7

Scaling Methodology and Procedures for the Mathematics and Science Literacy, Advanced Mathematics, and Physics Scales

Greg Macaskill
Raymond J. Adams
Margaret L. Wu
Australian Council for Educational Research

Student achievement is reported in TIMSS mainly through scale scores derived using Item Response Theory (IRT) scaling. This approach allows the performance of a sample of students in a subject area to be summarized on a common scale or series of scales even when different students have been administered different items. The common scale makes it possible to report on relationships between students' characteristics (based on responses to the background questionnaires) and their performance in mathematics and science.

For Population 3, as for Populations 1 and 2, each student was administered only a subset of items within each content area in the three areas examine – advanced mathematics, physics and mathematics and science literacy. In this situation, to obtain reliable indices of student proficiency "plausible values" methodology was used. Some references to this work are given in Adams, Wu and Macaskill (1997).

This chapter gives details of the IRT model used in TIMSS to scale the Population 3 achievement data and includes a description of the model and the estimation process. For more details, see also the reference above and papers cited within this chapter.

7.1 THE TIMSS SCALING MODEL

The scaling model used in TIMSS was the multidimensional random coefficients logit model described by Adams, Wilson, and Wang (1997), with the addition of a multivariate linear model imposed on the population distribution. The scaling was done with the ConQuest software (Wu, Adams, and Wilson, 1997) that was developed in part to meet the needs of the TIMSS study.

7.1.1 The Multidimensional Random Coefficients Model

Assume that I items are indexed $i=1,...,I$ with each item admitting $K_i + 1$ response alternatives $k=0,1,...,K_i$. Use the vector valued random variable, $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iK_i})'$ where

$$X_{ij} = \begin{cases} 1 \text{ if response to item } i \text{ is in category } j \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

to indicate the $K_i + 1$ possible responses to item i.

A response in category zero is denoted by a vector of zeroes. This effectively makes the zero category a reference category and is necessary for model identification. The choice of this as the reference category is arbitrary and does not affect the generality of the model. We can also collect the $\mathbf{X}_i$ together into the single vector $\mathbf{X}' = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_I)$ which we call the response vector (or pattern). Particular instances of each of these random variables are indicated by their lower-case equivalents; $\mathbf{x}$, $\mathbf{x}_i$ and x_{ik}.

The items are described through a vector $\xi^T = (\xi_1, \xi_2, \ldots, \xi_p)$ of p parameters. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. These linear combinations are defined by design vectors $\mathbf{a}_{jk}$, ($j = 1,\ldots,I$; $k = 1,\ldots,K_i$) each of length p that can be collected to form a design matrix $\mathbf{A}' = (\mathbf{a}_{11}, \mathbf{a}_{12}, \ldots, \mathbf{a}_{1K_1}, a_{21}, \ldots, a_{2K_2}, \ldots, a_{1K_I})$.

The multidimensional form of the model assumes that a set of D traits underlie the individuals' responses. The D latent traits define a D-dimensional latent space and the individuals' positions in the D-dimensional latent space are represented by the vector $\theta = (\theta_1, \theta_2, \ldots, \theta_D)$.

An additional feature of the model is the introduction of a scoring function which allows the specification of the score or "performance level" that is assigned to each possible response to each item. To do this we introduce the notion of a response score b_{ijd} that gives the performance level of an observed response in category j of item I in dimension d. The scores across D dimensions can be collected first into a column vector $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ik1D})^T$, then into the scoring sub-matrix for item i, $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iD})^T$, and then into a scoring matrix $\mathbf{B} = (\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T)^T$ for the whole test. (By definition, the score for a response in the zero category is zero, but other responses may also be scored zero.)

The probability of a response in category k of item i is modeled as

$$\Pr(\mathbf{X}_{ij} = 1; \mathbf{A}, \mathbf{B}, \xi | \theta) = \frac{\exp(\mathbf{b}_{ij}\theta + \mathbf{a}'_{ij}\xi)}{\sum_{k=1}^{K_i} \exp(\mathbf{b}_{ik}\theta + \mathbf{a}'_{ik}\xi)}. \tag{2}$$

And for a response vector we have

$$f(\mathbf{x}; \xi | \theta) = \Psi(\theta, \xi) \exp[\mathbf{x}'(\mathbf{B}\theta + \mathbf{A}\xi)] \tag{3}$$

with

$$\Psi(\theta, \xi) = \left\{ \sum_{z \in \Omega} \exp[\mathbf{z}^T(\mathbf{B}\theta + \mathbf{A}\xi)] \right\}^{-1} \tag{4}$$

where Ω is the set of all possible response vectors.

7.2 THE POPULATION MODEL

The item response model is a conditional model, in the sense that it describes the process of generating item responses conditional on the latent variable, θ. The complete definition of the TIMSS model, therefore, requires the specification of a density, $f_\theta(\theta;\alpha)$ for the latent variable θ. We use **a** to symbolize a set of parameters that characterize the distribution of θ. The most common practice when specifying unidimensional marginal item response models is to assume that the students have been sampled from a normal population with mean *m* and variance s^2. That is:

$$f_\theta(\theta;\alpha) \equiv f_\theta(\theta;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right] \tag{5}$$

or equivalently

$$\theta = \mu + E \tag{6}$$

where $E \sim N(0,\sigma^2)$.

A natural extension of (5) is to replace the mean, *m* with the regression model $\mathbf{Y}_n^T\beta$, where $\mathbf{Y}_n$ is a vector of *u*, fixed and known values for student *n*, and β is the corresponding vector of regression coefficients. For example, $\mathbf{Y}_n$ could be constituted of student variables such as gender, socio-economic status, or major. Then the population model for student *n* becomes

$$\theta_n = \mathbf{Y}_n^T\beta + E_n \tag{7}$$

where we assume that the E_n are independently and identically normally distributed with mean zero and variance s^2 so that (7) is equivalent to

$$f_\theta(\theta_n;\mathbf{Y}_n,\mathrm{b},\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(\theta_n-\mathbf{Y}_n^T\beta)^T(\theta_n-\mathbf{Y}_n^T\beta)\right] \tag{8}$$

a normal distribution with mean $\mathbf{Y}_n^T\beta$ and variance s^2. If (8) is used as the population model then the parameters to be estimated are ***b***, s^2 and **x.**

The TIMSS scaling model takes the generalization one step further by applying it to the vector valued θ rather than the scalar valued θ, resulting in the multivariate population model

$$f_\theta(\theta_n;\mathbf{W}_n,\gamma,\Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\theta_n-\gamma\mathbf{W}_n)^T\Sigma^{-1}(\theta_n-\gamma\mathbf{W}_n)\right] \tag{9}$$

where γ is a $u \times d$ matrix of regression coefficients, Σ is a $d \times d$ variance-covariance matrix and $\mathbf{W}_n$ is a $u \times 1$ vector of fixed variables. If (9) is used as the population model then the parameters to be estimated are γ, Σ and **x.** In TIMSS we refer to the $\mathbf{W}_n$ variables as conditioning variables.

7.3 ESTIMATION

The ConQuest software uses maximum likelihood methods to provide estimates of γ, Σ and $\mathbf{x}$. Combining the conditional item response model (3) and the population model (9) we obtain the unconditional or marginal response model

$$f(\mathbf{x};\xi,\gamma,\Sigma) = \int_{\theta} f_{x}(\mathbf{x};\xi|\theta) f_{\theta}(\theta;\gamma,\Sigma)d\theta \tag{10}$$

and it follows that the likelihood is

$$\Lambda = \prod_{n=1}^{N} f_{x}(\mathbf{x}_n;\xi,\gamma,\Sigma) \tag{11}$$

where N is the total number of sampled students.

Differentiating with respect to each of the parameters and defining the marginal posterior as

$$h_{\theta}(\theta_n;\mathbf{W}_n,\xi,\gamma,\Sigma|\mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n\xi|\theta_{\mathrm{n}}) f_{\theta}(\theta_n;\mathbf{W}_n\gamma,\Sigma)}{f_{\mathbf{x}}(\mathbf{x}_n;\mathbf{W}_n,\xi,\gamma,\Sigma)} \tag{12}$$

provides the following system of likelihood equations:

$$\mathbf{A}'\sum_{n=1}^{N}\left[\mathbf{x}_n - \int_{\theta_n} \mathrm{E}_z(z|\theta_n) h_{\theta}(\theta_n;\mathbf{Y}_n,\xi,\gamma,\Sigma|\mathbf{x}_n)d\theta_n\right] = 0 \tag{13}$$

$$\hat{\gamma} = \left(\sum_{n=1}^{N}\bar{\theta}_n\mathbf{W}_n^T\right)\left(\sum_{n=1}^{N}\mathbf{W}_n\mathbf{W}_n^T\right)^{-1} \tag{14}$$

and

$$\hat{\Sigma} = \frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}}\int_{\theta_{\mathrm{n}}}(\theta_n - \gamma\mathbf{W}_n)(\theta_n - \gamma W_n)^T h_{\theta}(\theta_n;\mathbf{Y}_n,\xi,\gamma,\Sigma|\mathbf{x}_n)d\theta_n \tag{15}$$

where

$$\mathrm{E}_{\mathbf{z}}(\mathbf{z}|\theta_n) = \Psi(\theta_n,\xi)\sum_{\mathbf{z}\in\Omega}\mathbf{z}\exp[\mathbf{z}'(\mathbf{b}\theta_n + \mathbf{A}\xi)] \tag{16}$$

and

$$\bar{\theta}_n = \int_{\theta_n} \theta_n h_\theta(\theta_n;\mathbf{Y}_n, \xi, \gamma, \Sigma|\mathbf{x}_n) d\theta_n . \tag{17}$$

The system of equations defined by (13), (14), and (15) is solved using an EM algorithm.

7.3.1 Quadrature and Monte Carlo Approximations

The integrals in equations (13), (14), and (15) are approximated numerically using either quadrature or Monte Carlo methods. In each case we define, Θ_p, $p=1,...,P$ a set of P D-dimensional vectors (which we call nodes) and for each node we define a corresponding weight $W_p(\gamma, \Sigma)$. The marginal item response probability (10) is then approximated using

$$f_{\mathbf{x}}(\mathbf{x};\xi, \gamma, \Sigma) = \sum_{p=1}^{P} f_{\mathbf{x}}(\mathbf{x};\xi|\Theta_p) W_p(\gamma, \Sigma) \tag{18}$$

and the marginal posterior (12) is approximated using

$$h_\Theta(\Theta_q;\mathbf{W}_n, \xi, \gamma, \Sigma|\mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n;\xi|\Theta_q) W_q(\gamma, \Sigma)}{\sum_{p=1}^{P} f_{\mathbf{x}}(\mathbf{x};\xi|\Theta_p) W_p(\gamma, \Sigma)} \tag{19}$$

for q=1,...,P.

The difference between the quadrature and Monte Carlo methods lies in the way the nodes and weights are prepared. For the quadrature case we begin by choosing a fixed set of Q points, $(\Theta_{d1}, \Theta_{d1}, \ldots, \Theta_{d1Q})$, for each latent dimension and then define a set of Q^D nodes that are indexed r = 1,…,Q^D, and are given by the Cartesian coordinates

$$\Theta_r = (\Theta_{1j_1}, \Theta_{2j_2}, \ldots, \Theta_{dj_d}) \text{ with } j_1 = 1,\ldots,Q;\ j_2 = 1,\ldots,Q;\ \ldots;\ j_d = 1,\ldots,Q.$$

The weights are then chosen to approximate the continuous latent population density (9), that is,

$$W_p = K(2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\Theta_p - \gamma\mathbf{W}_n)^T\Sigma^{-1}(\Theta_p - \gamma\mathbf{W}_n)\right] \tag{20}$$

where K is a scaling factor to ensure that the sum of the weights is one.

In the Monte Carlo case the nodes are drawn at random from the standard multivariate normal distribution, and at each iteration the nodes are rotated using standard methods so that they become random draws from a multivariate normal distribution with mean $\gamma\mathbf{W}_n$ and variance Σ. In the Monte Carlo case the weight for all nodes is $1/P$.

7.3.2 Latent Estimation and Prediction

The marginal item response (10) does not include parameters for the latent values θ_n and hence the estimation algorithm does not result in estimates of the latent values. For TIMSS, the expected *a-posteriori* (EAP) prediction of each student's latent achievement was produced. The EAP prediction of the latent achievement for case n is

$$\theta_n^{EAP} = \sum_{r=1}^{p} \Theta_r h_\Theta(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n).[1] \tag{21}$$

Variance estimates for these predictions were estimated using

$$\mathrm{var}(\theta_n^{EAP}) = \sum_{r=1}^{P} (\Theta_r - \theta_n^{EAP})(\Theta_r - \theta_n^{EAP})^T h_\Theta(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n). \tag{22}$$

7.3.3 Drawing Plausible Values

Plausible values are random draws from the marginal posterior of the latent distribution, (12), for each student. Unlike previously described methods for drawing plausible values ConQuest does not assume normality of the marginal posterior distributions. Recall from (12) that the marginal posterior is given by

$$h_\theta(\theta_n; \mathbf{W}_n, \xi, \gamma, \Sigma | \mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n; \xi | \theta_n) f_\theta(\theta_n; \mathbf{W}_n, \gamma, \Sigma)}{\int_\theta f(\mathbf{x}; \xi | \theta) f_\theta(\theta, \gamma, \Sigma) d\theta}. \tag{23}$$

The ConQuest procedure begins by drawing M vector valued random deviates, $\{\varphi_{nm}\}_{m=1}^{M}$, from the multivariate normal distribution $f_\theta(\theta_n, \mathbf{W}_n\gamma, \Sigma)$ for each case n. These vectors are used to approximate the integral in the denominator of (23) using the Monte Carlo integration

$$\int_\theta f_{\mathbf{x}}(\mathbf{x}; \xi | \theta) f_\theta(\theta, \gamma, \Sigma) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} f_{\mathbf{x}}(\mathbf{x}; \xi | \varphi_{mn}) \equiv \Im \tag{24}$$

At the same time the values

$$p_{mn} = f_{\mathbf{x}}(\mathbf{x}_n; \xi | \varphi_{mn}) f_\theta(\varphi_{mn}; \mathbf{W}_n, \gamma, \Sigma) \tag{25}$$

are calculated, so that we obtain the set of pairs $\langle \varphi_{nm}, \frac{p_{mn}}{\Im} \rangle_{m=1}^{M}$, which can be used as an approximation to the posterior density (23), and the probability that φ_{nm} could be drawn from this density is given by

[1] The current version of ConQuest uses the Monte Carlo method only when producing EAP predictions and variances for those predictions.

$$q_{nj} = \frac{p_{mn}}{\sum_{m=1}^{M} p_{mn}}. \tag{26}$$

At this point L uniformly distributed random numbers, $\{\eta_i\}_{i=1}^{L}$. are generated and for each random draw the vector φ_{n1_0} that satisfies the condition

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_{\mathrm{i}} \leq \sum_{s=1}^{i_0} q_{sn} \tag{27}$$

is selected as a plausible vector.

7.4 SCALING STEPS

The model was fitted to the data in two steps. First the items were calibrated using the combined data from most of the countries in the population. This was called the international calibration sample. In the second stage, the model was fitted separately for each country with the item parameters fixed at the values estimated in the first step.

7.4.1 Details of the Calibration Samples

The item calibration was carried out using almost the entire sample from each of the three areas- advanced mathematics, physics, and mathematics and science literacy- where students who attempted test booklets 1A and 1B formed the mathematics and science literacy calibration sample, those who did booklets 2A-2C formed the physics sample, and the students who took booklets 3A-3C made up the advanced mathematics sample. There was a further group of students who did booklet 4, which was a selection from all three topics, who were excluded from the calibration.

Six sets of item parameters were derived from these three samples. For mathematics and science literacy, a two- dimensional run was performed for mathematics literacy and science literacy. Because these scales are quite highly correlated (about .85) it was thought better to obtain the parameters from a two-dimensional run rather than two separate unidimensional runs. For another scale, the reasoning and social utility scale[2], which is composed of a subset of the mathematics and science literacy items and includes both mathematics and science literacy items, item parameters were also estimated from the mathematics and science literacy sample. Item parameters for full advanced mathematics and physics scales were obtained by unidimensional runs from their respective samples and item parameters for a 3-subscale model for advanced mathematics and a 5-subscale model for physics were also estimated.

2 Results for the reasoning and social utility scale were not reported in the TIMSS international report, but scores on this scale are available in the TIMSS international database (Gonzalez, Smith, and Sibberns, 1998).

Table 7.1 shows the countries which were included in the calibration for the three subject areas and the size of the sample they contributed to the calibration sample.

Table 7.1 Countries and Numbers of Students in the Population 3 Calibration Samples

Country	Mathematics and Science Literacy	Advanced Mathematics	Physics
Australia	1844	548	564
Austria	1779	599	594
Canada	4832	2381	1967
Cyprus	473	330	307
Czech Republic	1899	833	819
Denmark	*	*	*
France	1590	796	835
Germany	2182	2189	616
Greece	*	346	349
Hungary	5091	–	–
Iceland	1703	–	–
Israel	*	*	*
Italy	1578	360	*
Latvia	–	–	708
Lithuania	2887	734	–
Netherlands	1470	–	–
New Zealand	1763	–	–
Norway	2518	–	1048
Russia	2289	1402	1129
Slovenia	1387	1301	512
South Africa	2757	–	–
Sweden	2816	749	760
Switzerland	2976	1072	1039
United States	5371	2349	2678
Total Sample	49205	15989	13925

(*) Administered test but not included in calibration sample.
(–) Did not participate in assessment.

Countries that were included in the study but wholly or partly omitted from the calibration samples for various reasons are indicated by an asterisk in the table above. For example, Italy was not used in the calibration sample, but was modelled in the second step of the scaling process. A dash indicates that the country did not participate in this part of the study. The table below shows the number of countries included in the study and the calibration samples.

Table 7.2 Number of Countries in TIMSS and in Calibration Samples

	Total	Mathematics and Science Literacy	Advanced Mathematics	Physics
In TIMSS	24	23	17	18
In Calibration	22	20	15	15

7.4.2 International Scaling Results

Tables 7.3 to 7.15 display basic statistics and item parameters, along with an indicator of the fit of each individual item parameter, for the scales derived from the six calibration runs described above. Most items were dichotomous, but 3- and 4 -category items were fitted with a partial-credit model. The item parameters here are given in the logit metric and in the item-step form described in Wu, Adams, and Wilson (1997). The mean square fit statistic is an index of the fit of the data to the assumed scaling model; the statistic given here was derived by Wu (1997). Under the null hypothesis that the data and model are consistent, the expected value of these statistics is one. Values that are less than one usually indicate items with greater than average discrimination, while values that are greater than one can result from lower than average discrimination, guessing, or some other deviation from the model.

Only some questions appeared in all booklets; for example, for advanced mathematics the I cluster items were given to all students, whereas the J, K, and L cluster items were each present in only one of the three booklets – 3A, 3B and 3C. Percent correct figures were calculated by summing the total of scores from all students who provided valid responses and dividing that by the number of students multiplied by the maximum score that could be achieved for that item. This reduces to the usual percent correct for the dichotomous items.

7.4.3 Fit of the Scaling Model

Tables 7.3 to 7.6 show the results for the overall advanced mathematics and physics scales and the mathematics literacy and science literacy scales. For the advanced mathematics scale (Table 7.3) items with fit statistics greater than or equal to 1.15 are J04, J12, J18, K08, K16, L16, and L18. Item J04, in Figure 7.1, seems to fit rather poorly, with markedly lower discrimination than the other items and a downward kink for some of the higher-ability students. This item proved to have a distractor with a positive biserial for several countries. Item J12 in Figure 7.2 shows some lower discrimination though not as dramatic as for J04, and also curvature in the response for lower- performing students. Figure 7. 3 demonstrates some lack of discrimination in item K08. No items were found to have fit statistics less than .85.

Table 7.3 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Advanced Mathematics Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI01	15987	58.7	-0.585	0.017	0.89
CSMMI02	15987	57.0	-0.652	0.017	0.89
CSMMI03	15987	61.4	-0.669	0.017	1.07
CSMMI04	15988	57.7	-0.602	0.017	0.96
CSMMI05	15975	35.3	0.529	0.018	1.04
CSMMI06	15975	47.6	-0.200	0.017	0.93
CSMMI07	15986	57.8	-0.498	0.017	0.98
CSMMI08	15987	74.9	-1.578	0.020	0.95
CSMMI09	15986	59.7	-0.745	0.018	1.00
CSMMI10	15985	58.2	-0.460	0.017	1.02
CSMMJ01	5391	53.3	-0.564	0.030	0.89
CSMMJ02	4807	35.3	0.431	0.033	1.03
CSMMJ03	5390	56.6	-0.630	0.030	0.94
CSMMJ04	5391	39.2	0.477	0.030	1.16
CSMMJ05	5394	56.1	-0.686	0.030	0.85
CSMMJ06	5392	33.5	0.597	0.030	0.96
CSMMJ07	5390	41.5	0.103	0.029	0.96
CSMMJ08	4606	67.7	-0.880	0.032	1.08
CSMMJ09	5387	22.5	1.168	0.033	1.05
CSMMJ10	5388	36.8	0.447	0.030	1.03
CSMMJ11	5393	68.8	-1.063	0.032	1.14
CSMMJ12	5393	82.5	-1.752	0.037	1.38
CSMMJ13	5392	46.2	-0.023	0.029	1.03
CSMMJ14	5392	47.4	-0.304	0.029	0.87
CSSMJ15A	5393	48.4	-0.118	0.029	0.91
CSSMJ15B	5394	7.5	2.701	0.052	0.99
CSSMJ16A	5393	67.3	-0.974	0.031	1.03
CSSMJ16B	5394	22.6	1.269	0.034	0.91
CSSMJ17	5032	27.1	0.540	0.019	1.01
CSSMJ17 (S1)			1.382	0.051	1.12
CSEMJ18	5125	14.7	0.933	0.021	1.24
CSEMJ18 (S1)			2.385	0.087	0.94
CSEMJ19	5392	34.1	0.249	0.018	1.03
CSEMJ19 (S1)			1.279	0.046	0.98
CSMMK01	5296	82.1	-2.012	0.040	1.08
CSMMK02	5296	23.3	1.022	0.033	1.03
CSMMK03	5295	63.7	-0.831	0.031	1.06
CSMMK04	5297	27.8	0.928	0.032	0.98
CSMMK05	5297	43.0	0.136	0.030	1.08
CSMMK06	5297	51.4	-0.418	0.030	0.97

Table 7.3 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Advanced Mathematics Scale (Continued)

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMK07	5297	53.4	-0.328	0.030	1.07
CSMMK08	5297	26.8	0.917	0.032	1.15
CSMMK09	5047	42.8	0.244	0.031	1.07
CSMMK10	5296	20.5	1.352	0.035	0.97
CSMMK11	5297	50.3	-0.061	0.030	0.96
CSSMK12	5295	50.2	-0.131	0.030	1.07
CSSMK13	5294	27.5	1.084	0.033	0.92
CSSMK14	5284	10.2	1.411	0.026	0.95
CSSMK14 (S1)			2.660	0.111	0.86
CSSMK15	5285	14.9	0.931	0.021	1.02
CSSMK15 (S1)			1.911	0.068	0.91
CSEMK16	5296	51.5	-0.173	0.014	1.27
CSEMK16 (S1)			0.538	0.029	0.94
CSEMK16 (S2)			-0.651	0.033	0.93
CSEMK17	5294	27.9	0.389	0.014	1.02
CSEMK17 (S1)			1.901	0.048	1.06
CSEMK17 (S2)			-0.167	0.068	1.03
CSEMK18	5294	36.2	0.157	0.018	1.10
CSEMK18 (S1)			0.897	0.040	0.93
CSMML01	5298	69.4	-1.226	0.033	1.04
CSMML02	5298	58.7	-0.689	0.030	0.93
CSMML03	5297	40.7	0.219	0.030	0.92
CSMML04	5297	45.0	-0.027	0.030	0.96
CSMML05	5298	41.1	0.107	0.030	0.92
CSMML06	5298	31.6	0.728	0.031	0.96
CSMML07	5297	30.7	0.637	0.031	1.02
CSMML08	5298	45.8	-0.075	0.030	1.02
CSMML09	5296	56.5	-0.340	0.030	1.01
CSMML10	5296	26.2	0.918	0.032	1.06
CSMML11	5298	74.1	-1.387	0.034	1.03
CSMML12	5298	63.9	-0.856	0.031	1.08
CSSML13	5294	25.6	0.987	0.033	0.95
CSSML14	5291	48.9	-0.132	0.030	0.88
CSSML15A	5297	48.4	-0.128	0.030	1.04
CSSML15B	5295	62.3	-0.757	0.031	1.03
CSEML16	5296	35.1	0.236	0.015	1.20
CSEML16 (S1)			0.295	0.030	1.05
CSEML16 (S2)			0.078	0.041	1.15
CSEML17	5296	17.7	0.892	0.021	1.04
CSEML17 (S1)			1.779	0.064	1.10
CSEML18	5292	49.8	-0.185	0.017	1.24
CSEML18 (S1)			1.755	0.057	1.08

Figure 7.1 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMJ04. Fit MNSQ=1.16

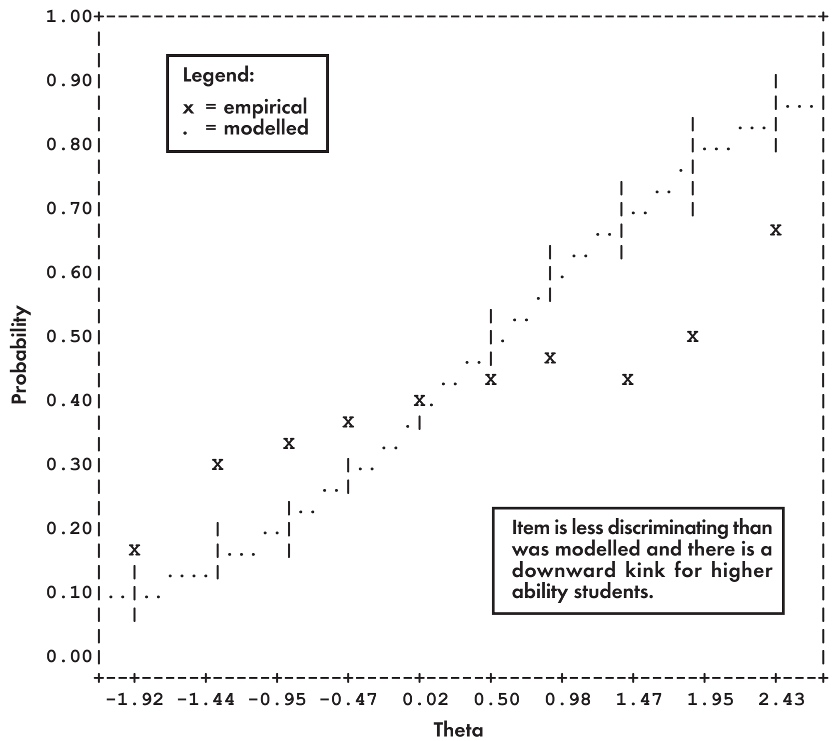


Figure 7.2 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMJ12. Fit MNSQ=1.38

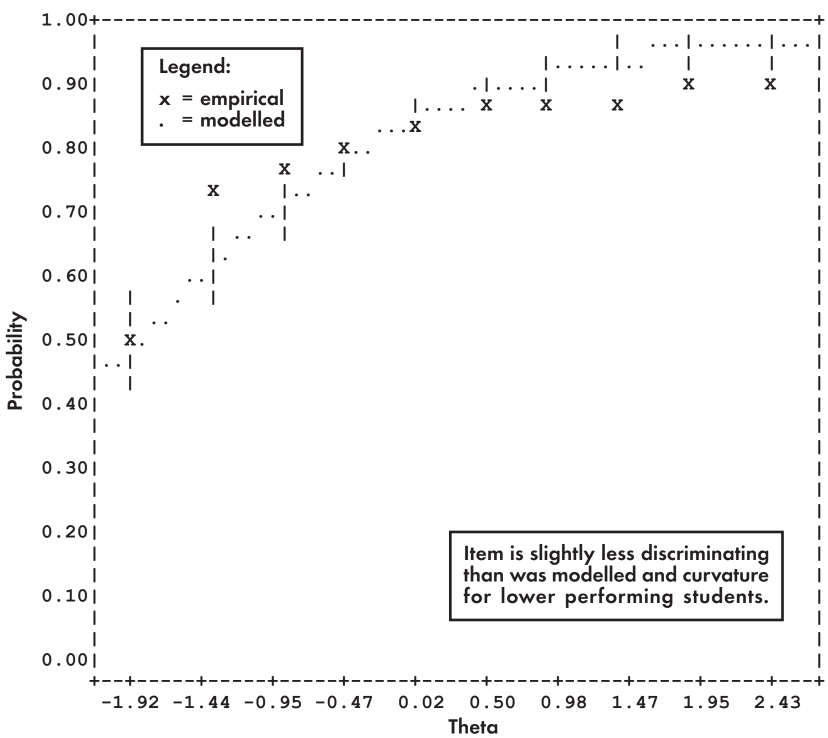
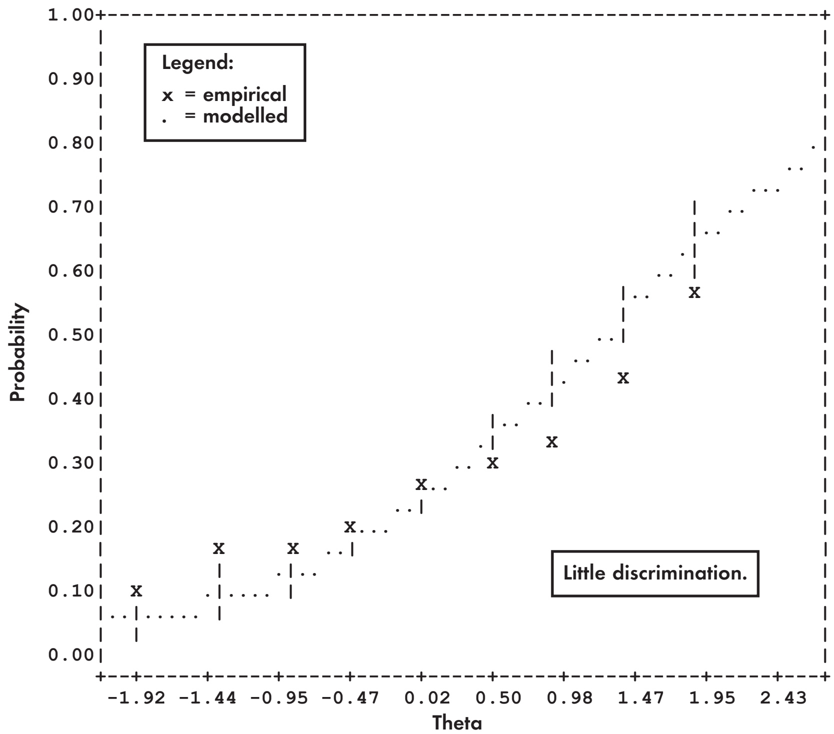


Figure 7.3 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMK08. Fit MNSQ=1.15



For the physics scale (Table 7.4) items with fit statistics greater than or equal to 1.15 are F16, F17B, G09, G11 and H18. There were also four items with fit statistics less than .85. Figure 7.4 shows that item F17B is a relatively hard item, although it is not clear why it is so difficult. For item G09, Figure 7.5 shows a dip among some of the better-performing students. Further investigation showed a positive point-biserial correlation with one of the distractors in most of the countries. Four items with fit less than .85 were more discriminating than average, especially among the higher-ability students.

Table 7.4 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Physics Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE01	13919	75.3	-1.773	0.020	1.10
CSMPE02	13921	58.9	-0.939	0.018	0.95
CSMPE03	13923	65.6	-1.408	0.019	0.95
CSMPE04	13925	84.4	-2.514	0.025	1.02
CSMPE05	13925	77.9	-2.034	0.022	1.03
CSMPE06	13925	34.6	0.083	0.019	1.07
CSMPE07	13924	45.8	-0.354	0.018	1.06
CSMPE08	13922	48.4	-0.469	0.018	1.07
CSMPE09	13919	35.0	0.083	0.019	1.00
CSMPE10	13917	45.0	-0.449	0.018	0.95
CSMPF01	4679	48.9	-0.523	0.031	1.12
CSMPF02	4679	16.8	1.097	0.039	1.11
CSMPF03	4679	39.0	-0.066	0.031	1.01
CSMPF04	4675	46.7	-0.592	0.031	0.93
CSMPF05	4675	60.3	-1.189	0.032	1.04
CSMPF06	4410	26.1	0.409	0.034	1.00
CSMPF07	4679	57.4	-0.808	0.031	1.09
CSMPF08	4679	43.3	-0.368	0.031	1.03
CSMPF09	4679	26.6	0.548	0.034	1.02
CSMPF10	4678	32.5	0.129	0.032	1.11
CSMPF11	4673	37.2	-0.050	0.031	0.95
CSSPF12	4670	15.9	0.666	0.024	0.95
CSSPF12 (S1)			1.037	0.052	1.06
CSSPF13	4671	61.6	-1.124	0.031	0.97
CSSPF14	4673	21.3	0.388	0.022	0.96
CSSPF14 (S1)			0.702	0.043	1.05
CSEPF15	4395	15.0	0.638	0.024	0.86
CSEPF15 (S1)			1.326	0.059	0.83
CSEPF16	4670	9.4	0.794	0.025	1.29
CSEPF16 (S1)			2.119	0.086	1.32
CSEPF17A	4669	25.7	0.297	0.033	1.04
CSEPF17B	4645	7.8	1.868	0.050	1.23
CSMPG01	4654	36.3	-0.123	0.032	1.11
CSMPG02	4654	65.3	-1.155	0.032	0.90
CSMPG03	4654	41.0	-0.177	0.031	1.17
CSMPG04	4654	32.7	0.148	0.032	1.09
CSMPG05	4651	37.0	0.021	0.032	0.98
CSMPG06	4652	60.1	-0.872	0.031	0.99
CSMPG07	4654	27.7	0.419	0.034	1.04
CSMPG08	4653	30.8	0.108	0.032	1.08
CSMPG09	4653	17.6	1.018	0.039	1.16
CSMPG10	4654	29.8	0.264	0.033	1.12

Table 7.4 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Physics Scale (Continued)

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSSPG11	4652	19.9	0.495	0.023	1.25
CSSPG11 (S1)			0.747	0.045	0.94
CSSPG12	4652	38.9	-0.304	0.019	0.99
CSSPG12 (S1)			0.898	0.042	1.00
CSSPG13	4652	30.2	0.087	0.032	0.95
CSSPG14	4652	24.2	0.613	0.035	0.83
CSSPG15	4652	15.6	1.292	0.042	0.80
CSSPG16	4444	35.2	0.171	0.028	1.14
CSSPG16 (S1)			-1.502	0.032	1.00
CSSPG17	4652	26.5	0.388	0.034	1.11
CSEPG18	4652	15.9	0.737	0.025	0.98
CSEPG18 (S1)			0.476	0.043	0.97
CSEPG19	4652	14.9	0.568	0.023	0.86
CSEPG19 (S1)			1.282	0.056	0.93
CSMPH01	4591	41.9	-0.086	0.032	1.12
CSMPH02	4592	52.2	-0.660	0.031	1.07
CSMPH03	4592	37.4	0.036	0.032	0.97
CSMPH04	4591	34.2	0.247	0.033	1.00
CSMPH05	4591	45.9	-0.309	0.031	1.03
CSMPH06	4591	31.4	0.394	0.034	1.00
CSMPH07	4591	35.2	-0.079	0.032	1.00
CSMPH08	4592	27.4	0.360	0.034	0.96
CSMPH09	4591	25.5	0.625	0.035	0.94
CSMPH10	4592	29.9	0.347	0.034	1.01
CSSPH12	4586	21.8	0.659	0.036	0.97
CSSPH13	4588	29.9	0.211	0.033	0.85
CSSPH14	4402	29.5	0.296	0.024	1.05
CSSPH14 (S1)			-0.478	0.033	1.01
CSSPH15	4588	25.6	0.705	0.036	0.86
CSSPH16	4585	20.6	0.322	0.021	1.04
CSSPH16 (S1)			1.711	0.064	0.83
CSEPH17	4587	15.1	0.535	0.023	1.13
CSEPH17 (S1)			1.978	0.076	1.00
CSEPH18	4587	24.7	0.378	0.023	1.17
CSEPH18 (S1)			-0.053	0.035	0.90
CSEPH19A	4412	28.0	0.258	0.022	0.93
CSEPH19A (S1)			0.308	0.039	0.93
CSEPH19B	4584	43.4	-0.276	0.032	0.92

Figure 7.4 Empirical and Modelled Item Characteristic Curves for Physics Population 3 Item: CSEPF17B. Fit MNSQ=1.23

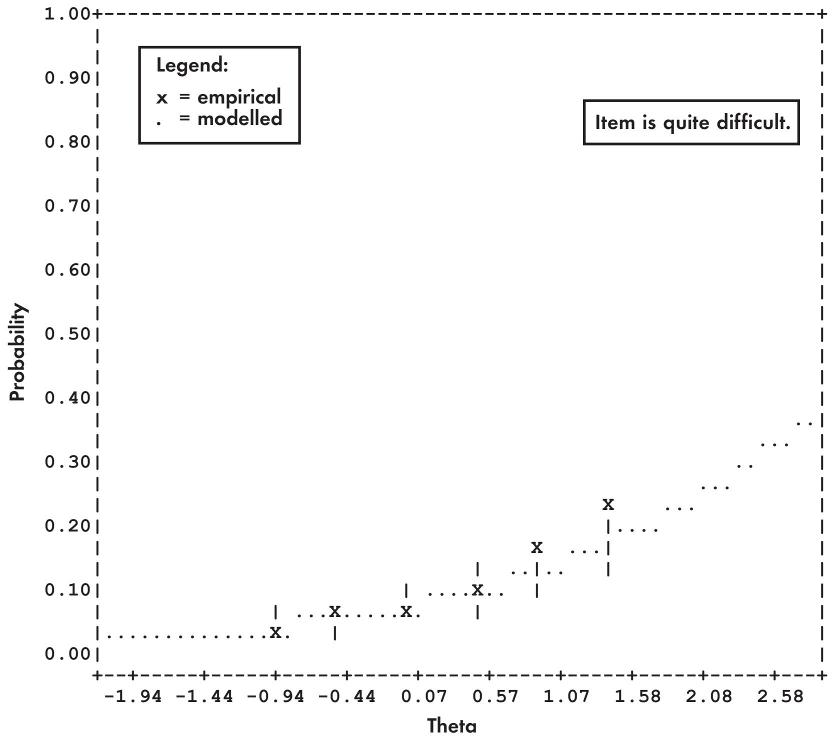
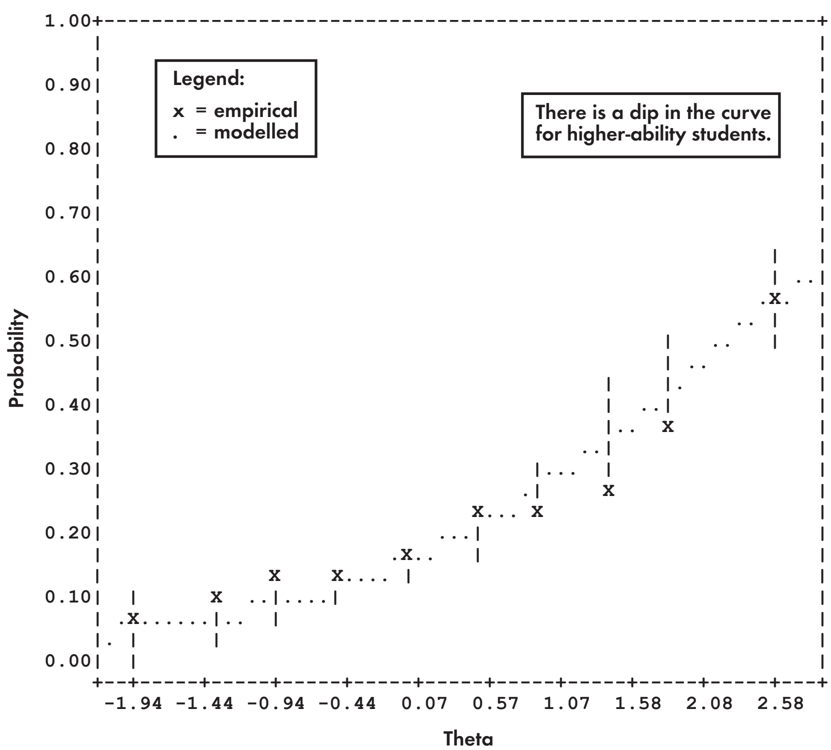


Figure 7.5 Empirical and Modelled Item Characteristic Curves for Physics Population 3 Item: CSMPG09. Fit MNSQ=1.16



For the mathematics literacy scale (Table 7.5), items with fit statistics greater than or equal to 1.15 are A04, A12, B20, B21 and B24. Figure 7.6, illustrating item A04, shows a tailing off of observed results compared with the modelled, for higher-ability students. Item B20, shown in Figure 7.7, exhibits a marked lack of discrimination, similar to J04 on the advanced mathematics scale. Both B21 and B24 in Figures 7.8 and 7.9 show similar behavior but the effect is smaller. Item A12, a partial credit item, consistently shows a low response in category 2 compared to categories 1 and 3 across the different countries. There are five items with fit statistics below .85. All of these show greater than average discrimination, but this sort of misfit was not deemed to be of concern.

Table 7.5 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Mathematics Literacy Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA03	49191	63.8	-0.398	0.010	0.95
CSMGA04	49191	70.5	-0.797	0.011	1.18
CSMGA05	49188	48.8	0.344	0.010	0.99
CSSGA08	49170	50.8	0.254	0.006	1.04
CSSGA08 (S1)			0.538	0.012	1.09
CSEGA10	49182	33.1	1.144	0.007	1.09
CSEGA10 (S1)			-0.295	0.010	0.99
CSEGA12	49178	55.5	0.065	0.005	1.36
CSEGA12 (S1)			0.564	0.012	1.00
CSEGA12 (S2)			0.787	0.019	0.89
CSMGB14	49194	71.5	-0.859	0.011	1.05
CSMGB15	49199	58.5	-0.158	0.010	0.98
CSMGB16	49198	78.2	-1.344	0.012	1.00
CSMGB17	49197	47.4	0.384	0.010	0.86
CSMGB18	49189	37.0	0.938	0.010	1.02
CSMGB19	49196	71.6	-0.951	0.011	0.91
CSMGB20	49194	50.8	0.340	0.010	1.20
CSMGB21	44107	36.1	1.009	0.011	1.21
CSMGB22	49196	69.6	-0.747	0.011	0.95
CSMGB23	49194	53.1	0.164	0.010	1.05
CSMGB24	49193	42.0	0.725	0.010	1.17
CSSGB25	49196	36.2	0.947	0.010	0.98
CSSGB26	44094	39.7	0.629	0.006	1.09
CSSGB26 (S1)			1.816	0.021	1.12
CSMGC01	24789	68.1	-0.637	0.015	0.86
CSMGC02	24784	69.0	-0.716	0.015	0.87
CSMGC03	24781	60.5	-0.280	0.015	0.85
CSMGC04	24784	66.5	-0.563	0.015	0.92
CSMGC05	24535	70.4	-0.795	0.016	0.96
CSMGD13	24407	62.6	-0.399	0.015	0.89
CSMGD14	24404	63.9	-0.512	0.015	0.92
CSSGD15A	24405	73.7	-0.987	0.016	0.99
CSSGD15B	21915	59.0	-0.130	0.015	0.96
CSSGD16A	21916	39.2	0.841	0.015	0.93
CSSGD16B	21871	32.4	1.221	0.016	0.94
CSSGD17	24381	30.8	1.166	0.010	1.00
CSSGD17 (S1)			-0.178	0.015	1.04
CSMGC06	24790	79.4	-1.593	0.018	0.84
CSMGC07	24792	51.0	0.185	0.014	0.83
CSMGC08	24784	65.6	-0.516	0.015	0.92
CSMGC09	24790	70.7	-0.706	0.015	1.03
CSMGC11	24784	65.5	-0.613	0.015	0.81
CSSGC12	24783	24.1	1.742	0.016	0.97
CSSGC13	24788	21.5	1.909	0.017	0.90
CSMGD06	24402	63.4	-0.419	0.015	0.93
CSMGD07	24406	69.8	-0.818	0.015	0.98
CSMGD08	24406	53.6	0.066	0.014	0.93
CSMGD09	24407	69.6	-0.826	0.016	0.82
CSMGD10	24407	58.9	-0.199	0.014	0.96
CSMGD11	24406	43.4	0.553	0.014	0.88
CSMGD12	24184	28.8	1.338	0.016	0.84

Figure 7.6 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGA04 Fit MNSQ=1.18

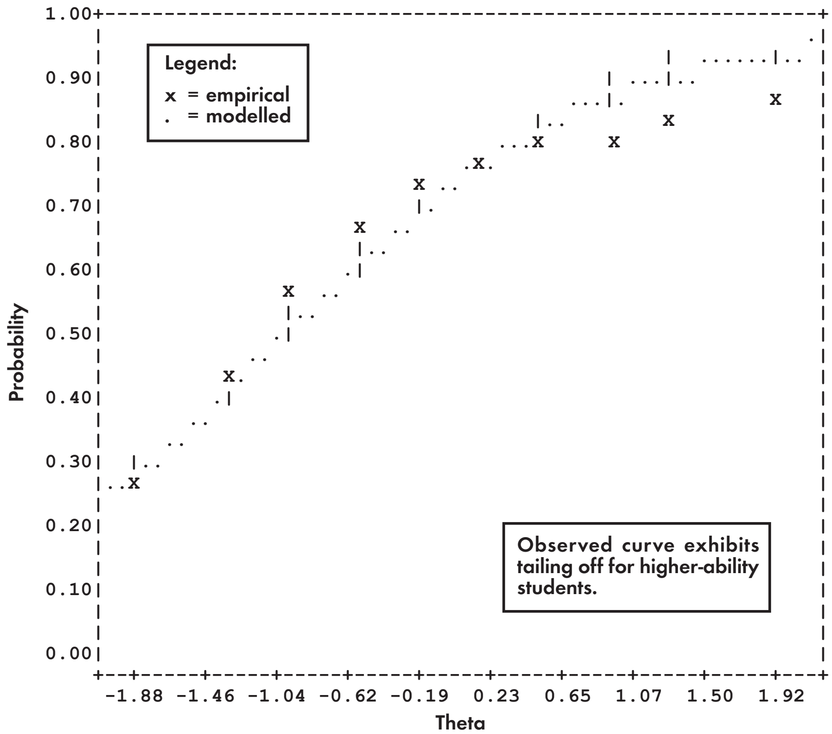


Figure 7.7 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB20. Fit MNSQ=1.2

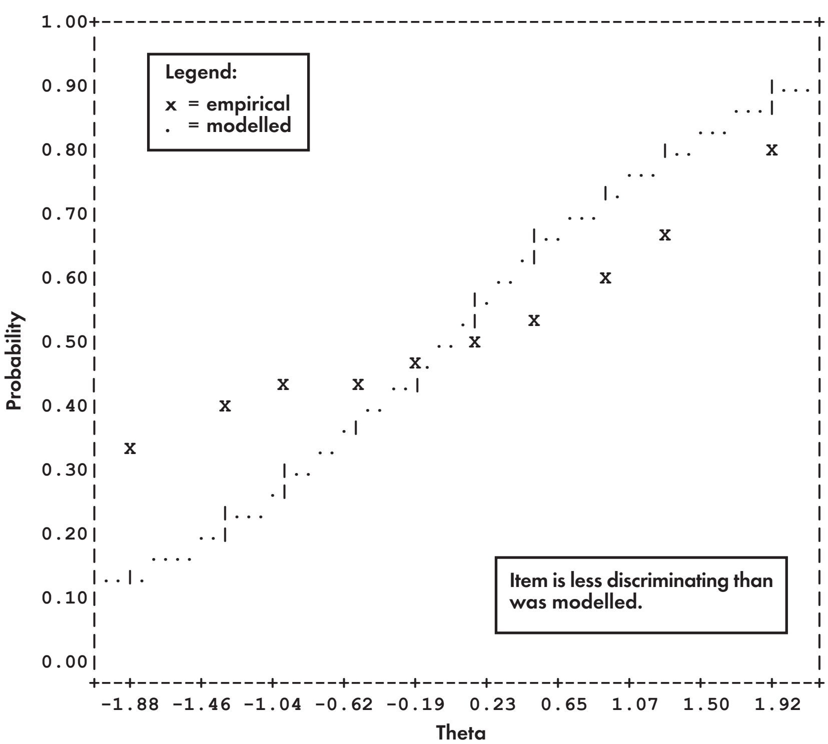


Figure 7.8 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB21. Fit MNSQ=1.21

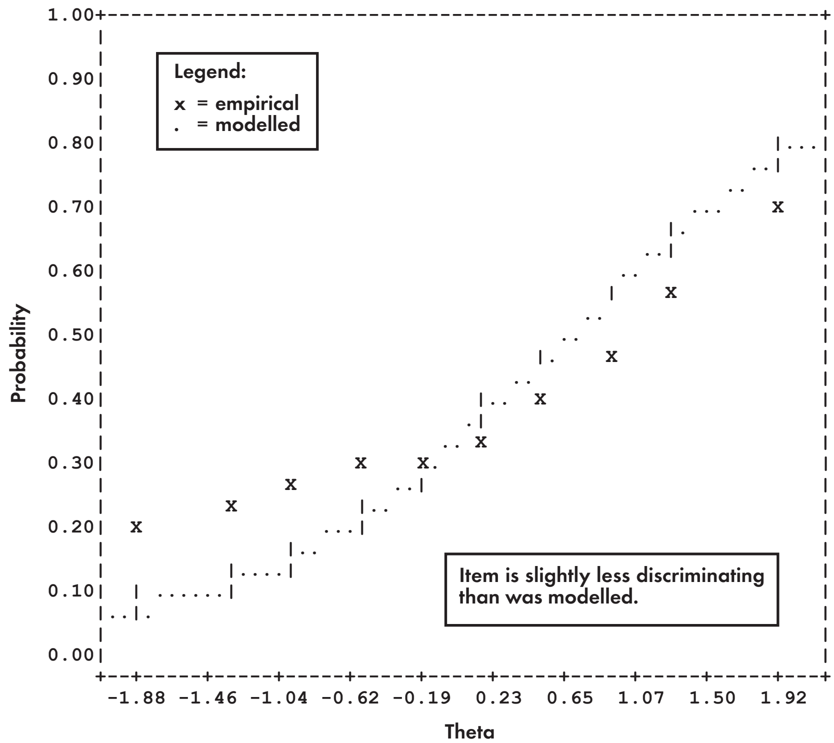
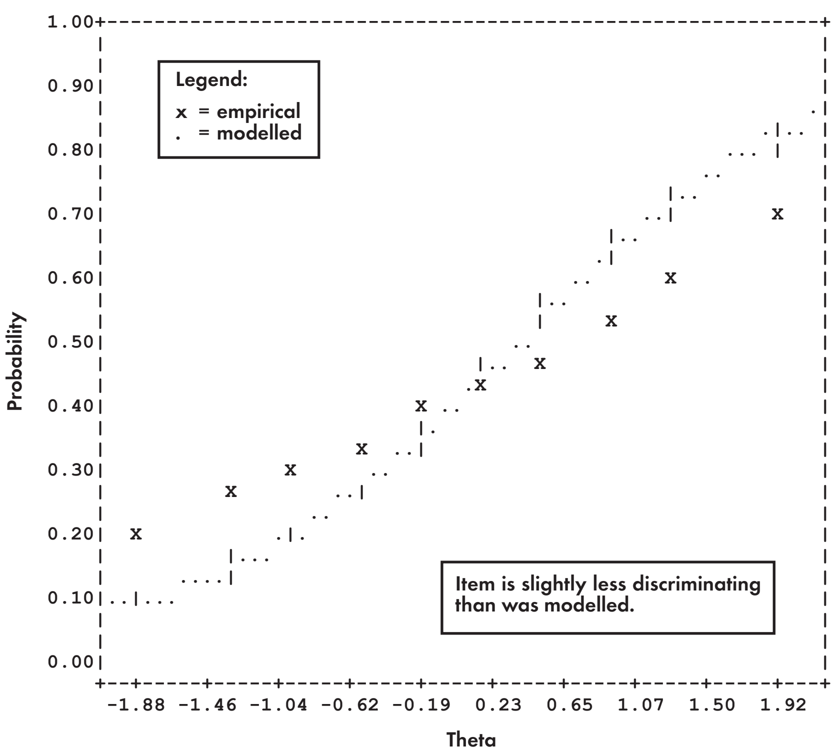


Figure 7.9 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB24. Fit MNSQ=1.17



For the science literacy scale (Table 7.6), items with fit statistics greater than or equal to 1.15 are B02, C20 and C21. Item B02 in Figure 7.10 shows evidence of less than usual discrimination. Both C20 and C21 also show slightly less than usual discrimination. There are no items with fit statistics less than .85 on this scale. This scale seemed to fit slightly better than the others.

Tables 7.7 through 7.14 present statistics for the reasoning and social utility (RSU) subscale, the three advanced mathematics subscales, and the five physics subscales. The fit statistics of the items on the subscales are very similar to the fit statistics for the overall scales, as would be expected.

Table 7.6 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Science Literacy Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA01	49194	39.9	0.844	0.010	1.14
CSMGA02	49195	74.9	-1.060	0.011	1.00
CSSGA06A	49191	36.8	0.954	0.010	0.96
CSSGA06B	49174	43.0	0.606	0.010	0.95
CSSGA07	49186	49.4	0.296	0.006	1.04
CSSGA07 (S1)			0.404	0.011	1.04
CSEGA09B	49180	32.4	1.233	0.010	1.00
CSEGA11A	49191	72.2	-0.880	0.011	0.93
CSEGA11B	49187	60.1	-0.314	0.010	0.87
CSEGA11C	46101	43.1	0.545	0.010	0.97
CSMGB01	49196	65.0	-0.511	0.010	1.05
CSMGB02	49197	85.0	-1.625	0.013	1.19
CSMGB03	49200	60.6	-0.112	0.010	1.08
CSMGB04	47607	54.6	-0.011	0.010	1.07
CSMGB05	49198	62.6	-0.311	0.010	1.08
CSMGB06	44629	31.7	1.257	0.011	0.99
CSMGB07	49198	91.3	-2.347	0.016	1.05
CSMGB08	44109	71.4	-0.704	0.011	1.10
CSMGB09	49196	30.3	1.261	0.011	1.10
CSMGB10	49196	49.7	0.281	0.010	0.96
CSMGB11	49195	54.1	0.154	0.010	1.03
CSSGB12	49195	32.7	1.143	0.010	0.89
CSSGB13	49193	81.6	-1.451	0.012	1.03
CSMGC14	24784	66.5	-0.455	0.015	1.00
CSMGC15	24787	73.3	-0.883	0.016	0.96
CSMGC16	24787	77.3	-1.160	0.016	0.89
CSMGC17	24781	57.3	-0.041	0.014	0.93
CSSGC18	24789	32.1	1.226	0.015	0.85
CSSGC19	24788	42.0	0.590	0.008	0.96
CSSGC19 (S1)			1.210	0.021	1.07
CSSGC20	22185	30.1	0.988	0.009	1.21
CSSGC20 (S1)			1.516	0.026	0.95
CSEGC21	24781	25.4	1.515	0.011	1.15
CSEGC21 (S1)			-0.473	0.014	1.04
CSMGD01	24405	85.3	-1.803	0.020	1.02
CSSGD02	24403	57.0	-0.005	0.014	1.01
CSSGD03	24405	66.8	-0.504	0.015	0.97
CSEGD04	24397	20.2	1.860	0.017	0.87
CSEGD05A	24407	72.4	-0.838	0.015	0.98
CSEGD05B	24407	53.1	0.263	0.014	0.99

Figure 7.10 Empirical and Modelled Item Characteristic Curves for Science Literacy Population 3 Item: CSMGB02. Fit MNSQ=1.19

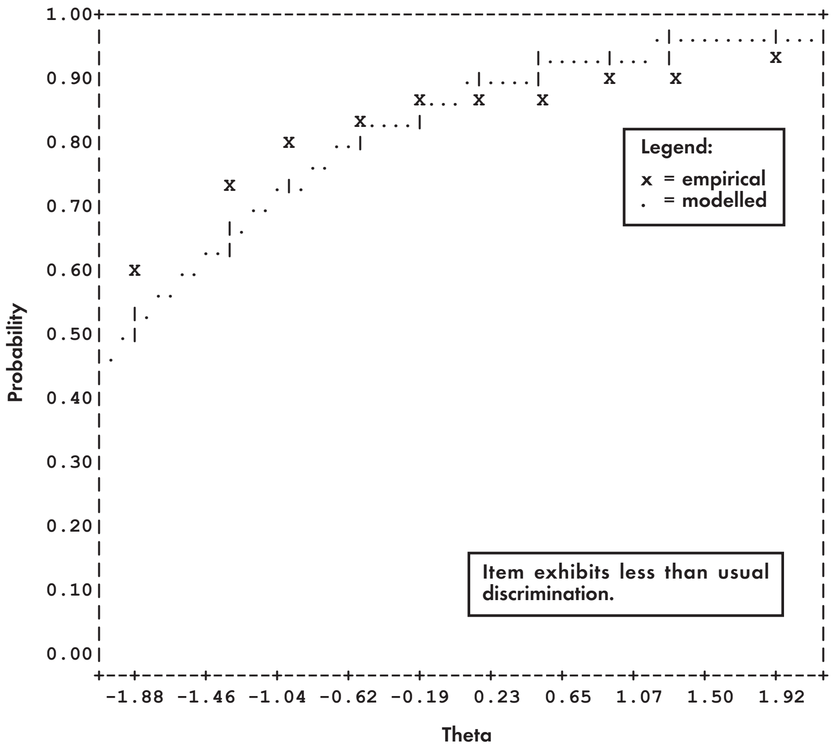


Table 7.7 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Reasoning and Social Utility Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA01	49194	39.9	0.659	0.010	1.13
CSMGA02	49195	74.9	-1.293	0.012	1.01
CSMGA03	49191	63.8	-0.556	0.010	0.95
CSMGA04	49191	70.5	-0.945	0.011	1.12
CSMGA05	49188	48.8	0.168	0.010	1.00
CSSGA06A	49191	36.8	0.775	0.010	0.96
CSSGA06B	49174	43.0	0.415	0.010	0.95
CSSGA07	49186	49.4	0.103	0.006	1.03
CSSGA07 (S1)			0.367	0.011	1.02
CSSGA08	49170	50.8	0.083	0.006	0.93
CSSGA08 (S1)			0.582	0.012	1.07
CSEGA09B	49180	32.4	1.048	0.011	0.96
CSEGA10	49182	33.1	0.917	0.007	1.03
CSEGA10 (S1)			-0.234	0.010	0.97
CSEGA11A	49191	72.2	-1.118	0.011	0.96
CSEGA11B	49187	60.1	-0.532	0.010	0.87
CSEGA11C	46101	43.1	0.363	0.010	0.95
CSEGA12	49178	55.5	-0.802	0.005	1.14
CSEGA12 (S1)			0.610	0.012	0.95
CSEGA12 (S2)			0.805	0.019	0.87

Table 7.8 Parameter Estimates for the International Calibration Sample Population 3 Numbers and Equations Scale:

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI01	-0.430	0.018	0.91
CSMMI02	-0.499	0.018	0.92
CSMMI03	-0.516	0.018	1.12
CSMMJ01	-0.406	0.030	0.88
CSMMJ02	0.627	0.034	1.01
CSMMJ03	-0.472	0.030	0.97
CSMMJ04	0.671	0.030	1.14
CSMMK01	-1.897	0.040	1.09
CSMMK02	1.205	0.033	1.00
CSSMK13	1.256	0.034	0.94
CSSMK15	1.137	0.021	1.04
CSSMK15 (S1)	1.874	0.068	0.89
CSEMK16	-0.019	0.015	1.37
CSEMK16 (S1)	0.485	0.029	0.97
CSEMK16 (S2)	-0.650	0.033	0.94
CSMML01	-1.101	0.033	1.01
CSMML02	-0.543	0.031	0.91
CSMML03	0.399	0.030	0.93
CSMML04	0.146	0.030	1.02
CSEML16	0.441	0.015	1.14
CSEML16 (S1)	0.220	0.031	1.05
CSEML16 (S2)	0.074	0.041	1.16

Table 7.9 Parameter Estimates for the International Calibration Sample Population 3 Calculus Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI04	-0.878	0.018	0.98
CSMMI06	-0.452	0.018	0.96
CSMMJ05	-0.966	0.031	0.87
CSMMJ06	0.399	0.031	1.03
CSMMJ14	-0.573	0.030	0.90
CSSMJ15A	-0.375	0.030	0.93
CSSMJ15B	2.600	0.053	0.92
CSSMJ17	0.373	0.020	1.14
CSSMJ17 (S1)	1.290	0.052	1.13
CSMMK03	-1.123	0.032	1.11
CSMMK04	0.722	0.033	1.01
CSMMK05	-0.109	0.030	1.13
CSMMK06	-0.692	0.030	1.05
CSEMK17	0.195	0.015	0.97
CSEMK17 (S1)	1.802	0.048	1.05
CSEMK17 (S2)	-0.165	0.068	1.00
CSMML05	-0.118	0.031	0.93
CSMML06	0.546	0.032	1.01
CSMML07	0.452	0.032	1.04

Table 7.10 Parameter Estimates for the International Calibration Sample Population 3 Geometry Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI07	-0.551	0.017	0.99
CSMMI08	-1.618	0.020	0.95
CSMMI09	-0.796	0.017	0.99
CSMMJ07	0.041	0.029	0.94
CSMMJ08	-0.933	0.032	1.04
CSMMJ09	1.095	0.033	1.05
CSMMJ10	0.377	0.030	1.04
CSMMJ11	-1.120	0.031	1.12
CSSMJ16A	-1.034	0.031	1.00
CSSMJ16B	1.190	0.034	0.88
CSEMJ19	0.173	0.017	0.99
CSEMJ19 (S1)	1.295	0.046	0.98
CSMMK07	-0.389	0.029	1.04
CSMMK08	0.837	0.032	1.09
CSMMK09	0.172	0.031	1.04
CSMMK10	1.264	0.035	0.92
CSSMK12	-0.199	0.029	1.03
CSSMK14	1.297	0.026	0.92
CSSMK14 (S1)	2.687	0.111	0.84
CSEMK18	0.082	0.018	0.96
CSEMK18 (S1)	0.922	0.040	0.91
CSMML08	-0.124	0.029	0.99
CSMML09	-0.385	0.030	1.00
CSMML12	-0.895	0.031	1.09
CSSML13	0.932	0.033	0.91
CSEML17	0.823	0.020	1.03
CSEML17 (S1)	1.788	0.064	1.09
CSEML18	-0.237	0.017	1.12
CSEML18 (S1)	1.771	0.057	1.06

Table 7.11 Parameter Estimates for the International Calibration Sample Population 3 Mechanics Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE03	-1.618	0.019	1.01
CSMPE05	-2.264	0.022	1.06
CSMPF02	0.979	0.040	0.99
CSMPF04	-0.775	0.031	0.93
CSMPF10	-0.026	0.033	1.12
CSEPF17A	0.146	0.033	0.97
CSEPF17B	1.761	0.051	1.12
CSMPG07	0.270	0.035	1.02
CSMPG08	-0.051	0.033	1.06
CSMPG09	0.891	0.039	1.12
CSSPG12	-0.466	0.019	1.01
CSSPG12 (S1)	0.836	0.042	1.00
CSSPG15	1.176	0.043	0.78
CSSPG16	0.029	0.028	1.18
CSSPG16 (S1)	-1.583	0.032	1.05
CSMPH01	-0.237	0.033	1.23
CSMPH04	0.112	0.034	0.96
CSSPH13	0.073	0.034	0.92

Table 7.12 Parameter Estimates for the International Calibration Sample Population 3 Electricity and Magnetism Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE04	-2.639	0.025	1.05
CSMPE06	-0.034	0.019	1.05
CSMPE09	-0.035	0.019	1.00
CSMPF06	0.296	0.035	0.96
CSMPF08	-0.487	0.031	1.01
CSSPF14	0.276	0.022	0.94
CSSPF14 (S1)	0.692	0.043	1.04
CSEPF16	0.687	0.026	1.20
CSEPF16 (S1)	2.107	0.086	1.30
CSMPG01	-0.248	0.032	1.09
CSMPG04	0.024	0.033	1.09
CSSPG17	0.269	0.034	1.10
CSEPG19	0.453	0.023	0.88
CSEPG19 (S1)	1.272	0.056	0.90
CSMPH06	0.290	0.034	1.04
CSMPH08	0.255	0.034	0.96
CSMPH10	0.242	0.034	0.95
CSSPH16	0.217	0.021	0.85
CSSPH16 (S1)	1.699	0.064	0.81
CSEPH17	0.434	0.023	1.09
CSEPH17 (S1)	1.965	0.076	0.98

Table 7.13 Parameter Estimates for the International Calibration Sample Population 3 Heat Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE08	-0.195	0.018	1.03
CSMPF05	-0.906	0.031	1.02
CSSPF12	0.873	0.023	0.87
CSSPF12 (S1)	1.092	0.052	1.05
CSMPG02	-0.882	0.032	0.91
CSMPG03	0.073	0.031	1.15
CSSPG11	0.708	0.022	1.09
CSSPG11 (S1)	0.799	0.045	0.94
CSMPH02	-0.386	0.031	1.06
CSMPH07	0.180	0.032	0.96
CSSPH14	0.535	0.023	1.02
CSSPH14 (S1)	-0.423	0.033	1.01

Table 7.14 Parameter Estimates for the International Calibration Sample Population 3 Wave Phenomena Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE01	-1.613	0.021	1.13
CSMPE10	-0.238	0.018	0.97
CSMPF01	-0.312	0.031	1.11
CSMPF11	0.179	0.032	0.95
CSSPF13	-0.938	0.032	1.02
CSMPG05	0.253	0.033	1.09
CSSPG13	0.322	0.033	1.03
CSMPH09	0.905	0.036	1.01
CSSPH12	0.940	0.036	0.96
CSEPH19A	0.546	0.023	0.98
CSEPH19A (S1)	0.221	0.039	0.93
CSEPH19B	-0.042	0.032	0.90

Table 7.15 Parameter Estimates for the International Calibration Sample Population 3 Particle, Quantum, Astrophysics, and Relativity Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE02	-0.993	0.018	0.97
CSMPE07	-0.401	0.018	1.03
CSMPF03	-0.111	0.031	0.99
CSMPF07	-0.861	0.031	1.10
CSMPF09	0.508	0.034	1.00
CSEPF15	0.612	0.024	0.90
CSEPF15 (S1)	1.307	0.060	0.82
CSMPG06	-0.934	0.032	1.03
CSMPG10	0.217	0.033	1.15
CSSPG14	0.570	0.035	0.84
CSEPG18	0.716	0.025	1.00
CSEPG18 (S1)	0.451	0.043	0.97
CSMPH03	-0.003	0.033	0.99
CSMPH05	-0.350	0.032	1.07
CSSPH15	0.677	0.036	0.87
CSEPH18	0.354	0.023	1.18
CSEPH18 (S1)	-0.077	0.035	0.91

7.4.4 The Population Model For Population 3

The population model equation (9) specifies that the latent variable θ has a distribution that is partly a function of a range of background variables. In order to derive reliable proficiency estimates, therefore, it is necessary to condition on these background variables before drawing the plausible values. A large set of background variables was used in the conditioning, including all of the questions from the student questionnaire. The information in these student variables was summarized through a principal components analysis in order to avoid multicollinearity problems and to keep the number of variables in the conditioning to a manageable level. A principal component analysis was run for each country on all students and as many components retained as explained 90 percent of the variance. Table 7.16 shows the number of components for each country. For the principal components analysis each student variable was recoded into a set of dummy variables which represented all categories of the variable as well as a missing data indicator.

For all scaling runs the variable sex was used as a conditioning variable. Additionally, preliminary national scores in mathematics and science literacy, reasoning and social utility (RSU), advanced mathematics, and physics were computed for each country using basic Rasch scaling methodology. These national scores were used in the conditioning process. As may be seen from Table 7.17, conditioning for the mathematics and science literacy scales included sex of student, the advanced mathematics national score, the physics national score, the school mean on the mathematics and science literacy national score (mathematics and science literacy and RSU combined), the principal components of the questionnaire variables, and the product of the mathematics and science literacy school mean and the principal components. Conditioning for the RSU scale was very similar, except that the RSU national score was substituted for the mathematics and science literacy national score. For advanced mathematics, the sex of student, the physics national score, the mathematics and science literacy national score (excluding RSU), the school mean on the advanced mathematics national score, the principal components, and the product of the school mean on the advanced mathematics score and the principal components. The physics conditioning was similar, and included the sex of student, the advanced mathematics national score, the mathematics and science literacy national score (excluding RSU), the school mean on the physics national score, the principal components, and the product of the physics score and the principal components.

Table 7.16 Number of Principal Components Retained in Conditioning - Population 3

Country	Retained Components
Australia	66
Austria	84
Canada	81
Cyprus	103
Czech Republic	90
Denmark	81
France	68
Germany	60
Greece	74
Hungary	103
Iceland	70
Israel	87
Italy	96
Latvia	82
Lithuania	91
Netherlands	60
New Zealand	78
Norway	81
Russia	113
Slovenia	90
South Africa	128
Sweden	79
Switzerland	91
United States	71

Table 7.17 Variables Used in Conditioning - Population 3

Variables	Mathematics and Science Literacy	RSU	Mathematics	Physics
Sex	Y	Y	Y	Y
Advanced Mathematics Score	Y	Y	N	Y
Physics Score	Y	Y	Y	N
Mathematics and Science Literacy	N	N	Y	Y
RSU Score	N	N	N	N
School Mean Advanced Mathematics	N	N	Y	N
School Mean Physics Score	N	N	N	Y
School Mean Mathematics and Science Literacy/RSU Score	Y	N	N	N
School Mean RSU Score	N	Y	N	N
Principal Components	Y	Y	Y	Y
Principal Components by School Mean Advanced Mathematics Score	N	N	Y	N
Principal Components by School Mean Physics Score	N	N	N	Y
Principal Components by School Mean Mathematics and Science Literacy	Y	N	N	N
Principal Components by School Mean RSU Score	N	Y	N	N

REFERENCES

Adams, R.J., Wilson, M.R., and Wang, W.C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement, 21,* 1-24.

Adams, R.J., Wilson, M.R., and Wu, M.L. (1997). Multilevel item responses models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22 (1),* 46-75.

Adams, R.J., Wu, M.L., and Macaskill G. (1997). Scaling methodology and procedures for the mathematics and science scales. In M.O. Martin and D.L. Kelly (Eds.), *TIMSS technical report, volume II: Implementation and analysis – primary and middle school years.* Chestnut Hill, MA: Boston College.

Gonzalez, E. J., Smith, T. A., and Sibberns, H. (Eds.). (1998). *User guide for the TIMSS international database: Final year of secondary school – 1995 assessment*. Chestnut Hill, MA.: Boston College.

Wu, M.L. (1997). *The development and application of a fit test for use with marginal estimation and generalized item response models.* Unpublished Master's dissertation, University of Melbourne.

Wu, M.L., Adams, R.J., and Wilson, M. (1997). *Conquest: Generalized item response modelling software - Manual.* Melbourne: Australian Council for Educational Research.