called "analogousness". As stated before, items are grouped by subjects and each subject domain is divided into a number of topics in accordance with school curriculum. For each item subject experts determine the most corresponding topic. They also assign the degree of difficulty and cognitive level to each item. Subject experts use 3 difficulty levels (easy, moderate and difficult) to characterize item difficulty. Further, their estimations are checked and revised when an item is used on a test administered either at preparatory or admission examination. Subject experts use 7 cognitive levels to distinguish the different thought processes required to respond. The cognitive level reflects the manner in which knowledge is being assessed, rather than the technical difficulty of the content. Items coded to cognitive level I (knowledge of terminology) elicit knowledge of terminology. Items coded to cognitive level II (knowledge of facts) elicit knowledge of specific facts, methodology, principles, theories and structures, and ability to memorize, repeat and identify. Items coded to cognitive level III (reasoning based on generalization) elicit the ability to generalize and differentiate specific facts, principles, theories and structures, and the skills such as comparison and classification. Items coded to cognitive level IV (interpretative reasoning)

elicit the ability to interpret events or facts rather than naming them, often involving cause/effect relationships. Items coded to cognitive level V (predictive reasoning) elicit ability to predict consequences. Such items describe a specific or hypothetical situation, and require forecasting consequences when certain period of time passes or when some factors determining a situation are changed. Items coded to cognitive level VI (modeling or application) elicit ability to use information, methods, concepts and theories in new situations, to solve problems that have single or best answers by applying acquired knowledge, and techniques and rules in a different way. Items coded to cognitive level VII (calculation) elicit ability to solve problems using mathematical principles and formulas. In this sense calculation can be regarded as one of the forms of application.

Each item has also a usage marker assigned on the basis of its latest usage and indicator of similarity assigned to indicate whether an item has an analog or parallel, i.e. an item or items with content, difficulty and cognitive level similar to its. Generally, a used item is not included in the tests administered in the next two years. Analogous or parallel items are used to build parallel (equivalent) test forms.

At present SSAC administers paper and pencil tests using traditional multiple-choice items. Items are written by teachers and reviewed comprehensively by the experts in the field. Each item consists of a unique number called “item code” that identifies the item and all of its parameters, a stem containing the introductory question or partial statement that the examinee must answer or complete, five alternatives, marked by the letters A, B, C, D, and E with a comment. One of the alternatives, the key, is the correct, or best, response to that stem. A comment contains encoded information about an item writer and reviewers, as well as usage information and statistics of the tests where the item has appeared. It may also include solution of a problem presented in the item.

Automated test development

The SSAC staff has developed a software titled “Test for Windows” that consists of several integrated procedures designed to handle an entire spectrum of test development process. This software enables to sort and register items, to get statistical report about the state of item bank, to retrieve items from the bank in order to construct tests with prespecified characteristics, to edit drawn items, to build parallel test forms, to compile and format test booklets, to get a list of answers to a test (key) in different formats, and to add usage information and statistics of the test where the item has appeared.

Prior to using the item bank, it is processed by special software to verify that each item has a correct code and all the required elements (i.e. stem, alternatives and comment). After that each item in the bank is registered. Registration consists of performing code-based sort on the items (optional) and marking all elements of all items for an easy access. Registration procedure also builds a database containing full information about each item necessary for running the

following procedures. This database is first used to gain statistical report called “map of the bank” – a three dimensional table reflecting the distribution of items by difficulty and cognitive levels within each subject topic. This information together with the information received from experts in the field is used for automated test construction. First, a group of subject experts prepares a blueprint of the objectives across a given content area and defines the number of items representing each difficulty and cognitive level at the test, as well as subject topics or topic groups that should be represented in the test. Generally, for each subject the number of topics is greater than the number of items that represent given subject in the test. For this reason experts assemble close topics into groups so that the test covers the whole curriculum. On the basis of information provided by subject experts a special software randomly generates a request called “technical task” using map of the bank. This request contains parameters of each item to be included in the test. Based on this request, procedure called “Generator” randomly retrieves items from the bank. “Generator” is capable of simultaneously developing parallel test forms, i.e. test forms containing items that are similar in terms of content and parameters by making use of analogous items. Test forms generated by “Generator” are called “initial” or "original forms”. These forms are reviewed by the experts once more. After this final expertise these forms are processed by the procedure called “Generator of Forms”. “Generator of Forms” generates a specified number of test forms called “derivative forms” from each initial form by randomly changing both the position of items and the position of alternatives within each item in the initial form. This contributes to preventing cribbing during an examination while providing identical items on the test. “Generator of Forms” also provides for control over the frequency of occurrence of each of the alternatives as a key. “Generator of Forms” generates derivative forms for each subject administered in the test. After that a procedure called “Booklet Compiler” compiles test booklets from derivative forms. This procedure also formats test booklets and makes them ready for use at the examination.

“Test for Windows” also contains three procedures that are used after an examination. “Key Generator”, first of them, enables to get a list of answers to a test (key) in different formats. First format is designed for experts who are invited immediately after the examination to confirm correctness of items and the key. Second format is designed for publishing and, thus, is more compact than the first one. Third format is designed to provide the source information for software which handles scoring of each examinee's responses. “Usage Markers”, the second procedure, enables to add usage information to items used on the test. As stated before, a used item is not included in the tests administered in the next two years. Third procedure, “Post-examination statistics” enables to add statistical data to the item reflecting the performance of that item on the test. This data consists of the percentages of examinees responding correctly and incorrectly to that item and is later used to revise the difficulty level of that item.

Future plans

As stated, SSAC's Automated Item Banking and Test Development system consists of two parts: item bank stored in computer storage as MS Word files, and item bank management and test development software. Storing test items in MS Word files was preferred due to the ease of use as MS Word provides tools for text editing, creating mathematical formulas (Equation Editor) and graphics (Word Picture), and insertion of bitmap images (OLE standard). But on the other side, this model has several limitations. It is difficult to maintain the rapidly growing item bank using this model. Another difficulty is with adding new fields of data associated with the item. These limitations can be overcome by building central item database using one of the DBMS software (Oracle, Sybase, etc.). This solution will contribute to designing a more flexible structure for item bank, and provide more advanced data integrity and data security mechanisms. In the nearest future SSAC is going to build a central item database on the Oracle platform. Associated with each item stored in the central item database, will be the data stored in fields related to the revisions the item has undergone, a list of the test forms in which the item has previously appeared, and statistical data indicating how the item performed at each previous administration. Other descriptive data fields will include information related to the item type (i.e., multiple choice or open-ended), the item's author and editors, solution of a problem presented in the item, subject and topic information, and both difficulty and cognitive levels.

At present, SSAC administers only paper and pencil tests. But in the nearest future it is going to provide computer-based tests for admission to master's level of higher schools. Computer-based testing has several advantages over traditional paper and pencil testing. The biggest advantage of using computers to deliver tests is the level of control over the conditions of the testing session it gives to instructors. For example, the amount of time used for the test and the type of feedback (either on individual questions or on the entire test) can be varied. Such timing information can assist test designers in eliminating bad or poorly worded questions and in comparing the knowledge level of test-taking candidates. A second set of advantages includes the variety of multimedia learning objects that can be incorporated into test questions and the fact that most question types can be scored instantly. Using computer based testing also enables the use of adaptive tests. Adaptive tests are tailored to the individual ability of the test-taker and are able to provide somewhat better estimates of ability with shorter tests. SSAC staff is working on implementation of a computer-based testing in admission examinations in a few years to come.