

2011

User's Manual: *SimulCAT*

Kyung (Chris) T. Han



Header: User's Manual: *SimulCAT*

User's Manual for *SimulCAT*:
Windows Software for Simulating Computerized Adaptive Test Administration

Kyung T. Han
Graduate Management Admission Council®

The views and opinions expressed in this article are those of the author and do not necessarily reflect those of the Graduate Management Admission Council®.

I. Introduction

Most, if not all, computerized adaptive testing (CAT) programs use simulation techniques to develop and evaluate CAT program administration and operations, but such simulation tools are rarely available to the public. Up to now, several software tools have been available to conduct CAT simulations for research purposes; however, these existing tools, for the most part, oversimplify the CAT algorithms and are not powerful enough to simulate operational CAT environments. *SimulCAT*, a new CAT simulation software tool, was developed to serve various purposes ranging from fundamental CAT research to technical CAT program evaluations. The new CAT simulation software tool offers many advantages, including a wide range of item selection algorithms, adaptability for a variety of CAT administration environments, and a user-friendly graphical interface, among others, as described in the following commentary.

SimulCAT offers a variety of item selection algorithms.

The most widely used item selection algorithms—including maximized Fisher information (MFI; Weiss, 1982), a -stratification (Chang & Ying, 1999; Chang, Qian, & Ying, 2001), global information (Chang & Ying, 1996), interval information, likelihood weighted information (Veerkamp & Berger, 1997), gradual maximum information ratio (GMIR; Han, 2009), and efficiency balanced information (EBI; Han, 2010)—are available within. Along with a choice of item selection criteria, a variety of item exposure control options are available within *SimulCAT*, including, the randomesque strategy (Kingsbury & Zara, 1989), Sympton and Hetter method (1985), the multinomial methods—both conditional and unconditional (Stocking & Lewis, 1995, 1998)—and the fade-away method (FAM; Han, 2009). For content balancing, *SimulCAT* supports the content script method and the constrained CAT method (Kingsbury & Zara, 1989).

SimulCAT simulates various CAT environments.

To create CAT environments that are as realistic as possible, *SimulCAT* supports various CAT administration options. First, the interim and final score estimates can be accomplished using the maximum likelihood (ML), Bayesian maximum a posteriori (MAP), Bayes expected a posteriori (EAP) estimations, or any combination of those. Software users also can set the initial score value, range of score estimates, and restriction in estimate change. The length of CAT administration can be either fixed or variable; for variable-length testing, *SimulCAT* supports multiple termination rules including standard error of estimation and score estimate consistency. Within *SimulCAT*, the number of test takers who are administered simultaneously at each test time slot and the frequency of communication between a test server and client computers (i.e., terminals) can also be conditioned according to the user's choice.

SimulCAT provides powerful research tools.

SimulCAT can read user-specified existing data and can generate new data sets as well. Score distribution can be drawn from a normal, uniform, or beta distribution, and item parameters for an item pool can be generated from normal, uniform, and/or lognormal distributions. The *SimulCAT* tool also offers several graphical analysis tools such as distribution density functions,

item response functions, and information functions at both item and pool levels. *SimulCAT* can generate reports on item pool usage and CAT administrations. For more advanced research, *SimulCAT* provides users with options to input differential item functioning (DIF) or item parameter drift (IPD) information as well as preexisting item exposure data. The software tool also supports the use of syntax files and a cue file for massive simulation studies.

SimulCAT has an intuitive graphical user interface.

As a Windows-based application, *SimulCAT* provides a user-friendly graphical interface. Most features of *SimulCAT* can be accessed by just a few simple point-and-clicks. The main interface of *SimulCAT* consists of easy-to-follow three steps: Examinee/Item Data, Item Selection, and Test Administration.

System Requirements, Availability, and Distribution

SimulCAT runs on a Microsoft Windows-based operating system with .NET framework 2.0 or above. Microsoft's Windows Vista and Windows 7 include .NET framework, but a machine running an older version of the Windows OS will need to have .NET framework installed first. The software package, a copy of the manual in PDF format, and example files can be found and downloaded at the following web site: <http://www.hantest.net>. The software package is free of charge and may be distributed to others without the author's permission for noncommercial uses only. *SimulCAT* always checks for the latest version and automatically updates itself as long as it is running on a machine with an active Internet connection.

II. Item Selection Algorithms Used Within *SimulCAT*

SimulCAT supports nine item selection algorithm/criteria as follows:

- 1) Maximum Fisher Information (MFI)
- 2) a -Stratification (with/without b -blocking)
- 3) Best matching b -value
- 4) Randomization
- 5) Interval Information Criterion
- 6) Likelihood Information Criterion
- 7) Kullbak-Leibler Information (Global Information)
- 8) Gradual Maximum Information Ratio (GMIR)
- 9) Efficiency Balanced Information (EBI)

1. Maximum Fisher Information

One of the most widely used—and probably the oldest—item selection methods in computerized adaptive testing (CAT) involves selecting an item with the *maximized Fisher information* (MFI) at the interim proficiency estimate based on test items previously administered to the examinee. Basically this involves finding item x maximizing $I[\hat{\theta}_{m-1}]$ for an examinee with the interim proficiency estimate $\hat{\theta}$ and $m-1$ as the number of items administered so far (Weiss, 1982). Taking a typical case of a multiple-choice item pool, where item characteristics are defined by the three-parameter logistic model, or 3PLM (Birnbaum, 1968), the item selection method based on the MFI criterion looks for item i that results in the largest value of

$$I_i[\hat{\theta}_{m-1}] = \frac{(Da_i)^2(1-c_i)}{[c_i + e^{Da_i(\hat{\theta}_{m-1}-b_i)}][1 + e^{-Da_i(\hat{\theta}_{m-1}-b_i)}]^2}, \quad (1)$$

where a_i , b_i , and c_i are the discrimination, difficulty, and pseudo-guessing parameters in 3PLM, respectively, and D is a scaling constant whose value is 1.702. The MFI criterion is popular because it is an effective means of administering CAT that results in maximized test information for each individual. In the early stage of testing, however, when five or fewer items have been administered, for example, the interim proficiency estimates are rarely accurate. So as testing begins, items selected according to the MFI criterion tend not to provide as much information as

they were designed to do at interim proficiency estimates. Another issue with the MFI method is that it tends to select items with higher a -parameter values more frequently than it selects items with lower a -parameter values. This uneven item usage with the MFI method may create serious problems in item pool maintenance.

2. a -Stratification With/Without b -Blocking

To prevent wasting items with higher a -parameter values at the outset of CAT administration, Chang and Ying (1999) proposed stratifying items in the item pool by a -parameter values. In their method, known as *a-stratification*, items with b -parameter values closest to the interim $\hat{\theta}$ are selected from the item stratum with the lowest a -parameter values at CAT's outset. Using this method, items with higher a -parameter values are selected from the item strata as CAT administration proceeds.

A potential problem with this method is an observed correlational relationship between a - and b -parameter values, such that when items are stratified according to a -parameter values the item strata may not be equivalent to each other in terms of b -parameters. Chang, Qian, and Ying (2001) addressed this problem with their modification known as *a-stratification with b-blocking*.

3. Best Matching b -Value

The *best matching b-value* criterion, essentially is a special application of the *a-stratification* method that always uses only one item stratum. With this method, an item with a b -value closest to the interim theta estimate is selected regardless a - and c -parameter values.

4. Randomization

With this item selection method, an item is picked at random; thus the test is not adaptive.

5. Interval Information Criterion

Veerkamp and Berger (1997) designed two alternatives to MFI. In the first, called the *interval information criterion* (IIC), the information function is averaged across the confidence interval of an interim proficiency estimate. The mathematic expression of IIC for item i is

$$\int_{\hat{\theta}_L}^{\hat{\theta}_R} I_i[\theta] d\theta \quad (2)$$

where $\hat{\theta}_L$ and $\hat{\theta}_R$ are a confidence interval of θ . The actual mean value for IIC is Equation 2 divided by the length of the confidence interval; however taking an average of it is unnecessary for the purpose of item selection.

6. Likelihood Weight Information Criterion

In the second alternative approach proposed by Veerkamp and Berger (1997), the *likelihood weighted information criterion* (LWI), the information function is summed throughout the theta scale, weighted by the likelihood function after item administrations performed thus far. With the LWI criterion, the item to be selected is item i that results in the maximized value of

$$\int_{-\infty}^{\infty} L(\theta; x_{m-1}) I_i[\theta] d\theta \quad (3)$$

where $L(\theta; x_{m-1})$ is the likelihood function of the response vector x_{m-1} after $(m-1)$ th item administration.

7. Kullbak-Leibler Information (Global Information)

Chang and Ying (1996) came up with the *global information* approach that uses the moving average of *Kullback-Leibler information* (KLI) to select items (Cover & Thomas, 1991; Kullback, 1959). The KLI for any θ for the i th item with response X_i is defined by

$$K_i(\theta \| \theta_0) = P_i(\theta_0) \log \left[\frac{P_i(\theta_0)}{P_i(\theta)} \right] + [1 - P_i(\theta_0)] \log \left[\frac{1 - P_i(\theta_0)}{1 - P_i(\theta)} \right], \quad (4)$$

where $P_i(\theta_0)$ is the probability that a random examinee at the proficiency level θ_0 answers the item correctly. The moving average of KLI is then calculated and used as the item selection criterion, as follows,

$$K_i(\theta_0) = \int_{\theta_0 - \delta}^{\theta_0 + \delta} K_i(\theta \| \theta_0) d\theta, \quad (5)$$

where δ specifies the range of the moving average. Determining δ could yield ambiguous results, but Chang and Ying (1996) proposed c / \sqrt{m} as a reasonable choice for δ with constant c selected according to a specified coverage probability and with m being the number of items administered thus far. Chang and Ying (1996) found that replacing the MFI criterion with the KLI criterion

often reduced the biases and mean-squared errors of proficiency estimation when the test length was short or the CAT administration was in its early stage ($m < 30$).

8. Gradual Maximum Information Ratio (GMIR)

To promote use of items with low a -parameter values, Han (2009) proposed selecting items at the beginning of CAT administration based on expected item efficiency instead of the MFI criterion. Expected item efficiency is defined as the level of realization of an item's potential information at interim $\hat{\theta}$. Thus, if item i results in its maximum potential information at θ_i^* , the expected item efficiency at interim $\hat{\theta}$ is computed by

$$\frac{I_i[\hat{\theta}_{m-1}]}{I_i[\theta_i^*]}, \quad (6)$$

where θ_i^* is equal to b_i when either a 1PL or 2PL model is used. If a 3PL model is used and $c_i \neq 0$, θ_i^* can be computed using Birnbaum's solution (1968):

$$\theta_i^* = b_i + \frac{1}{Da_i} \log\left(\frac{1 + \sqrt{1 + 8c_i}}{2}\right). \quad (7)$$

Han (2009) suggested taking item effectiveness (i.e., expected item information) into account over the item efficiency as CAT administration proceeds and reaches its conclusion. This approach looks for an item that maximizes the criterion,

$$\frac{I_i[\hat{\theta}_{m-1}]}{I_i[\theta_i^*]} \left(1 - \frac{m}{M}\right) + I_i[\hat{\theta}_{m-1}] \frac{m}{M}, \quad (8)$$

where M is the test length, and m is 1 plus the number of items administered thus far. This method is referred to as the *gradual maximum information ratio* (GMIR) approach. The first part of Equation 8 is the item efficiency term; the second part is the item effectiveness term (the Fisher information from Equation 1). Each part of Equation 8 is inversely weighted by the progress of the CAT administration. Based on his simulation results, Han (2009) found that the GMIR approach could generally improve the item pool utilization compared to the MFI criterion.

9. Efficiency Balanced Information (EBI)

Han (2010) proposed the *efficiency balanced information* (EBI) criterion for item selection.

Unlike the GMIR approach (Han, 2009), the item efficiency (Equation 6) and information are not evaluated at a certain point of θ but assessed across the θ interval. The width of the θ interval for evaluation of item efficiency and information is determined by standard errors of estimation (SEE; ε) and set to 2SEE from $\hat{\theta}$ after j -th item administration ($\hat{\theta}_j + 2\varepsilon_j$).

$$EBI_i[\hat{\theta}_j] = \left(1 + \frac{1}{I_i[\theta_i^*]}\right) \int_{\hat{\theta}_j - 2\varepsilon_j}^{\hat{\theta}_j + 2\varepsilon_j} I_i[\theta] d\theta, \quad (9)$$

With this method, items with lower a -values are more likely to be selected at the beginning of CAT whereas items with higher a -values occur more frequently in CAT's later stages.

III. Item Exposure Controls Used Within *SimulCAT*

The SimulCAT software tool for CAT simulation offers five item exposure control methods: (1) randomesque, (2) Sympton-Hetter, (3) unconditional multinomial, (4) conditional multinomial, and (5) the fade-away method.

1. Randomesque

Kingsbury and Zara (1989) proposed the *randomesque* method to keep the best item from being solely (or excessively) used in CAT administration. In this method, a certain number of the best items are selected, one of which is administered randomly. The randomesque method may not be highly effective in limiting the overall item exposure rate to a target rate but it can prevent the same item from being used repeatedly for test takers with similar proficiency.

2. Sympton-Hetter Method

In the probabilistic approach developed by Sympton and Hetter (1985), the probability $P(A)$ that an item will be administered is differentiated from the probability $P(S)$ that the item will be selected by the item selection algorithm. In other words, the Sympton-Hetter method introduces the conditional probability $P(A|S)$ that the selected item actually will be administered. In order to keep the $P(A)$ at a desirable target level, the $P(A|S)$ that results in the target $P(A)$ is derived from iterative simulations. Once the $P(A|S)$ is computed for each item in the item pool, it is treated as the exposure parameter in the actual item selection process. During CAT administration, all eligible items are ordered by a choice of item selection criterion. Starting from the best item, the item exposure parameter is compared to a randomly generated value between 0 and 1 (following a uniform distribution). If the random value is smaller than the exposure parameter, the item is administered; otherwise, the process proceeds to the next best item. This process is repeated until one item is finally administered. It is important to note that the computed exposure parameters are pool-specific—in other words, the exposure parameters should be recomputed whenever there is a change in the item pool.

3. Unconditional Multinomial Method

The unconditional multinomial (UM) method (Stocking & Lewis, 1995) is similar to the

Sympson-Hetter (SH) method in computing exposure parameters using iterative simulations. The major difference between the UM method and the SH method is that the UM method forms a multinomial distribution from each item's $P(A|S)$ first and then compares the distribution to a random value to determine which item to actually administer.

4. Conditional Multinomial Method

The SH and UM methods are useful for controlling overall exposure for each item but they don't guarantee a desired exposure rate within each group of test takers of similar proficiency. In the conditional multinomial (CM) method (Stocking & Lewis, 1998), each item has multiple exposure parameters corresponding to each proficiency group. The exposure parameters are computed separately for each proficiency group during the simulations. Once the exposure parameters are computed, the exposure parameter for the corresponding proficiency group based on the (interim) theta estimate is used to form a multinomial distribution. The rest of the procedure is the same as the UM method.

5. Fade-Away Method

Today's computer networking technology makes it possible for main computer servers and client computers (i.e., test terminals) in test centers to communicate before, during, and/or after CAT administration to reconfigure a variety of test information including item use. Complete item usage information maintained in the main server can be updated regularly by the client computers during or after each CAT administration via the online network. In addition, each client computer can access updated item usage information from the server just before the start of the next test administration. Such network technology enables the CAT system to use near real-time item exposure information for the exposure control, precluding the need to predict the item exposure by other means, for example, using the Sympson-Hetter method (Sympson & Hetter, 1985), which involves iterative simulations.

In the new item exposure control method, the item selection criterion value for each eligible item in the pool is inversely weighted by the ratio between the updated actual exposure rate and the target exposure rate. For example, with the MFI criterion displayed in Equation 6, CAT looked for an item that maximized

$$I_i[\hat{\theta}_{m-1}] \frac{U_i}{C}, \quad (6)$$

where C was the absolute item usage limit (of the first exposure control component), which was 3,000 in this study. U_i was the item usage for the life of item i . With this new method, rarely used items are expected to be promoted more frequently, and excessively used items are likely to “fade away” from the item selection. This method will be referred to hereafter as the *fade-away* (FAM) method.

IV. Content Balancing Methods Used Within *SimulCAT*

SimulCAT employs two content-balancing methods—the script method and weight method.

1. Script Method

In the script method, test content is controlled by a script that specifies the content area based on test administration progress. The program randomly selects one script among many available scripts to prevent test takers from predicting the sequence of content areas (Note: The current version of *SimulCAT* supports only one script). When the script is shorter than the actual test length, it will restart from the top after the last content area in the script is administered.

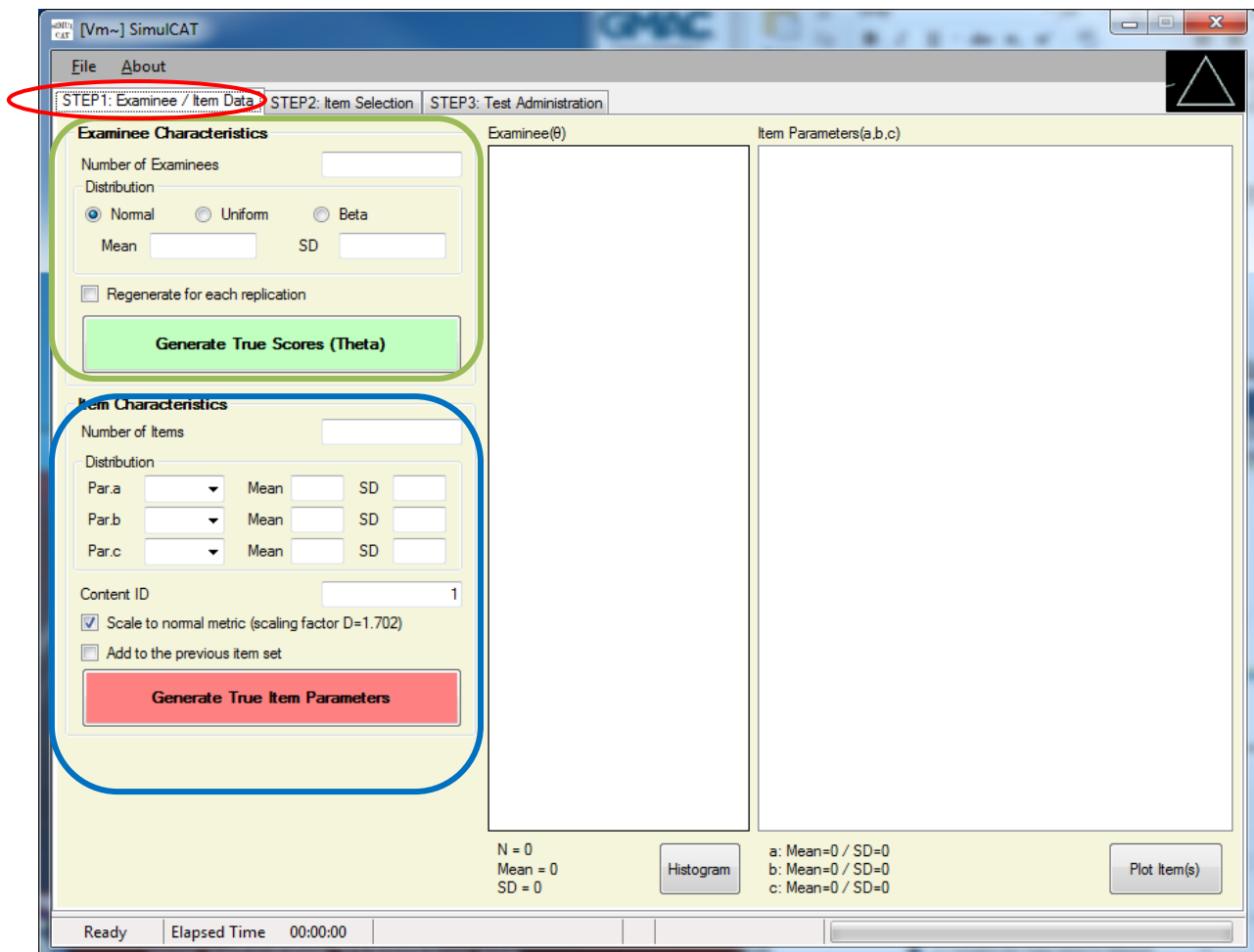
2. By Weight

Kingsbury and Zara (1989) proposed the constrained CAT (CCAT) method to balance content areas. In CCAT, the content area from which an item will be selected for administration is determined by the difference between the target weight and actual percentage of each content area thus far administered; thus the system selects the content area with a percentage farthest from the target weight.

V. Using *SimulCAT* With Graphical User Interface (GUI)

This section of the manual provides step-by-step instructions for setting up and simulating various CAT administration options, with illustrations of the actual graphical interface used in the SimulCAT program. Step 1 explains how to generate examinee and item pool data. Step 2 details how to specify an item selection algorithm by setting item selection criteria, item exposure control, test length, and content balancing. Step 3 includes instructions for specifying CAT administration rules regarding score estimation, test administration and pre-test administration, and allows the user to specify extra features and output formats before running a CAT simulation. It also contains information on running example scenarios.

Step 1. Generating Examinee and Item Pool Data



A. Examinee Characteristics (Green Box)

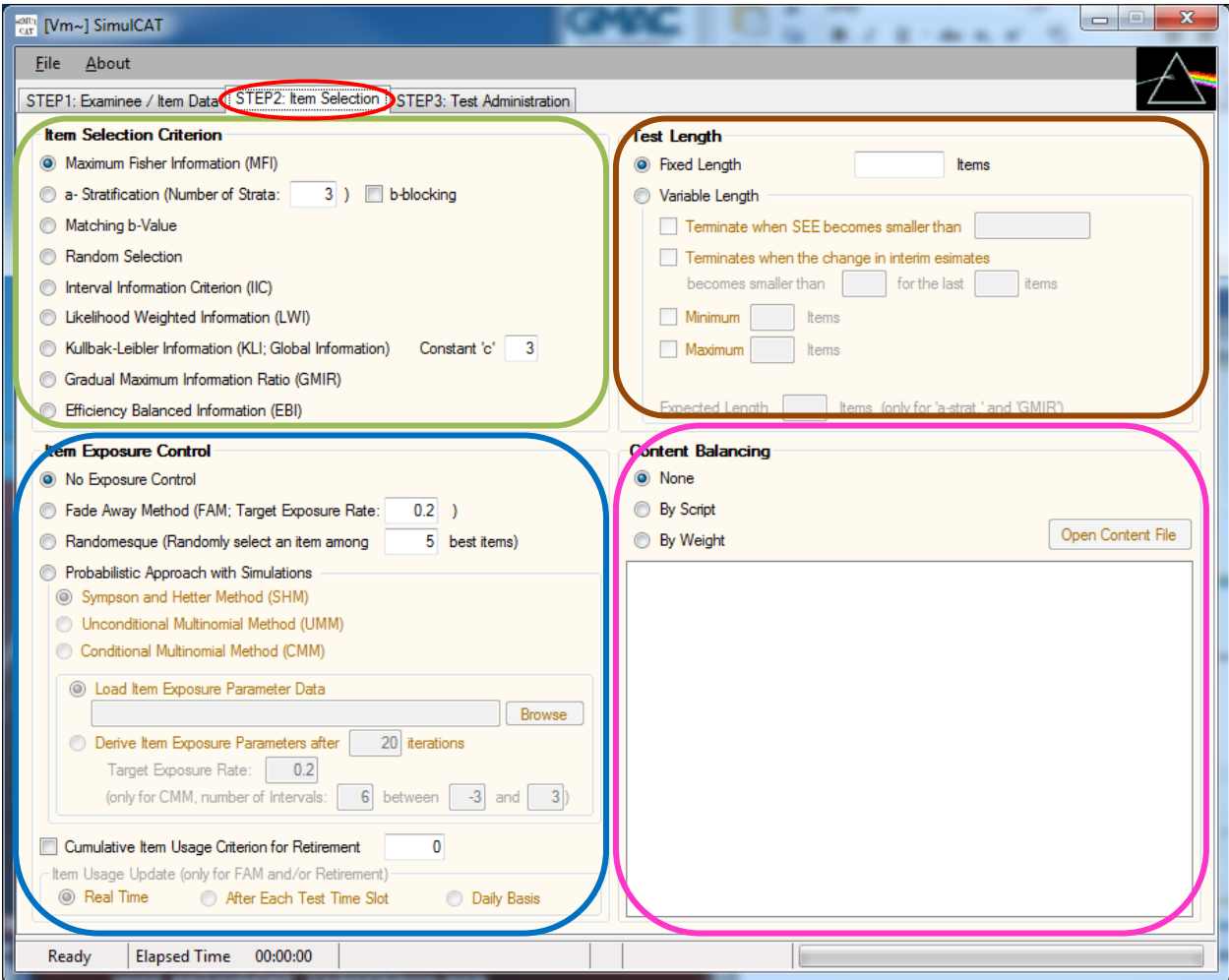
1. Specify the number of examinees.
2. Select type of score distribution .
3. Specify mean and standard deviation for a normal score distribution or, specify minimum and maximum value for a uniform score distribution or, specify a and b parameters for a beta score distribution.
4. Click on the green 'Generate True Scores' button.
5. Generated examinee theta scores should display in the box. The data set can be saved at 'File > Save > Examinee.'
6. Distribution of examinee thetas can be shown by clicking on the 'Histogram' button.

B. Item Characteristics (Blue Box)

1. Specify the number of items.
2. Select distribution of item parameters and specify properties of the distributions.
3. Specify the content ID (area code) for the items being generated.
4. Click on the red 'Generate' button.
5. Generated item parameter data should then display in the box. The data set can be saved at 'File > Save > Item.'
6. Item characteristic curves (ICCs), the item pool characteristic curve, item information function curves (IIF), and the pool information function curve (PIF) can be displayed by clicking on the 'Plot Item(s)' button.
7. Check the box labeled, 'Add to the previous item set', and repeat steps 1 through 4 if you need to add another set of items (or items with different content IDs) to a previous set of items. This option is particularly useful when simulating an item pool with multiple content areas.

[Note: The current version of *SimulCAT* only supports the three parameter logistic model (3PLM) .]

Step 2. Specifying Item Selection Algorithm



A. Item Selection Criterion (Green Box)

1. Select one of the nine item selection criterion (methods) listed in the box. (For detailed descriptions of each item selection criterion, see Chapter II.)

B. Item Exposure Control (Blue Box)

1. Select the level of item exposure control for your CAT simulation. You can choose to have no exposure control, or you can select one of several options for exposure control. (For detailed information about each item exposure control method, see Chapter III.)
2. (Optional) Check “Cumulative Item Usage Criterion for Retirement,” if you want items to retire when they reach a specified cumulative usage level. Once the items retire, they no longer will be considered in item selection.

3. If you select the Fade-Away method and/or choose the item retirement option, you must specify the “Item Usage Update” frequency. The item usage update frequency is the frequency of the reconsolidation between the test server and test terminals. (For more information, see Chapter III, Section 5.)

C. Test Length (Brown Box)

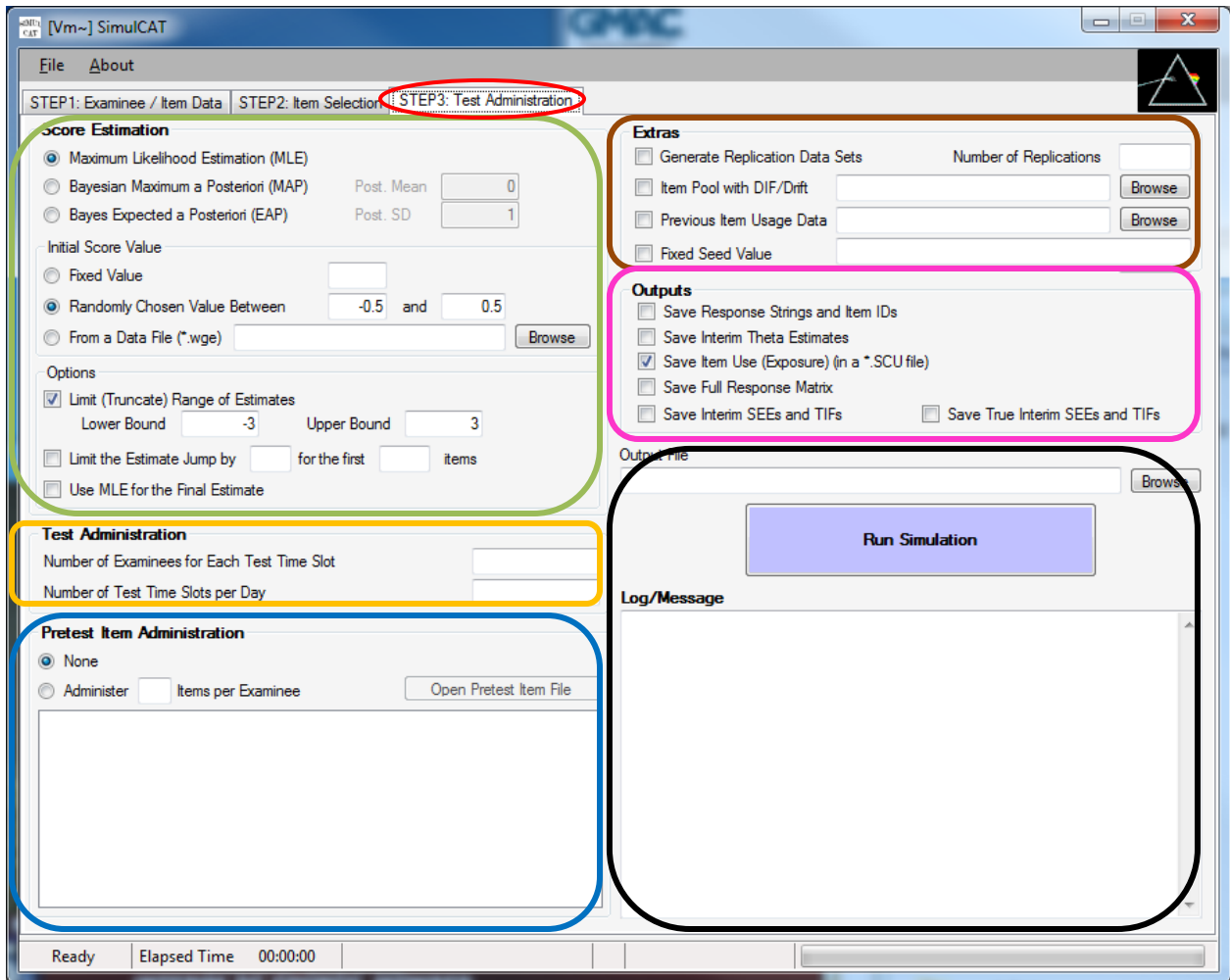
1. If “Fixed Length” is selected, specify the number of items for each individual.
2. If “Variable Length” is selected, specify the CAT termination rules.
 - a) The termination rules are combined by an “or” operation, which means the CAT administration is terminated when any of the CAT termination rules are satisfied.
 - b) Unlike the other three rules, the “minimum” rule will not terminate CAT until the test length reaches the specified minimum length even if one or more other termination rules were satisfied.
 - c) If the a-Stratification or GMIR methods were chosen for item selection, the expected test length must be specified. Both of these item selection methods compute test progress based on expected test length.

D. Content Balancing (Pink Box)

1. Select your preferred content balancing format, either “By Script” or “By Weight,” or select None. The input file formats for Script and Weight differ. See Chapter VI, Section 2, for detailed information about each content balancing method.

[IMPORTANT NOTE: The content area (content ID) is determined before the item selection criterion and exposure control are selected.]

Step 3. Specifying CAT Administration Rules



A. Score Estimation (Green Box)

1. Select MLE, MAP, or EAP for estimating interim and final scores.
 - a. Specify posterior mean and SD values if you selected MAP or EAP. (The default values for posterior mean and SD are 0 and 1, respectively).
2. Specify the initial score value. The initial score value can be fixed, randomly drawn, or loaded with a preexisting data (*.wge file).
 - a. The default setting randomly draws a value from a uniform distribution (-0.5, 0.5).
3. (Optional) You can opt to specify the range of score estimates. Estimates that are out of the specified range will be truncated.

4. (Optional) In the early stage of CAT, score estimates often are unstable (especially with the MLE method). You can choose to limit the level of fluctuation in estimates during the first several item administrations.
5. (Optional) You can choose to have final score estimates computed using MLE even if you selected MAP or EAP as the main estimation method.

B. Test Administration (Orange Box)

1. Specify the number of examinees that for each time slot in the simulated CAT administration. If you want to administer CAT with all examinees simultaneously, put "0" (default) in "Number of Examinees for Each Test Time Slot." Note: The program assumes that all examinees for each test time slot make CAT progress at the same pace.
2. Specify "Number of Test Time Slots per Day." A default value is "1."
(These inputs are closely related to "Item Usage Update" in "Exposure Control.")

C. Pretest Item Administration (Blue Box)

1. Specify the number of pretest items to be administered with each examinee. The pretest item pool data file (in *.wgi or *.wgix format) should be loaded using the "Open Pretest Item File" button. The pretest items are randomly selected for each examinee, and examinees' responses will not be used for scoring. The pretest item administration results will be stored in a separate file (*.scp).

D. Extras (Brown Box)

1. Generate Replication Data Sets: *SimulCAT* will replicate as many CAT simulations as specified here.
2. Item Pool with DIF/Drift: To simulate CAT with DIF/item parameter drift (IPD), check this box and provide an item pool data file containing the DIF/IPD affected item parameter values. The DIF/IPD item pool data file (*.wgi or *.wgix) must have item parameters for all items (even if items are not all of DIF/IPD). *SimulCAT* uses the DIF/IPD item parameters only to generate responses. During the item selection process, *SimulCAT* uses the original item pool data.

3. Previous Item Usage Data: Check this box only if you selected the item retirement option in Exposure Control and there is preexisting item usage data.
4. Fixed Seed Value: Fix the seed value for simulation. This is useful if you want to replicate the exact same study.

E. Outputs (Pink Box)

1. Select how you want to store the simulation results in the output file (*.sca). The item use information will be stored in a separate file (*.scu). A full response matrix (optional) will be stored in a separate file (*.dat).

F. Simulation Run (Black Box)

1. Specify the filename of the main output file (*.sca).
2. After reviewing all your selections in Steps 1, 2, and 3, click the “Run Simulation” button to run the CAT simulation.
3. Messages from *SimulCAT* and the progress of CAT simulation will be displayed in the “Log/Message” box.

G. Examples

To run examples, select “File>Open>Syntax” and choose an example syntax file. Once a syntax file is successfully loaded, review all settings throughout Steps 1, 2 and 3. Click “Run Simulation” in Step 3. For more information about file formats used in *SimulCAT* see Chapter VI. For more information about *SimulCAT* syntax commands, see Chapter VII.

Example Scenario 1 (Example syntax file: Example_syntax_I1E3V.scs)

1,000 examinees from a normal distribution $N(0,1)$ (from a data file, “Example_simulee1000.wge”)
 Item pool with 500 items (from a data file, “Example_ItemPool500.wgix”)
 Maximum Fisher Information Criterion for item selection
 Randomesque method for item exposure control
 Initial theta estimate is a random value between -0.5 and 0.5
 Interim theta is estimated using the EAP method.
 Final theta is estimated using the MLE method.

VI. *SimulCAT* File Formats

1. File Extensions

SimulCAT uses and produces several kinds of input and output files. Unique extensions are assigned to files according to their purpose. Several file formats are the same as ones that are used with WinGen (Han, 2007a, 2007b). Table 6.1 summarizes the types of files associated with *SimulCAT*.

Table 6.1 *Extensions of SimulCAT Files*

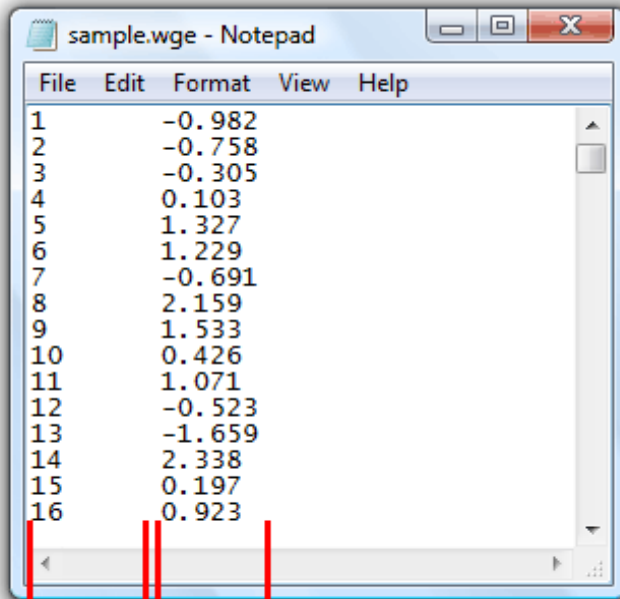
Extension	Description	Type
*.cue	<i>SimulCAT</i> cue file for executing sets of syntax files	Input only
*.log	<i>SimulCAT</i> log (for each 'Run Simulation')	Output only
*.wge	WinGen/ <i>SimulCAT</i> data file for examinees	Input and output
*.wgi	WinGen data file for Item Parameters (without content variables)	Input only
*.wgix	<i>SimulCAT</i> data file for Item Parameters (with content variables)	Input and output
*.dat	<i>SimulCAT</i> output for full response data matrix	Output only
*.sca	<i>SimulCAT</i> output for CAT administration	Output only
*.scc	<i>SimulCAT</i> input for content balancing information (two different formats exist)	Input only
*.sce	<i>SimulCAT</i> data file for item exposure parameters	Input and output
*.scs	<i>SimulCAT</i> syntax file	Input only
*.scp	<i>SimulCAT</i> output for response data for pretesting items	Output only
*.scu	<i>SimulCAT</i> output for item usage information Can be used as 'previous item exposure' data.	Input and output

2. *SimulCAT* File Formats

All input and output files of *SimulCAT* are ASC text file format and can be opened and edited with Notepad, TextPad, MS Excel, SPSS, SAS, etc.

A. Sample File Format : WinGen/*SimulCAT* Examinee Data File (*.wge) – ‘tab-delimited’

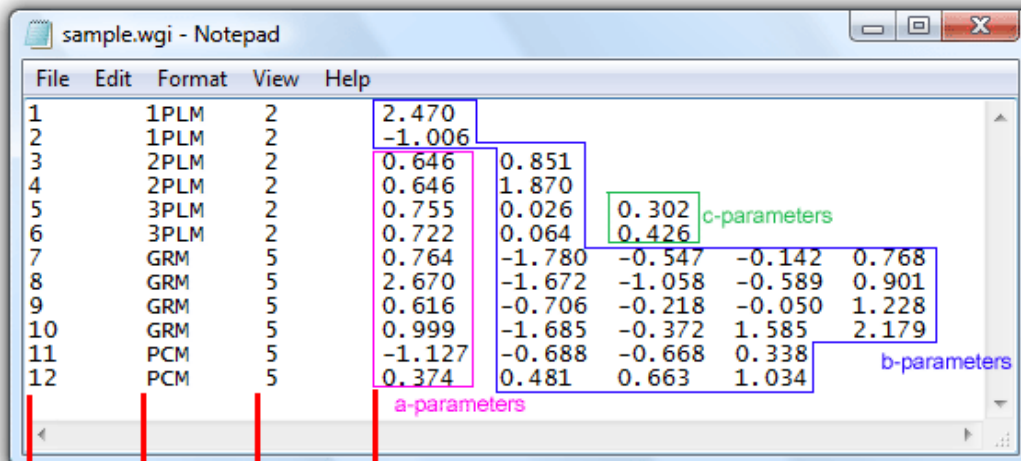
Format: [Examinee #][Theta]



Trait scores: 5 digits
Tab-delimited
ID: 8 characters

B. Sample File Format: WinGen/*SimulCAT* Item Parameter Data File (*.wgi) – ‘tab-delimited’

Format: [Item#][Model][# of categories][a-parameters][b-parameters][c-parameters]



a-parameters
b-parameters
c-parameters
of categories
Model
ID: 8 char.
... a-parameters, b-parameters, c-parameters

Models available [**NOTE:** *SimulCAT* currently does not support any polytomous response models, e.g., GRM, PCM, GPCM, NRM, and RSM.]:

- 1PLM: One Parameter Logistic Model
- 2PLM: Two Parameter Logistic Model
- 3PLM: Three Parameter Logistic Model
- GRM: Graded Response Model
- PCM: Partial Credit Model
- GPCM: General Partial Credit Model
- NRM: Nominal Response Model

RSM: Rating Scale Model

C. *SimulCAT* Extended Item Parameter Data File (*.wgix) – ‘tab-delimited’

Format: [Item#][Content Code][Model][# of categories][a-parameters][b-parameters][c-parameters]

The only difference between the *.wgix format and *.wgi is additional information about content (content code, which always has to be integer) after the item number. This is a mandatory format if content balancing is performed in the simulation.

Example File > *Example_ItemPool500.wgix*

D. *SimulCAT* Administration Result File (*.sca) – ‘tab-delimited’ (partially ‘comma-delimited’ for a list of interim values)

Format:[Replication # (only if there is more than one replication)]

[Day #]

[Slot #]

[Examinee #]

[True theta value]

[# of items administered]

[Final theta estimate]

[SEE for the final theta estimate]

[Response string (if the output option was selected)]

[Administered item IDs (if the output option was selected)]

[Initial & interim theta estimates (if the output option was selected)]

[Interim SEEs (if the output option was selected)]

[Interim test information (if the output option was selected)]

[True interim SEEs (if the output option was selected)]

[True interim test information (if the output option was selected)]

Example File> *Example_output_I2E4F.SCA* (with an option for saving response strings and administered item IDs)

E. *SimulCAT* Item Usage (Exposure) Data File (*.scu) – ‘tab-delimited’

Format: [Replication # (only if there are more than one replication)][Item #][# of item administration] [Retirement day if the item was retired during the test window]

Example File> *Example_output_I1E3V.SCU*

F. *SimulCAT* Item Exposure Parameter File (*.sce) – ‘tab-delimited’

Format: [Item #][Exposure parameter]

Example File> *Example_SCE_for_SHM.SCE*

[NOTE: *.sce file for the conditional matrimonial method (CMM) has additional information in the first line; [# of intervals], [lower bound], [upper bound]]

Example File> *Example_SCE_for_MCC.SCE*

G. *SimulCAT* pretesting item administration data file (*.scp) – ‘tab-delimited’

Format: [Examinee ID (8 characters)][blank (2 spaces)]

[True theta value (6 characters)][blank (2 spaces)]

[Final score estimate (6 characters)][blank (2 spaces)]

[Response data]

H. *SimulCAT* Full Response Matrix File (*.dat) – fixed format

Format: [Examinee ID(8 characters)][blank (2 spaces)][Response data]

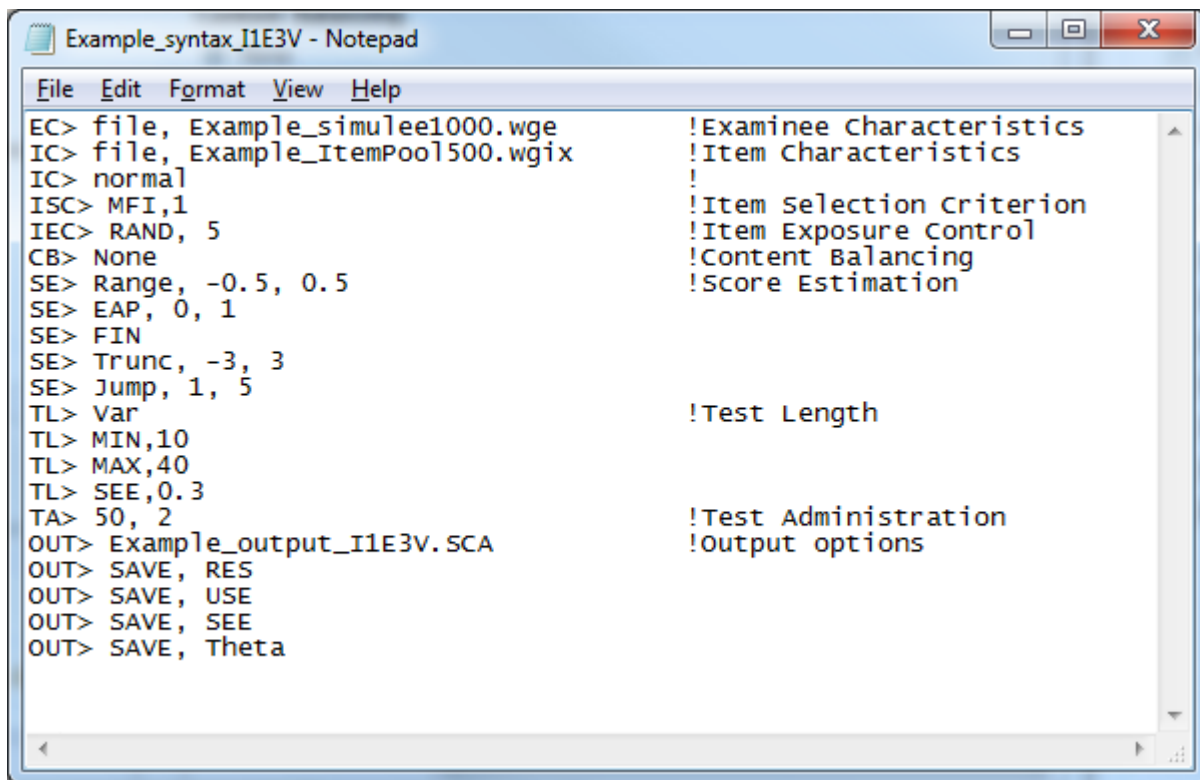
Example File> *Example_output_I2E4F.DAT*

VII. Advanced Uses of *SimulCAT*

Using a Syntax File

A syntax file can be used to run *SimulCAT* instead of the point-and-click method of the graphical user interface. Syntax files for WinGen can be composited using any kind of text editing software such as 'Notepad' or 'TextPad'.

The structure of a syntax file is straightforward—there is one command/option per line. Each line starts with an abbreviation for the corresponding section in the interface, followed by “>” and a choice of options. If an option has multiple inputs, they should be delimited by “,” (comma). See the example below for an illustration.



```

Example_syntax_I1E3V - Notepad
File Edit Format View Help
EC> file, Example_simulee1000.wge      !Examinee Characteristics
IC> file, Example_ItemPool500.wgix    !Item Characteristics
IC> normal                             !
ISC> MFI,1                             !Item selection Criterion
IEC> RAND, 5                           !Item Exposure Control
CB> None                               !Content Balancing
SE> Range, -0.5, 0.5                   !Score Estimation
SE> EAP, 0, 1
SE> FIN
SE> Trunc, -3, 3
SE> Jump, 1, 5
TL> Var                                !Test Length
TL> MIN,10
TL> MAX,40
TL> SEE,0.3
TA> 50, 2                              !Test Administration
OUT> Example_output_I1E3V.SCA         !Output options
OUT> SAVE, RES
OUT> SAVE, USE
OUT> SAVE, SEE
OUT> SAVE, Theta

```

It should be noted that when *SimulCAT* runs with a syntax file, it can only read existing data for examinee and item characteristics. To generate random examinee and/or item data, *SimulCAT* should be used with the graphical user interface, not with a syntax file. Text/syntax after “!” is

recognized as a 'comment' and ignored by *SimulCAT*. Table 7.1 displays the complete list of abbreviations and options for a syntax file.

Table 7.1 Abbreviations /Options for SimulCAT Syntax Files	
Abbreviation	Option
<u>EC</u> (Examinee Characteristics)	<p>['file,' a full filename with a complete directory name] e.g. EC> file, c:\simulcatStudy\examinee.wge</p>
<u>IC</u> (Item Characteristics)	<p>['file,' a full filename with a complete directory name] e.g. IC> file, c:\simulcatStudy\item.wgix</p> <p>['normal'] – scale to normal metric (D=1.702 instead of 1.0) e.g. IC> normal</p>
<u>ISC</u> (Item Selection Criteria)	<p>['MFI'] – the maximum Fisher information. e.g. ISC> MFI !MFI method</p> <p>['STRA,' # of strata, 'BB'] – the a-stratification method. The number of strata should be specified after a comma. e.g. ISC> STRA, 4 !a-stratification with 4 strata</p> <p>To select 'b-blocking', ['BB'] should be followed after the number of strata. e.g. ISC> STRA, 4, BB !a-stratification with b-blocking (with 4 strata)</p> <p>['MAT'] – the matching <i>b</i>-value method. ['RAN'] – the randomization method. ['IIC'] – the interval information criterion (IIC). ['LWI'] – the likelihood weighted information (LWI). ['KLI,' constant <i>c</i> value, # of items for randomesque] – the Kullbak-Leibler (global) information (KLI). For this option, the value of the constant 'c' should be specified. e.g. ISC> KLI, 3, 2 !KLI with constant c=3 with the randomesque with 2 items.</p> <p>['GMIR'] – the gradual maximum information ratio method. ['EBI'] – the efficiency balanced information method.</p>
<u>IEC</u> (Item Exposure Control)	<p>['NON'] – no exposure control.</p> <p>['FAM,' exposure target] – the fade-way method. For this option, the exposure rate target should be specified. e.g. IEC> FAM, 0.2</p> <p>['RAN,' # of items for randomesque] - the randomesque method. The number of items for the randomesque procedure also should be specified after a comma. e.g. IEC> RAN, 2 !the randomesque with 2 items.</p>

Table 7.1 Abbreviations /Options for SimulCAT Syntax Files

Abbreviation	Option
	<p data-bbox="526 352 1398 411">[‘SHM, FILE,’ the full filename of a *.sce file] – the Sympson and Hetter method with pre-computed exposure parameter data. e.g.</p> <pre data-bbox="526 415 1382 443">IEC> SHM, FILE, c:\simulcatStudy\expo.sce</pre> <p data-bbox="526 478 1382 537">[‘SHM,’ # of iterations, exposure target] – the Sympson and Hetter method after computing the exposure parameters for the specified exposure rate target. e.g.</p> <pre data-bbox="526 562 1382 590">IEC> SHM, 30, 0.2</pre> <p data-bbox="526 632 1398 690">[‘UCM,’ # of iterations, exposure target] – the unconditional multinomial method after computing the exposure parameters for the specified exposure rate target. e.g.</p> <pre data-bbox="526 716 1382 743">IEC> UCM, 30, 0.2</pre> <p data-bbox="526 785 1430 877">[‘CCM,’ # of iterations, exposure target, # of intervals for theta, lower bound of theta interval, upper bound of theta interval] – the conditional multinomial method after computing the exposure parameters for the specified exposure rate target . e.g.</p> <pre data-bbox="526 903 1382 930">IEC> CCM, 30, 0.2, 6, -3, 3</pre> <p data-bbox="526 972 1430 1064">[‘CCM, FILE,’ the full filename of a *.sce file] – the conditional multinomial method with pre-computed exposure parameter data. Note that *.sce files contain the interval information in the first line. e.g.</p> <pre data-bbox="526 1089 1382 1117">IEC> CCM, FILE, c:\simulcatStudy\expo.sce</pre> <p data-bbox="526 1138 1349 1197">[‘RET,’ target item usage] – letting items permanently retire once their usage reaches to the target e.g.</p> <pre data-bbox="526 1222 1382 1249">IEC> RET, 3000</pre> <p data-bbox="526 1255 1373 1283">[‘UPDATE,’ [‘REAL’ or ‘SLOT’ or ‘DAY’] – frequency of item usage update. e.g.,</p> <pre data-bbox="526 1308 1382 1335">IEC> UPDATE, SLOT</pre>
<p data-bbox="269 1354 418 1438" style="text-align: center;"><u>TL</u> (Test Length)</p>	<p data-bbox="526 1383 1162 1442">[‘FIX,’ X] – fixed length test. X items will be administered. e.g.</p> <pre data-bbox="526 1446 1382 1474">TL> FIX, 30</pre> <p data-bbox="526 1516 1406 1608">[‘VAR’] – variable length test. Additional information about termination rules should be specified <u>in separate lines</u>. At least one termination rule has be specified (among SEE, EST, and MAX).</p> <p data-bbox="526 1640 1398 1667">[‘SEE,’ X] – terminating CAT when SEE of examinee’s score estimate reaches X.</p> <p data-bbox="526 1703 1406 1761">[‘EST,’ X, Y] – terminating CAT when examinee’s score estimate does not change by X for the last Y item administrations.</p> <p data-bbox="526 1793 1406 1852">[‘MAX,’ X] – terminating CAT when the number of administered items reaches X. This rule overrules SEE and EST rules.</p>

Table 7.1 Abbreviations /Options for SimulCAT Syntax Files	
Abbreviation	Option
	<p>['MIN,' X] – keeping from terminating before the number of administered items reaches X. This rule overrules SEE and EST rules.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>TL> VAR !CAT is variable length.</p> <p>TL> SEE, 0.3 !CAT is terminated when SEE<=0.3</p> <p>TL> EST, 0.1 ,5 !CAT is terminated when theta estimate does not change by 0.1 for the last five item administrations.</p> <p>TL> MAX, 40 !CAT is terminated when 40 items are administered regardless of SEE and EST values.</p> <p>TL> MIN, 10 !CAT is not terminated before at least 10 items are administered.</p> </div> <p>['EXP,' #] – expected test length. This is required information when the a-stratification or GMIR method was selected for item selection and when the test length is variable.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>TL> EXP, 30 !User expects a typical length of CAT would be 30 items.</p> </div>
CB (Content Balancing)	<p>['NON'] – no content balancing</p> <p>['SCR,' a full filename with a complete directory name] – content balancing by a script.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>CB> SCR, c:\simulcatStudy\script.scc</p> </div> <p>['WGT,' a full filename with a complete directory name] – content balancing by a weight (or percentage).</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>CB> WGT, c:\simulcatStudy\script.scc</p> </div>
SE (Score Estimation)	<p>['MLE'] – the maximum likelihood Estimation</p> <p>['MAP,' X, Y] – the Bayesian maximum a posteriori estimation with a posterior distribution with mean of X and SD of Y.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>SE> MAP, 0, 1 !MAP estimation with a posteriori of N(0,1)</p> </div> <p>['EAP,' X, Y] – the Bayes expected a posteriori estimation with a posterior distribution with mean of X and SD of Y.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>SE> EAP, 0, 1 !EAP estimation with a posteriori of N(0,1)</p> </div> <p>['FIX,' X] – the initial score is fixed to X.</p> <p>['RAN,' X, Y] – the initial score is a random value between X and Y.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>SE> RAN, -0.5, 0.5 !Initial theta value is a random value between -0.5 and 0.5.</p> </div> <p>['FILE,' X, a full filename with a complete directory name] – the initial score is loaded from an existing data file (*.wge).</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>SE> FILE, c:\simulcatStudy\oldScore.wge</p> </div> <p>['TRUNC,' X, Y] – the score estimates are truncated to be between X and Y.</p> <p>e.g.,</p> <div style="border: 1px solid black; padding: 5px;"> <p>SE> TRUNC, -3, 3 !Theta estimates are truncated to be b/w -3 and 3.</p> </div>

Table 7.1 Abbreviations /Options for SimulCAT Syntax Files	
Abbreviation	Option
	<p>['JUMP,' X, Y] – the change in score estimate cannot exceed X from the previous estimate until Yth item administration. e.g., SE> JUMP, 1, 5</p> <p>['FINAL'] – the final score is estimated using the MLE regardless a choice of proficiency estimation method. e.g., SE> FINAL</p>
<u>TA</u> (Test Administration)	<p>[X, Y] – X is the number of examinees taking the exam at the same time (i.e., the maximum seating capacity of testing centers). Y is the number of testing slots per day. If all examinees are to be administered simultaneously, put "0, 0", which is the default. e.g., TA> 500, 2 !500 examinees are administered in each testing slot. Two testing slots are administered per day. (i.e., 1000 examinees per day)</p>
<u>EXT</u> (Extras)	<p>['REP,' X] – Replicating the simulation X times. ['DIF,' a full filename with a complete directory name] – Introducing DIF/IPD from the item parameter data (*.wgi or *.wgix). ['EXP,' a full filename with a complete directory name] – Setting the initial item usage values to the previous item usage value (*.scu). ['SEED,' X] – Using X as a SEED value for simulation. e.g., EXT> REP, 10 !Replicates 10 times EXT> DIF, c:\simulcatStudy\DIF_Param.wgi !Examinees' responses are simulated based on the DIF item parameters. EXT> EXP, c:\simulcatStudy\prevUsage.scu !item usage information incorporates the previous information. EXT> SEED, 61346125 !SEED value is 61346125</p> <p>['IP,' item pocket size] – simulating the worst case scenario for item pocket option.</p>
<u>PIA</u> (Pretest Item Administration)	<p>['NON'] – no precalibrated item to be administered. [X, a full filename with a complete directory name] – administering X pretesting items to each examinee from a pretesting item pool (*.wgi or *.wgix). e.g., PIA> 5, c:\simulcatStudy\preTestingItems.wgi !Each examinee takes 5 precalibrated items that are randomly selected from 'preCallItems.wgi'.</p>
<u>OUT</u> (Outputs)	<p>['SAVE, RES'] – Saving the response strings and item IDs in *.sca. ['SAVE, THE'] – Saving all interim theta estimates in *.sca. ['SAVE, SEE'] – Saving all interim SEE and test information values in *.sca. ['SAVE, TRU'] – Saving all interim SEE and test information values at the true theta in *.sca. ['SAVE, USE'] – Saving item usage information in *.scu. ['SAVE, USE'] – Saving item usage information in *.scu.</p>

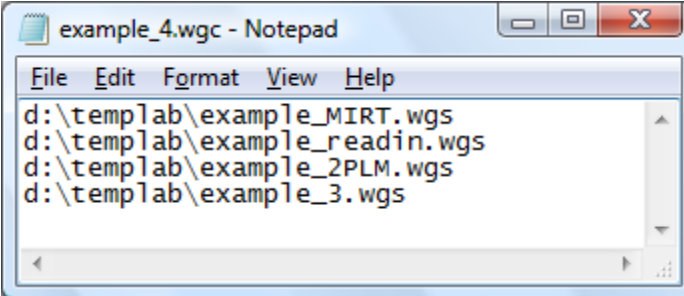
Table 7.1 Abbreviations /Options for SimulCAT Syntax Files

Abbreviation	Option
	['SAVE, FULL'] – Saving a full response matrix in *.dat. e.g., OUT> SAVE, RES OUT> SAVE, THE OUT> SAVE, SEE OUT> SAVE, USE

Cue File

A cue file is a batch file with which *SimulCAT* runs multiple syntax files. Basically, it is a list of the full names of the syntax files. A cue file can be executed at “File>Run a Cue File” on the program menu bar.

Example of a Cue File



```

example_4.wgc - Notepad
File Edit Format View Help
d:\templab\example_MIRT.wgs
d:\templab\example_readin.wgs
d:\templab\example_2PLM.wgs
d:\templab\example_3.wgs
  
```

References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (Chaps. 17-20). Reading, MA: Addison-Wesley.
- Chang, H.-H., Qian, J., & Ying, Z. (2001). Alpha-stratified multistage computerized adaptive testing with beta blocking. *Applied Psychological Measurement, 25*, 333–341.
- Chang, H.-H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20*, 213–229.
- Chang, H.-H., & Ying, Z. (1999). Alpha-stratified multistage computerized adaptive testing. *Applied Psychological Measurement, 23*, 211–222.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.
- Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.
- Han, K. T. (2007a). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement, 31*(5), 457-459.
- Han, K. T., & Hambleton, R. K. (2007b). User's Manual: WinGen (*Center for Educational Assessment Report No. 642*). Amherst, MA: University of Massachusetts, School of Education.
- Han, K. T. (2009). *A gradual maximum information ratio approach to item selection in computerized adaptive testing*. Research Reports 09–07, McLean, VA: Graduate Management Admission Council.
- Han, K. T. (2010). *SimulCAT*: Simulation software for computerized adaptive testing [computer program]. Retrieved March 20, 2010, from <http://www.hantest.net/>
- Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education, 2*(4), 359–375.
- Stocking, M. L., & Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing*. Research Report 95–25. Princeton, NJ: Educational Testing Service.
- Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 23*(1), 57–75.
- Sympson, J. B. & Hetter, R. D. (1985). *Controlling item-exposure rates in computerized adaptive testing*. In Proceedings of the 27th annual meeting of the Military Testing Association, (pp. 973–977), San Diego, CA: Navy Personnel Research and Development Centre.

Veerkamp, W. J. J., & Berger, M. P. F. (1997). Some new item selection criteria for adaptive testing, *Journal of Educational and Behavioral Statistics*, 22(2), 203–226.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473–492.

Acknowledgements

The author is very grateful to Ronald K. Hambleton of the University of Massachusetts in Amherst for important feedback. The author also wishes to thank Lawrence M. Rudner and Fanmin Guo of the Graduate Management Admission Council® for valuable comments and support.

Author's Address

Correspondence concerning *SimulCAT* should be addressed to Kyung T. Han, Graduate Management Admission Council, 11921 Freedom Dr., Suite 300, Reston, VA 20190; email: trueTheta@gmail.com.