Graduate
Management
Admission
Council®

The MEDPRO Project: An SBIR Project for a Comprehensive IRT and CAT Software System — CAT Software

Nathan A. Thompson
Assessment Systems Corporation

Presented at the U.S. Government Supported CAT Programs and Projects Paper Session, June 2, 2009

Abstract

Development of computerized adaptive tests (CAT) requires a number of appropriate software tools. This paper describes the development of two new CAT software programs. CATSIM has been designed specifically to conduct several different kinds of simulation studies, which are necessary for planning purposes as well as properly designing live CATs. FastCAT is a software system for banking items and publishing CAT tests as standalone files, to be administered anywhere. Both are available for public use.

Acknowledgment

CATSIM and FASTCAT are the CATPRO component of the MEDPRO system, which is being developed by Scientific Software International, Inc., in collaboration with Assessment Systems Corporation, and has been funded by SBIR contract HHSN-2612007-00013C with the National Cancer Institute of the National Institutes of Health. Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Citation

Thompson, N. A. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system — CAT software. In D. J. Weiss (Ed.), ***Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.*** **Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/**

Author Contact

Nathan A. Thompson, Assessment Systems Corporation, Suite 200, 2233 University Avenue, St. Paul MN 551`14, U.S.A. Email: nthnmopson@assess.com

The MEDPRO Project: An SBIR Project for a Comprehensive IRT and CAT Software System — CAT Software

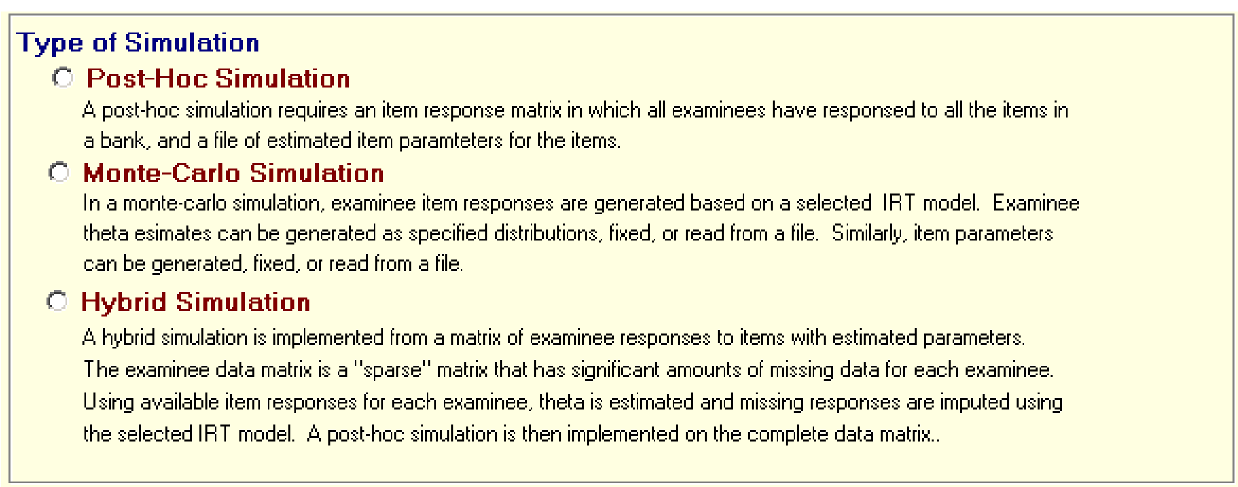
An important line of medical research concerns patient-reported outcomes, which entails the use of data provided directly by a patient regarding the results of a treatment, rather than only information from the treating physician. Computerized adaptive testing (CAT; van der Linden & Glas, 2000) is a method of delivering and scoring tests with numerous advantages, including decreased test length with an equivalent or better level of precision than conventional fixed form tests. Previous research has shown that the advantages of CAT are applicable to research on patient-reported outcomes (e.g., Immekus, Gibbons, & Rush, 2007).

An important step in being able to apply CAT to research on patient-reported outcomes is the development of necessary software for data analysis, research, and actual delivery of the tests. Thissen (2009) described the development of software for item response theory (IRT) analysis. This paper describes the software for CAT. There are two major components: CATSIM, a program for CAT simulation research, and FastCAT, a CAT delivery platform. Both components include options for using IRT models developed for polytomous items (e.g., rating scales), which are frequently used in research on patient-reported outcomes.

CATSIM

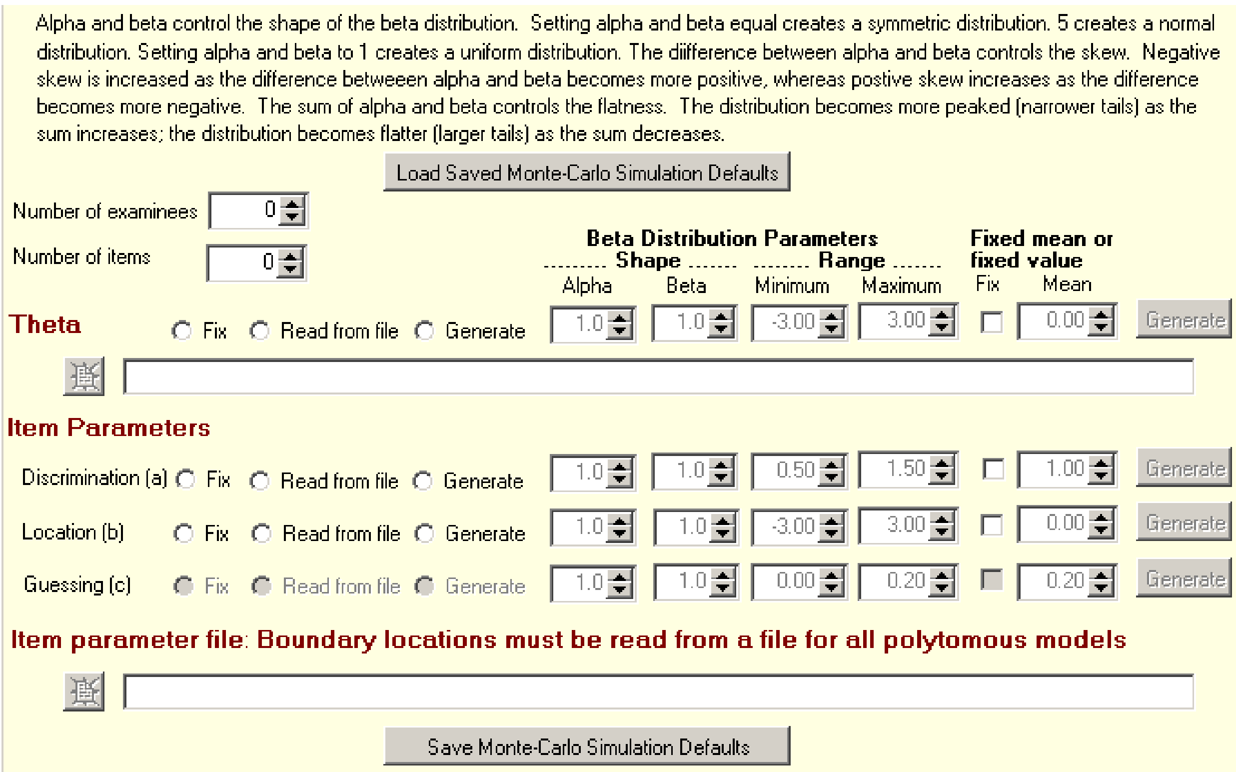
CATSIM is designed to perform simulation research necessary to develop and publish CATs. It is able to produce three types of simulations, as seen in Figure 1. Monte-carlo simulation is appropriate when no examinee data is available for research. Post-hoc simulation is appropriate when there is examinee a complete item response matrix matrix. Hybrid simulation (Nydick & Weiss, 2009) is appropriate for the common situation where examinee data is available but is in an incomplete matrix. For example, if a bank of 300 items is developed but each examinee is pilot tested with only 50 or 100 items.

Figure 1. CATSIM Simulation Options



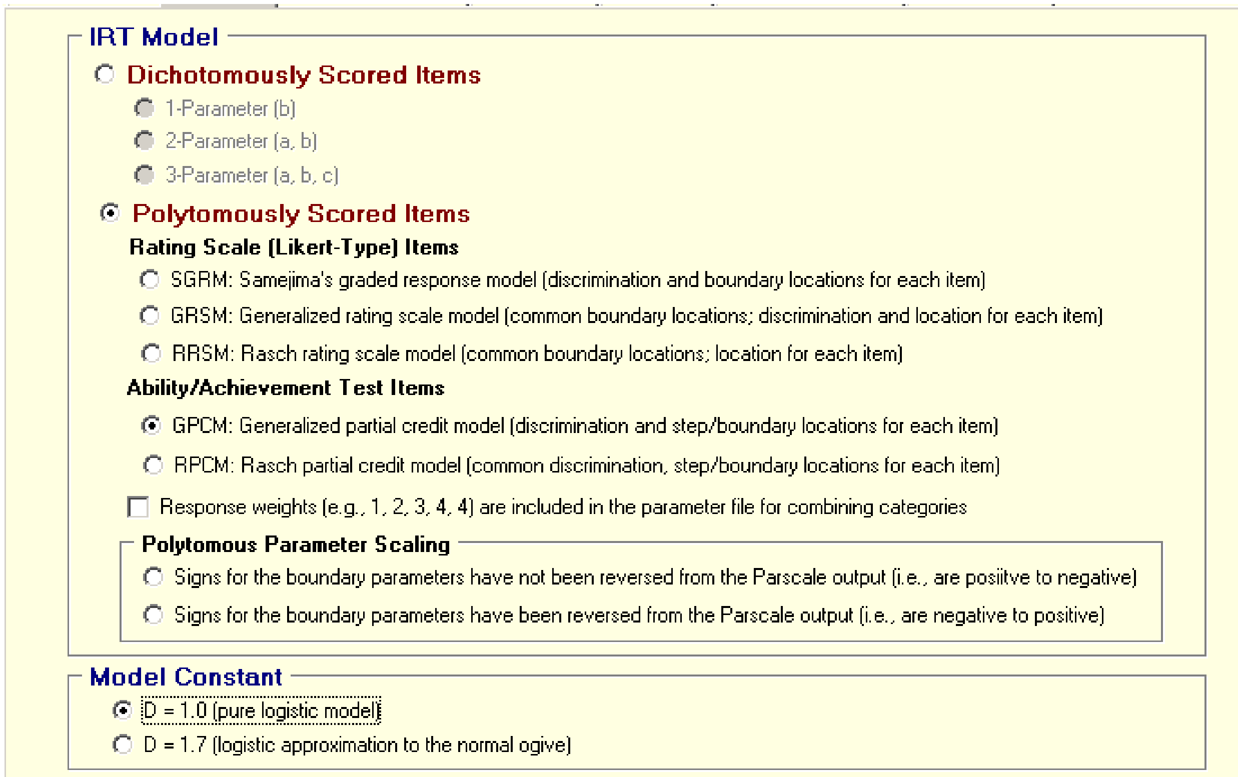
If monte-carlo simulations are being utilized, there are a number of parameters that must be provided in order to generate a set of item responses, namely person (θ) parameters and all item parameters for a given IRT model. These parameters can either be provided by the researcher based on previous research or data, or they can be randomly generated, or fixed to a constant value. Figure 2 presents the options for generating all relevant parameters with the beta family of distributions, which allows for a wide range of distributions.

Figure 2. CATSIM Monte-Carlo Options



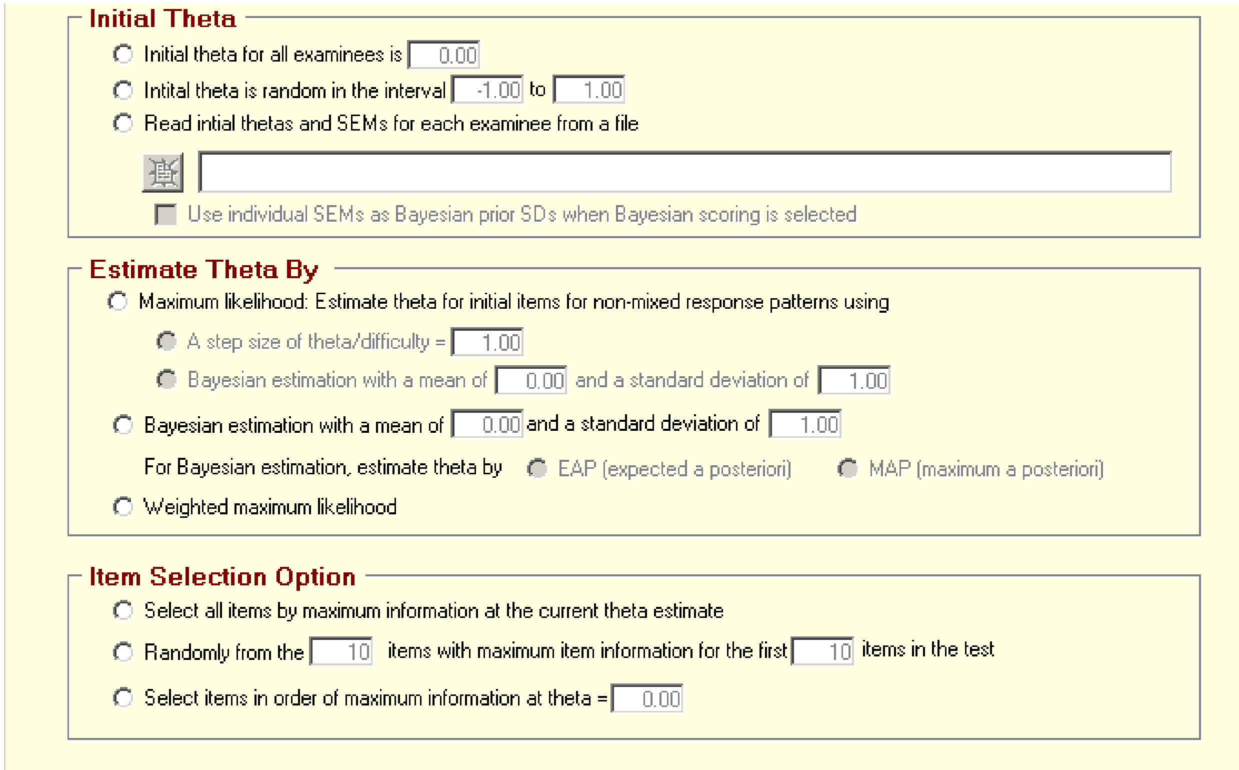
Once a data set is specified, either from real examinees or monte-carlo simulation, CATs can be simulated for the examinees in the data set. The first specification for simulating CATs is to determine the assumed IRT model. Figure 3 presents the options for eight IRT models supported by CATSIM.

Figure 3. CATSIM IRT Model Options



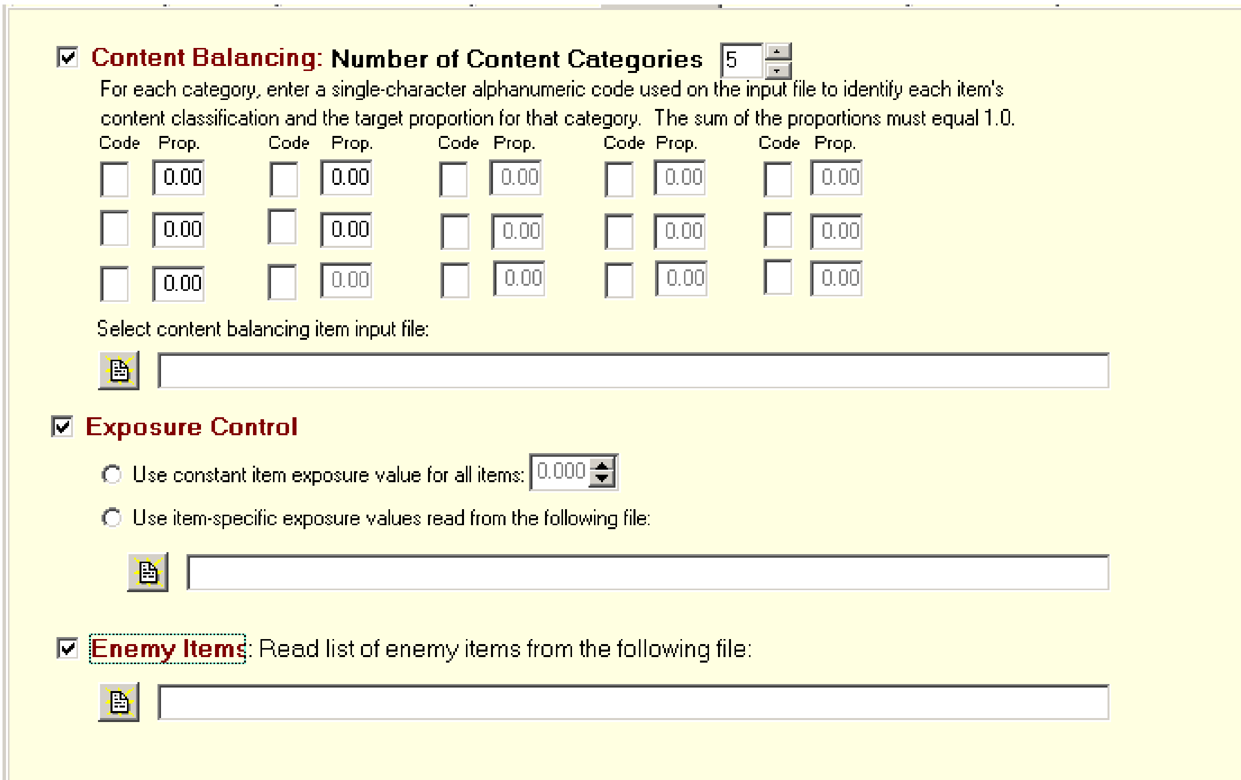
Next, the specifications for the CATs must be determined. A substantial portion of CAT research is designed to compare the results of CAT under varying conditions, such as competing item selection methods. In addition, before a CAT is implemented it frequently is important to investigate the specific set of CAT options to be implanted in the CAT. Figure 4 presents the options for the CAT specifications supported by CATSIM.

Figure 4. CATSIM CAT Options



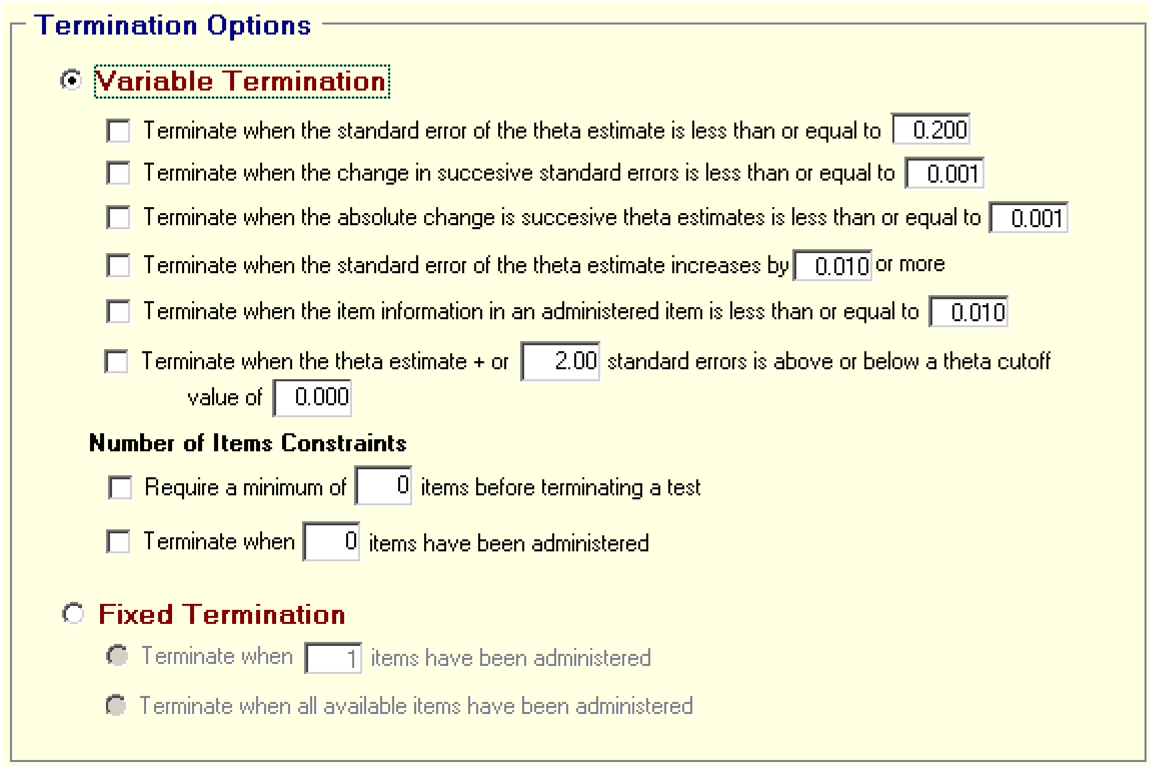
Perhaps the most complex of the CAT algorithms is the item selection algorithm. Often, this algorithm is constrained by subalgorithms designed to control various aspects of item bank usage. Content constraints refer to a test that requires a certain number of items from various content areas, such as a math test containing a specific balance of algebra and geometry. Item exposure constraints are designed to limit the percentage of examinees that see a given item. Enemy item constraints are designed to prevent items that are dependent on each other (for example, covering the same concept) from being administered to a certain examinee. Figure 5 presents the options for the item selection constraints subalgorithms supported by CATSIM.

Figure 5. CATSIM Item Selection Constraints Options



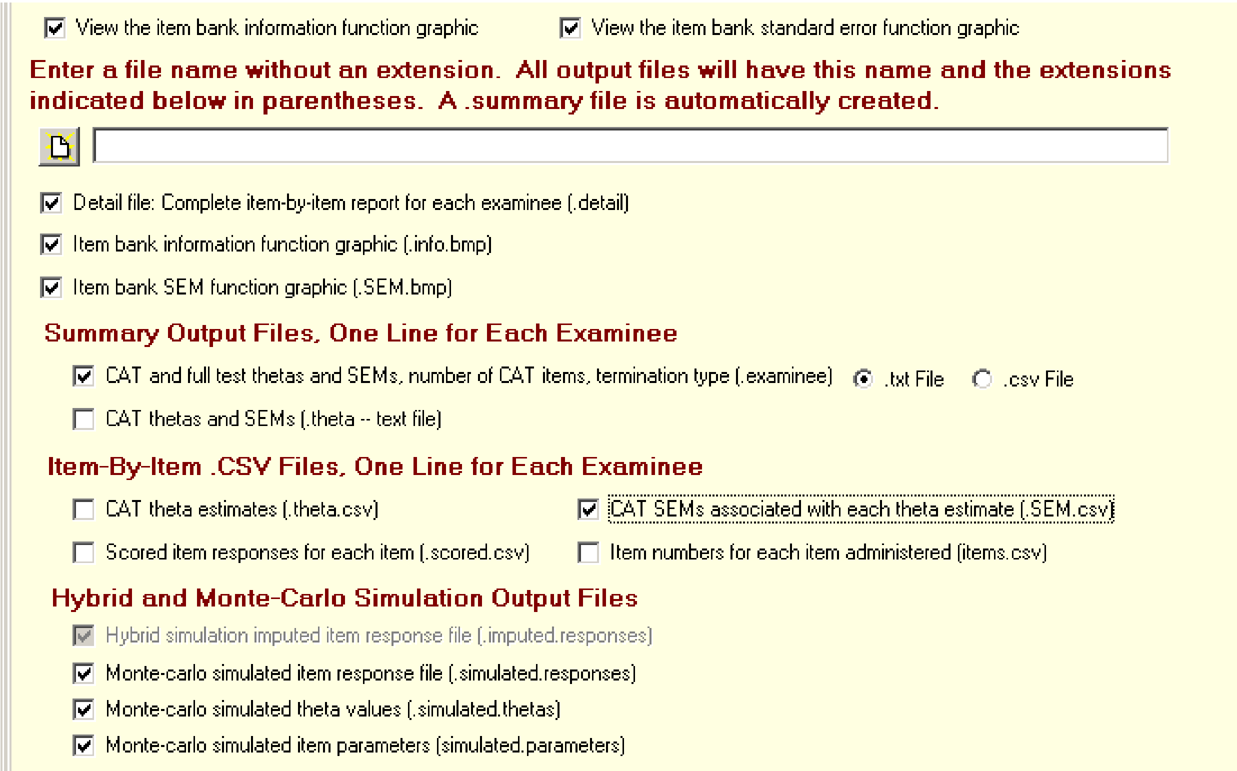
The final component in a CAT is the termination criterion. CATs can be terminated at a fixed number of items, or with variable length by evaluating variables such as convergence of the examinee θ estimate, a fixed value of the conditional standard error of measurement or its convergence, or the amount of information in the next item to be administered. These termination criteria may be used in combination or in conjunction with a minimum or maximum number of items. Figure 6 presents the options for the termination criteria supported by CATSIM.

Figure 6. CATSIM Termination Options



CATSIM provides a wide variety of output options. These include item bank information and SEM graphics, summary reports for examinees, item-by-item reports for every examinee, several types of .CSV files with item-by-item results, and both hybrid and monte-carlo data and parameter files used in those types of simulations. All use a single file name provided by the user and have a variety of file extensions that identify their contents. Figure 7 displays CATSIM's output options.

Figure 7. CATSIM Output Options

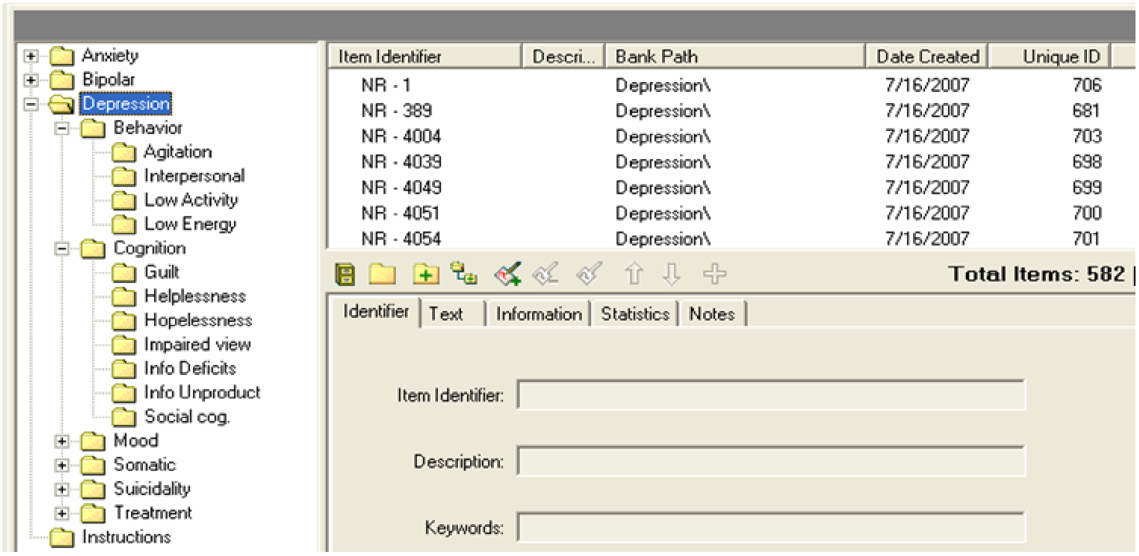


FastCAT

FastCAT is a software system for item banking, test assembly, and the delivery of CAT assessments. It is designed to have a similar set of options as CATSIM, so that the results of simulation studies can easily be implemented in live CATs. FastCAT supports CATS with both dichotomous and polytomous IRT models, and the ubiquitous multiple-choice item as well as open response and survey items. It is also able to implement multistage testing with testlets.

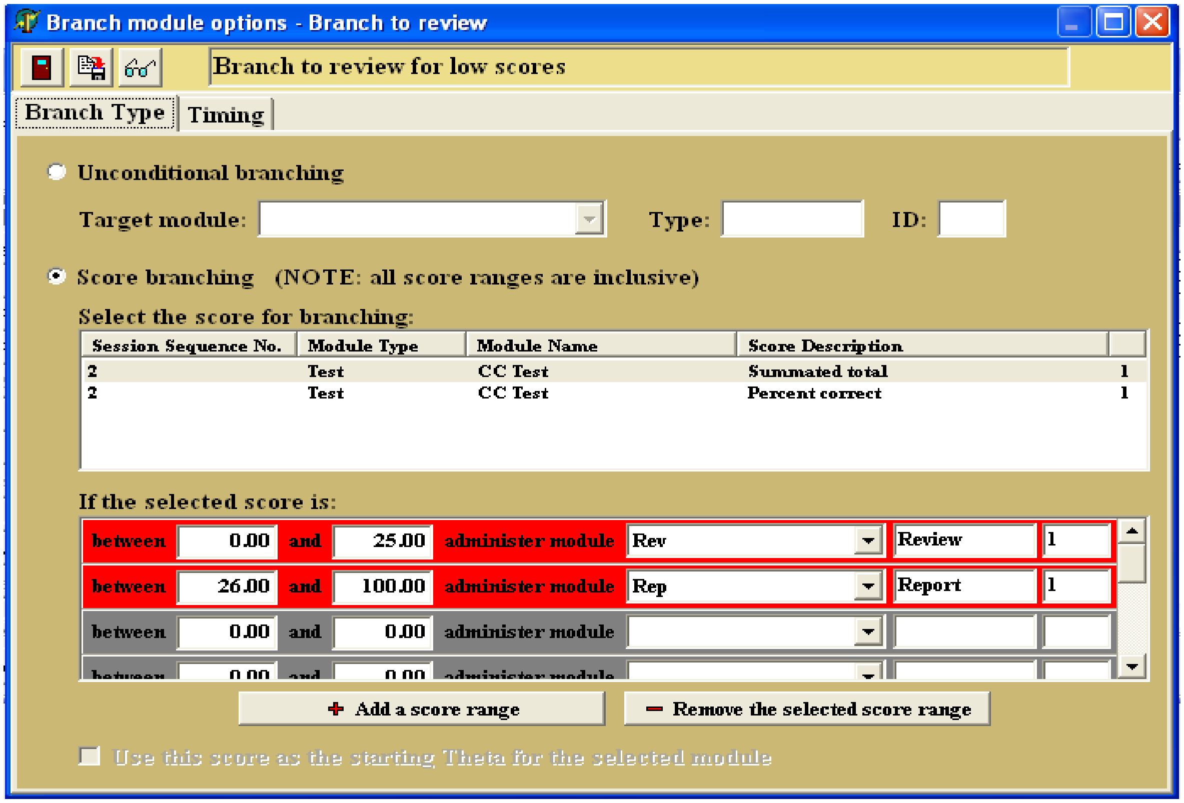
Items are stored in an item banking module that is organized with a familiar folder structure, as seen for a CAT Depression Inventory in Figure 8. The item banker stores a wide range of information, including sources, author, and rationales, in addition to item text and statistics or IRT parameters. All item parameters are easily importable from calibrations performed by IRTPRO (Thissen, 2009).

Figure 8. The CAT-DI Item Bank in FastCAT



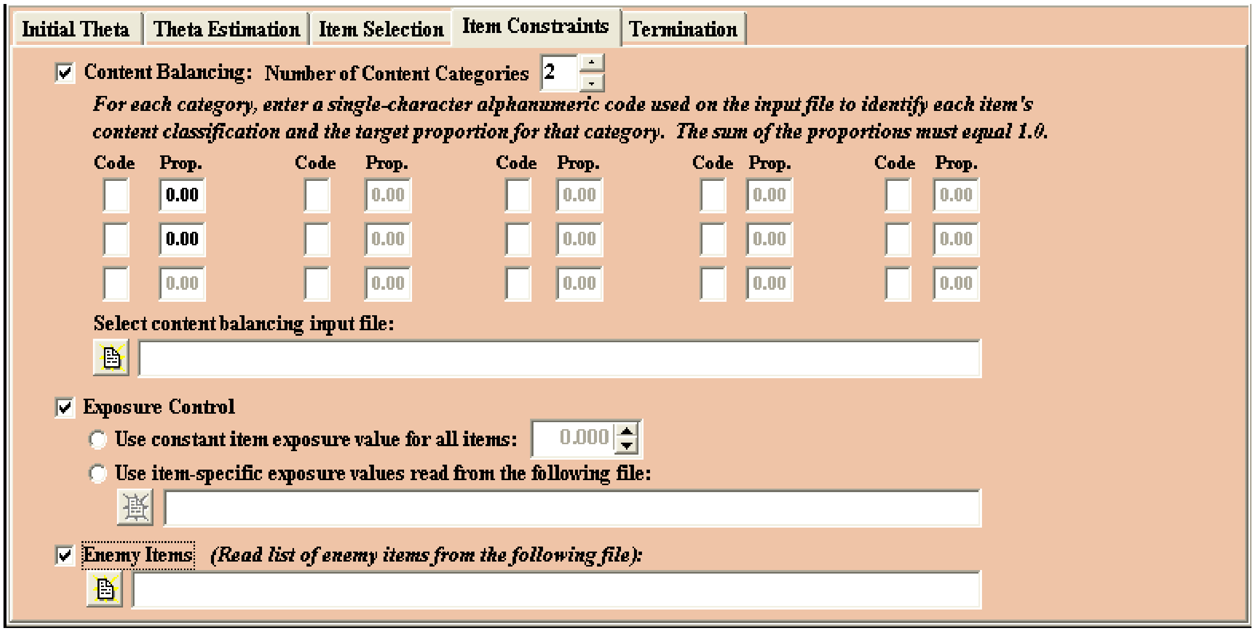
Items are then assembled into test sessions, which are modularized. This allows for multiple tests or subtests to be given easily at one sitting. Moreover, results from one test can be used as a starting point for a subsequent test. Branching can also occur between modules (Figure 9), thus allowing for multistage implementations of CAT. In addition, the modular structure of test sessions allows the use of several additional types of modules including separate (unscored) instruction modules and score conversion modules which allow conversion of the θ metric to other scales, e.g., a T score with a defined mean and standard deviation or a percentile score.

Figure 9. Module Branching in FastCAT



To facilitate use with CATSIM, the options available in FastCAT, as well as the layout of the interface, are designed to be similar to CATSIM. Therefore the results of simulation studies, which are essential for determining appropriate specifications for a published CAT, can be easily implemented. Figure 10 displays the item selection constraints in FastCAT; note the similarity to Figure 5.

Figure 10. Item Selection Constraints in FastCAT



The test delivery portion of FastCAT provides several types of reports. Reports can be displayed to the examinee at any point in the session to provide necessary feedback. Item-by-item reports with psychometric details can also be produced (Figure 11). Reports can be displayed on screen or saved to a file for later access. An additional feature is specific warning reports about certain items that contain very important content. For example, items that indicate high suicide risk can produce an automatic report for the proctor to identify the patient before they leave the clinic.

Figure 11. Item-by-Item Report in FastCAT

```
Item-By-Item Report of Maximum Information CAT for file: SESSION1.PREVIEW_42

               This test will terminate when the standard error of theta
                        is equal to or less than 0.200
            Minimum number of items = 5        Maximum number of items = 40
                   Theta was estimated by maximum likelihood.

Examinee Name : John Q. Public

The standard error band plotted as ---- is plus or minus 2.00 standard errors.
      X = Initial theta value    C = Correct answer    I = Incorrect answer

Item  Theta  SE   -3.......-2........-1.........0........+1........+2........+3
  0  -0.24* 1.00*        --------------------X--------------------
  1   4.00* 1.00*                               .         -------------------->
  2   4.00* 1.00*                               .         -------------------->
  3   2.52  0.84                                .       -----------------I-----
  4   2.77  0.68                                .             -------------C---
  5   2.38  0.61                                .          ------------I-------
  6   2.09  0.61                                .       ------------I----------
  7   1.49  0.89                             -----------------I----------------
  8   0.36  1.00               --------------------I--------------------
  9   0.88  0.63                            ------------C-------------
 10   1.13  0.56                                -----------C-----------
 11   1.34  0.49                                .  ----------C----------
 12   1.44  0.46                                .    ---------C---------
 13   1.55  0.43                                .     ---------C---------
 14   1.67  0.41                                .       --------C--------
 15   1.54  0.38                                .      --------I--------
 16   1.60  0.36                                .       --------C-------
 17   1.70  0.35                                .        -------C-------
 18   1.76  0.34                                .         -------C-------
 19   1.65  0.32                                .         ------I------
 20   1.52  0.31                                .       -------I------
 21   1.40  0.30                                .      -------I------
 22   1.27  0.30                                .     ------I------
 23   1.30  0.28                                .      ------C-----
 24   1.32  0.28                                .      ------C-----
 25   1.36  0.27                                .       -----C------
```

Item response data and item time data are recorded, and can be exported at a later time for additional analysis. This also allows FastCAT to be used during the pilot testing phase, not just the full CAT delivery phase. Item response data from either pilot or operational testing can be exported in a format ready for analysis by IRTPRO, which then produces item parameter output that is easily importable back into FastCAT.

FastCAT delivers tests from a standalone PC file, enabling maximum portability without being Web-based. Tests can be delivered from a PC hard drive, a PC network drive, a USB (thumb) drive, a CD, downloaded from the Internet, or send as an email attachment. Furthermore, data can also be consolidated via these same media. However, a Web-based version (FastTEST Web) is also under development.

Summary

The MEDPRO project entails the development of software for the major steps of CAT development and implementation, including item banking, pilot testing, IRT calibration, simulation studies, and CAT delivery. The IRTPRO software (Thissen, 2009) provides extensive functionality for the IRT calibration step. The remaining steps are implemented by the two software systems described in this paper: CATSIM for CAT simulation research, and

FastCAT for item banking, pilot testing, and CAT delivery. The development of this software enables the application of CAT in more areas, as it advantages become more well known in patient-reported outcomes research as well as in medical assessment in general.

References

Immekus, J. C., Gibbons, R.D., & Rush, J. A. (2007). Patient-reported outcomes measurement and computerized adaptive testing: An application of post-hoc simulation to a diagnostic screening instrument. In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. Retrieved July 1, 2010 from www.psych.umn.edu/psylabs/CATCentral/.

Nydick, S. W., & Weiss, D. J. (2009). A hybrid simulation procedure for the development of CATs. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved July 1, 2010 from www.psych.umn.edu/psylabs/CATCentral/.

Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system—IRT software. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved July 1, 2010 from www.psych.umn.edu/psylabs/CATCentral/.

van der Linden, W. J., & Glas, C.A.W. (Eds.). (2000). *Computerized adaptive testing: Theory and practice*. Boston, MA: Kluwer.