

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*



2009 GMAC® Conference on Computerized Adaptive Testing

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®.

Copyright © 2009 by the Authors

All rights reserved. Permission is granted for non-commercial use.

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 55114, 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

Type of Simulation

- Post-Hoc Simulation**
A post-hoc simulation requires an item response matrix in which all examinees have responded to all the items in a bank, and a file of estimated item parameters for the items.
- Monte-Carlo Simulation**
In a monte-carlo simulation, examinee item responses are generated based on a selected IRT model. Examinee theta estimates can be generated as specified distributions, fixed, or read from a file. Similarly, item parameters can be generated, fixed, or read from a file.
- Hybrid Simulation**
A hybrid simulation is implemented from a matrix of examinee responses to items with estimated parameters. The examinee data matrix is a "sparse" matrix that has significant amounts of missing data for each examinee. Using available item responses for each examinee, theta is estimated and missing responses are imputed using the selected IRT model. A post-hoc simulation is then implemented on the complete data matrix..

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

Alpha and beta control the shape of the beta distribution. Setting alpha and beta equal creates a symmetric distribution. 5 creates a normal distribution. Setting alpha and beta to 1 creates a uniform distribution. The difference between alpha and beta controls the skew. Negative skew is increased as the difference between alpha and beta becomes more positive, whereas positive skew increases as the difference becomes more negative. The sum of alpha and beta controls the flatness. The distribution becomes more peaked (narrower tails) as the sum increases; the distribution becomes flatter (larger tails) as the sum decreases.

Load Saved Monte-Carlo Simulation Defaults

Number of examinees

Number of items

		Beta Distribution Parameters				Fixed mean or fixed value		
		Shape		Range				
		Alpha	Beta	Minimum	Maximum	Fix	Mean	
Theta	<input type="radio"/> Fix <input type="radio"/> Read from file <input type="radio"/> Generate	<input type="text" value="1.0"/>	<input type="text" value="1.0"/>	<input type="text" value="-3.00"/>	<input type="text" value="3.00"/>	<input type="checkbox"/>	<input type="text" value="0.00"/>	Generate
<input type="text"/>								
Item Parameters								
Discrimination (a)	<input type="radio"/> Fix <input type="radio"/> Read from file <input type="radio"/> Generate	<input type="text" value="1.0"/>	<input type="text" value="1.0"/>	<input type="text" value="0.50"/>	<input type="text" value="1.50"/>	<input type="checkbox"/>	<input type="text" value="1.00"/>	Generate
Location (b)	<input type="radio"/> Fix <input type="radio"/> Read from file <input type="radio"/> Generate	<input type="text" value="1.0"/>	<input type="text" value="1.0"/>	<input type="text" value="-3.00"/>	<input type="text" value="3.00"/>	<input type="checkbox"/>	<input type="text" value="0.00"/>	Generate
Guessing (c)	<input checked="" type="radio"/> Fix <input type="radio"/> Read from file <input type="radio"/> Generate	<input type="text" value="1.0"/>	<input type="text" value="1.0"/>	<input type="text" value="0.00"/>	<input type="text" value="0.20"/>	<input checked="" type="checkbox"/>	<input type="text" value="0.20"/>	Generate
Item parameter file: Boundary locations must be read from a file for all polytomous models								
<input type="text"/>								
Save Monte-Carlo Simulation Defaults								

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

IRT Model

- Dichotomously Scored Items**
 - 1-Parameter (b)
 - 2-Parameter (a, b)
 - 3-Parameter (a, b, c)
- Polytomously Scored Items**
 - Rating Scale (Likert-Type) Items**
 - SGRM: Samejima's graded response model (discrimination and boundary locations for each item)
 - GRSM: Generalized rating scale model (common boundary locations; discrimination and location for each item)
 - RRSM: Rasch rating scale model (common boundary locations; location for each item)
 - Ability/Achievement Test Items**
 - GPCM: Generalized partial credit model (discrimination and step/boundary locations for each item)
 - RPCM: Rasch partial credit model (common discrimination, step/boundary locations for each item)
 - Response weights (e.g., 1, 2, 3, 4, 4) are included in the parameter file for combining categories

Polytomous Parameter Scaling

- Signs for the boundary parameters have not been reversed from the Parscale output (i.e., are positive to negative)
- Signs for the boundary parameters have been reversed from the Parscale output (i.e., are negative to positive)

Model Constant

- D = 1.0 (pure logistic model)
- D = 1.7 (logistic approximation to the normal ogive)

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

Initial Theta

- Initial theta for all examinees is
- Initial theta is random in the interval to
- Read initial thetas and SEMs for each examinee from a file



Use individual SEMs as Bayesian prior SDs when Bayesian scoring is selected

Estimate Theta By

- Maximum likelihood: Estimate theta for initial items for non-mixed response patterns using
 - A step size of theta/difficulty =
 - Bayesian estimation with a mean of and a standard deviation of
- Bayesian estimation with a mean of and a standard deviation of
For Bayesian estimation, estimate theta by EAP (expected a posteriori) MAP (maximum a posteriori)
- Weighted maximum likelihood

Item Selection Option

- Select all items by maximum information at the current theta estimate
- Randomly from the items with maximum item information for the first items in the test
- Select items in order of maximum information at theta =

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

Content Balancing: Number of Content Categories

For each category, enter a single-character alphanumeric code used on the input file to identify each item's content classification and the target proportion for that category. The sum of the proportions must equal 1.0.

| Code | Prop. |
|----------------------|-----------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|----------------------|-----------------------------------|
| <input type="text"/> | <input type="text" value="0.00"/> |
| <input type="text"/> | <input type="text" value="0.00"/> |
| <input type="text"/> | <input type="text" value="0.00"/> |

Select content balancing item input file:

Exposure Control

Use constant item exposure value for all items:

Use item-specific exposure values read from the following file:

Enemy Items: Read list of enemy items from the following file:

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

Termination Options

Variable Termination

- Terminate when the standard error of the theta estimate is less than or equal to
- Terminate when the change in successive standard errors is less than or equal to
- Terminate when the absolute change in successive theta estimates is less than or equal to
- Terminate when the standard error of the theta estimate increases by or more
- Terminate when the item information in an administered item is less than or equal to
- Terminate when the theta estimate + or standard errors is above or below a theta cutoff value of

Number of Items Constraints

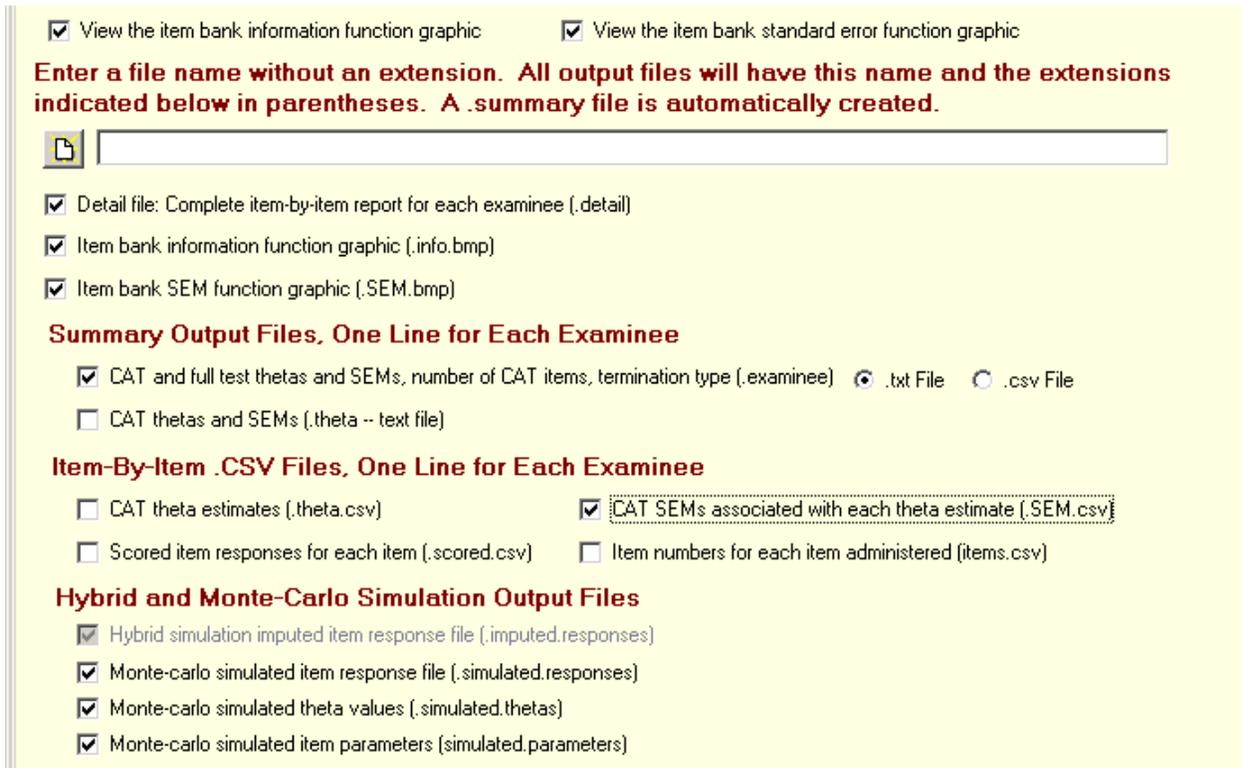
- Require a minimum of items before terminating a test
- Terminate when items have been administered

Fixed Termination

- Terminate when items have been administered
- Terminate when all available items have been administered

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



View the item bank information function graphic View the item bank standard error function graphic

Enter a file name without an extension. All output files will have this name and the extensions indicated below in parentheses. A .summary file is automatically created.

Detail file: Complete item-by-item report for each examinee (.detail)
 Item bank information function graphic (.info.bmp)
 Item bank SEM function graphic (.SEM.bmp)

Summary Output Files, One Line for Each Examinee

CAT and full test thetas and SEMs, number of CAT items, termination type (.examinee) .txt File .csv File
 CAT thetas and SEMs (.theta -- text file)

Item-By-Item .CSV Files, One Line for Each Examinee

CAT theta estimates (.theta.csv) CAT SEMs associated with each theta estimate (.SEM.csv)
 Scored item responses for each item (.scored.csv) Item numbers for each item administered (.items.csv)

Hybrid and Monte-Carlo Simulation Output Files

Hybrid simulation imputed item response file (.imputed.responses)
 Monte-carlo simulated item response file (.simulated.responses)
 Monte-carlo simulated theta values (.simulated.thetas)
 Monte-carlo simulated item parameters (.simulated.parameters)

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

Item Identifier	Descri...	Bank Path	Date Created	Unique ID
NR - 1		Depression\	7/16/2007	706
NR - 389		Depression\	7/16/2007	681
NR - 4004		Depression\	7/16/2007	703
NR - 4039		Depression\	7/16/2007	698
NR - 4049		Depression\	7/16/2007	699
NR - 4051		Depression\	7/16/2007	700
NR - 4054		Depression\	7/16/2007	701

Total Items: 582

Identifier | Text | Information | Statistics | Notes

Item Identifier:

Description:

Keywords:

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

Branch module options - Branch to review

Branch to review for low scores

Branch Type | Timing

Unconditional branching

Target module: Type: ID:

Score branching (NOTE: all score ranges are inclusive)

Select the score for branching:

Session Sequence No.	Module Type	Module Name	Score Description	
2	Test	CC Test	Summated total	1
2	Test	CC Test	Percent correct	1

If the selected score is:

between	0.00	and	25.00	administer module	Rev	Review	1
between	26.00	and	100.00	administer module	Rep	Report	1
between	0.00	and	0.00	administer module			
between	0.00	and	0.00	administer module			

+ Add a score range - Remove the selected score range

Use this score as the starting Theta for the selected module

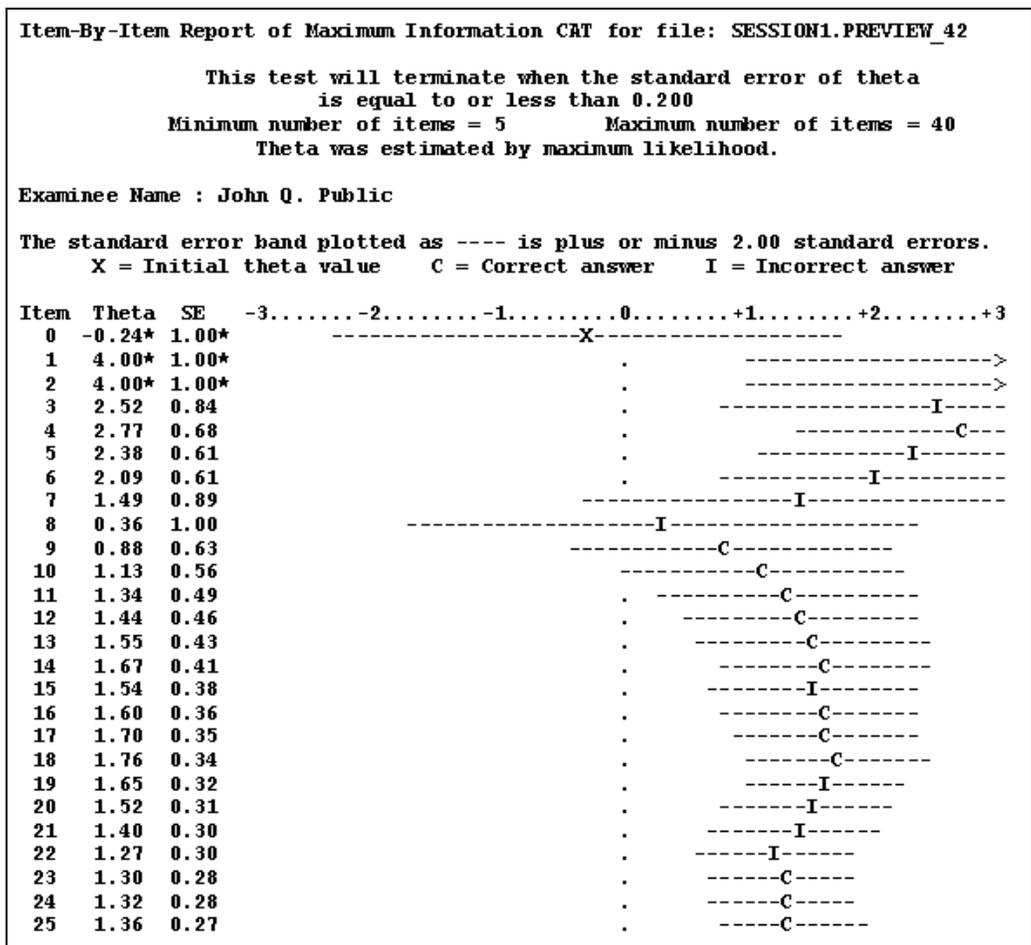
To facilitate use with CATSIM, the options available in FastCAT, as well as the layout of the interface, are designed to be similar to CATSIM. Therefore the results of simulation studies, which are essential for determining appropriate specifications for a published CAT, can be easily implemented. Figure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

Figure 10. Item Selection Constraints in FastCAT

Initial Theta	Theta Estimation	Item Selection	Item Constraints	Termination
<input checked="" type="checkbox"/> Content Balancing: Number of Content Categories <input type="text" value="2"/>				
<i>For each category, enter a single-character alphanumeric code used on the input file to identify each item's content classification and the target proportion for that category. The sum of the proportions must equal 1.0.</i>				
Code	Prop.	Code	Prop.	Code
<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>
<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>
<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>	<input type="text" value="0.00"/>	<input type="text"/>
Select content balancing input file:				
<input type="text"/>				
<input checked="" type="checkbox"/> Exposure Control				
<input type="radio"/> Use constant item exposure value for all items: <input type="text" value="0.000"/>				
<input type="radio"/> Use item-specific exposure values read from the following file:				
<input type="text"/>				
<input checked="" type="checkbox"/> Enemy Items: <i>(Read list of enemy items from the following file):</i>				
<input type="text"/>				

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 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 its 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.