Appendix A

An Expectation-Maximization Algorithm for the SISM Model

By referring to the derivation of formulas for the expectation-maximization algorithm illustrated in de la Torre (2009), the probability of correctly answering item j can be reexpressed as

$$P_j(\alpha_l)=\begin{cases} h_j & \text{if } \eta_{\mathrm{j}}=1, \gamma_j=0 \\ \omega_j & \text{if } \eta_{\mathrm{j}}=1, \gamma_j=1 \\ g_j & \text{if } \eta_{\mathrm{j}}=0, \gamma_j=0 \\ \varepsilon_j & \text{if } \eta_{\mathrm{j}}=0, \gamma_j=1 \end{cases}. \tag{A1}$$

According to the de la Torre (2009), $\partial l(X)/\partial\beta_{j(\eta,\gamma)}$ can be defined as

$$\frac{\partial l(X)}{\partial\beta_{j(\eta,\gamma)}}=\sum_{l=1}^{L}\frac{\partial P_j(\alpha_l)}{\partial\beta_{j(\eta,\gamma)}}\left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)]}\right][R_{jl}-P_j(\alpha_l)I_l], \tag{A2}$$

where $\beta_{j(\eta,\gamma)}=h_j$ if $\eta=1$ and $\gamma=0$; $\beta_{j(\eta,\gamma)}=\omega_j$ if $\eta=1$ and $\gamma=1$; $\beta_{j(\eta,\gamma)}=g_j$ if $\eta=0$ and $\gamma=0$; and $\beta_{j(\eta,\gamma)}=\varepsilon_j$ if $\eta=0$ and $\gamma=1$. $I_l=\sum_{i=1}^{I}p(\alpha_l|\mathbf{X}_i)$ is the expected number of students whose attribute pattern is α_l, and $R_{jl}=\sum_{i=1}^{I}p(\alpha_l|\mathbf{X}_i)X_{ij}$ is the expected number of students who have the attribute pattern α_l and answer item *j* correctly. $p(\alpha_l|\mathbf{X}_i)=L(\mathbf{X}_i|\alpha_l)p(\alpha_l)$ is the posterior probability of the attribute pattern α_l, and $L(\mathbf{X}_i|\alpha_l)$ and $p(\alpha_l)$ are the likelihood function of $\mathbf{X}_i$ conditioned on α_l and the prior probability, respectively.

To obtain estimators of parameters, Equation A2 can be written as

$$\frac{\partial l(\mathbf{X})}{\partial\beta_{j(\eta,\gamma)}}=\sum_{\{\alpha_l:\eta_{lj}=1,\gamma_{lj}=0\}}\frac{\partial P_j(\alpha_l)}{\partial\beta_{j(\eta,\gamma)}}\left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)]}\right]\left[R_{jl}-P_j(\alpha_l)I_l\right]+$$

$$\sum_{\{\alpha_l:\eta_{lj}=1,\gamma_{lj}=1\}}\frac{\partial P_j(\alpha_l)}{\partial\beta_{j(\eta,\gamma)}}\left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)]}\right]\left[R_{jl}-P_j(\alpha_l)I_l\right]+$$

$$\sum_{\{\alpha_l:\eta_{lj}=0,\gamma_{lj}=0\}} \frac{\partial P_j(\alpha_l)}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)]}\right] \left[R_{jl} - P_j(\alpha_l) I_l\right] +$$

$$\sum_{\{\alpha_l:\eta_{lj}=0,\gamma_{lj}=1\}} \frac{\partial P_j(\alpha_l)}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)]}\right] \left[R_{jl} - P_j(\alpha_l) I_l\right]$$

$$= \frac{\partial h_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{h_j[1-h_j]}\right] \sum_{\{\alpha_l:\eta_{lj}=1,\gamma_{lj}=0\}} \left[R_{jl} - h_j I_l\right] +$$

$$\frac{\partial \omega_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{\omega_j[1-\omega_j]}\right] \sum_{\{\alpha_l:\eta_{lj}=1,\gamma_{lj}=1\}} \left[R_{jl} - \omega_j I_l\right] +$$

$$\frac{\partial g_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{g_j[1-g_j]}\right] \sum_{\{\alpha_l:\eta_{lj}=0,\gamma_{lj}=0\}} \left[R_{jl} - g_j I_l\right] +$$

$$\frac{\partial \varepsilon_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{\varepsilon_j[1-\varepsilon_j]}\right] \sum_{\{\alpha_l:\eta_{lj}=0,\gamma_{lj}=1\}} \left[R_{jl} - \varepsilon_j I_l\right] +$$

$$= \frac{\partial h_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{h_j[1-h_j]}\right] \left[R_{jl}^{\eta_{lj}=1,\gamma_{lj}=0} - h_j I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}\right] +$$

$$\frac{\partial \omega_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{\omega_j[1-\omega_j]}\right] \left[R_{jl}^{\eta_{lj}=1,\gamma_{lj}=1} - \omega_j I_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}\right] +$$

$$\frac{\partial g_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{g_j[1-g_j]}\right] \left[R_{jl}^{\eta_{lj}=0,\gamma_{lj}=0} - g_j I_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}\right] +$$

$$\frac{\partial \varepsilon_j}{\partial \beta_{j(\eta,\gamma)}} \left[\frac{1}{\varepsilon_j[1-\varepsilon_j]}\right] \left[R_{jl}^{\eta_{lj}=0,\gamma_{lj}=1} - \varepsilon_j I_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}\right], \tag{A3}$$

where $I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}$ is the expected number of students whose $\eta_{lj}=1$ and $\gamma_{lj}=0$; $I_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}$ is the expected number of students whose $\eta_{lj}=1$ and $\gamma_{lj}=1$; $I_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}$ is the expected number of students whose $\eta_{lj}=0$ and $\gamma_{lj}=0$; and $I_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}$ is the expected number of students whose $\eta_{lj}=0$ and $\gamma_{lj}=1$. The parameters $R_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}$, $R_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}$, $R_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}$, and $R_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}$ respectively represent the expected number of students among $I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}$, $I_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}$, $I_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}$, and $I_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}$ who correctly answer item *j*.

When $\eta_{lj}=1$ and $\gamma_{lj}=0$, $\beta_{j(\eta,\gamma)}=h_j$, the first term $\partial h_j/\partial\beta_{j(\eta,\gamma)}$ of Equation A3 is 1 and the other terms are 0s. Thus, maximization $\partial l(X)/\partial\beta_{j(\eta,\gamma)}$ simplifies to solving for h_j in the equation $\left[\frac{1}{h_j[1-h_j]}\right][R_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}-h_j I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}]=0$, which gives the estimator of $\hat{h}_j=R_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}/I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}$. Similarly, the estimators of $\hat{\omega}_j$, $\hat{g}_j$, and $\hat{\varepsilon}_j$ are $R_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}/I_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}$, $R_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}/I_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}$, and $R_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}/I_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}$.

Because $I_l=\sum_{i=1}^{I}p(\alpha_l|X_i)$ and $R_{jl}=\sum_{i=1}^{I}p(\alpha_l|X_i)X_{ij}$, the formulas of $\hat{h}_j$, $\hat{\omega}_j$, $\hat{g}_j$, and $\hat{\varepsilon}_j$ can be written as

$$\hat{h}_j=\frac{R_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}}{I_{jl}^{\eta_{lj}=1,\gamma_{lj}=0}}=\frac{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)X_{ij}\eta_{lj}(1-\gamma_{lj})}{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)\eta_{lj}(1-\gamma_{lj})}, \tag{A4}$$

$$\hat{\omega}_j=\frac{R_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}}{I_{jl}^{\eta_{lj}=1,\gamma_{lj}=1}}=\frac{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)X_{ij}\eta_{lj}\gamma_{lj}}{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)\eta_{lj}\gamma_{lj}}, \tag{A5}$$

$$\hat{g}_j=\frac{R_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}}{I_{jl}^{\eta_{lj}=0,\gamma_{lj}=0}}=\frac{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)X_{ij}(1-\eta_{lj})(1-\gamma_{lj})}{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)(1-\eta_{lj})(1-\gamma_{lj})}, \tag{A6}$$

$$\hat{\varepsilon}_j=\frac{R_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}}{I_{jl}^{\eta_{lj}=0,\gamma_{lj}=1}}=\frac{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)X_{ij}(1-\eta_{lj})\gamma_{lj}}{\sum_{l=1}^{L}\sum_{i=1}^{N}p(\alpha_l|\mathbf{X}_i)(1-\eta_{lj})\gamma_{lj}}. \tag{A7}$$

For details regarding the estimation process using EM algorithm for estimating the SISM model parameters, refer to Appendix 5 of Kuo et al. (2016).

Standard Errors (SEs) for Item Parameter Estimates

Let $\boldsymbol{\beta}$ denote the vector of the item parameters of the SISM model. The information matrix $I(\boldsymbol{\beta})$ is defined as $I(\boldsymbol{\beta}) = -E\left[\partial^2 l(\mathbf{X}) / \partial^2 \boldsymbol{\beta}^2\right]$, where $l(\mathbf{X})$ is the log-marginalized likelihood of the observed data **X** (de la Torre, 2009). Let β and β' respectively denote $P_j(\boldsymbol{\alpha}_l)$ and $P_{j'}(\boldsymbol{\alpha}_l)$. The second derivative of the log-marginalized likelihood regarding to $P_j(\boldsymbol{\alpha}_l)$ and $P_{j'}(\boldsymbol{\alpha}_l)$ is given as

$$\frac{\partial^2 l(\mathbf{X})}{\partial P_j(\boldsymbol{\alpha}_l)\partial P_{j'}(\boldsymbol{\alpha}_l)} = -\sum_{i=1}^{I}\left\{\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l|\mathbf{X}_i)\frac{X_{ij}-P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)[1-P_j(\boldsymbol{\alpha}_l)]}\frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta}\right\}$$

$$\times\left\{\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l|\mathbf{X}_i)\frac{X_{ij'}-P_{j'}(\boldsymbol{\alpha}_l)}{P_{j'}(\boldsymbol{\alpha}_l)[1-P_{j'}(\boldsymbol{\alpha}_l)]}\frac{\partial P_{j'}(\boldsymbol{\alpha}_l)}{\partial \beta'}\right\}. \quad \text{(A8)}$$

After $p_j(\eta,\gamma|\mathbf{X}_i) = \sum_{\{\boldsymbol{\alpha}_l:\eta_{lj}=\eta,\gamma_{lj}=\gamma\}} p(\boldsymbol{\alpha}_l|\mathbf{X}_i)$ and $P_j(\eta,\gamma) = P_j(\boldsymbol{\alpha}_l)$ are defined, Equation A8 can be written as

$$\frac{\partial^2 l(\mathbf{X})}{\partial P_j(\boldsymbol{\alpha}_l)\partial P_{j'}(\boldsymbol{\alpha}_l)} = -\sum_{i=1}^{I}\left\{\sum_{\eta=0}^{1}\sum_{\gamma=0}^{1} p_j(\eta,\gamma|\mathbf{X}_i)\frac{X_{ij}-P_j(\eta,\gamma)}{P_j(\eta,\gamma)[1-P_j(\eta,\gamma)]}\frac{\partial P_j(\eta,\gamma)}{\partial \beta}\right\}$$

$$\times\left\{\sum_{\eta'=0}^{1}\sum_{\gamma'=0}^{1} p_{j'}(\eta',\gamma'|\mathbf{X}_i)\frac{X_{ij'}-P_{j'}(\eta',\gamma')}{P_{j'}(\eta',\gamma')[1-P_{j'}(\eta',\gamma')]}\frac{\partial P_{j'}(\eta',\gamma')}{\partial \beta'}\right\}, \quad \text{(A9)}$$

which sums the products of the expected values according to the posterior probabilities of students with η and γ for item *j*. $I(\boldsymbol{\beta})$ can be approximated by $I(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ represents the estimated item parameters. Therefore, the estimated item parameter standard error $SE(\hat{\boldsymbol{\beta}})$ can be obtained by calculating the square roots of the diagonal elements of the estimated item parameter covariance matrix $Cov(\hat{\boldsymbol{\beta}})$, which is approximated by the inverse information matrix $I^{-1}(\hat{\boldsymbol{\beta}})$.

Appendix B

Table B1. The Q-matrix for the short test.

Item	Attribute						
	Skill				Misconception		
	S1	S2	S3	S4	B1	B2	B3
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1
8	1	0	0	0	1	0	0
9	0	1	0	0	1	0	0
10	0	0	1	0	0	0	1
11	0	0	0	1	0	1	0
12	1	1	0	0	1	0	0
13	1	0	1	0	0	0	1
14	1	0	0	1	0	0	1
15	0	1	1	0	0	0	1
16	0	1	0	1	0	1	1
17	0	0	1	1	0	1	1
18	1	0	1	0	1	1	0
19	1	1	0	1	1	1	0
20	0	1	1	1	1	1	0

Note. S1, S2, S3, and S4 are skills. B1, B2, and B3 are misconceptions.

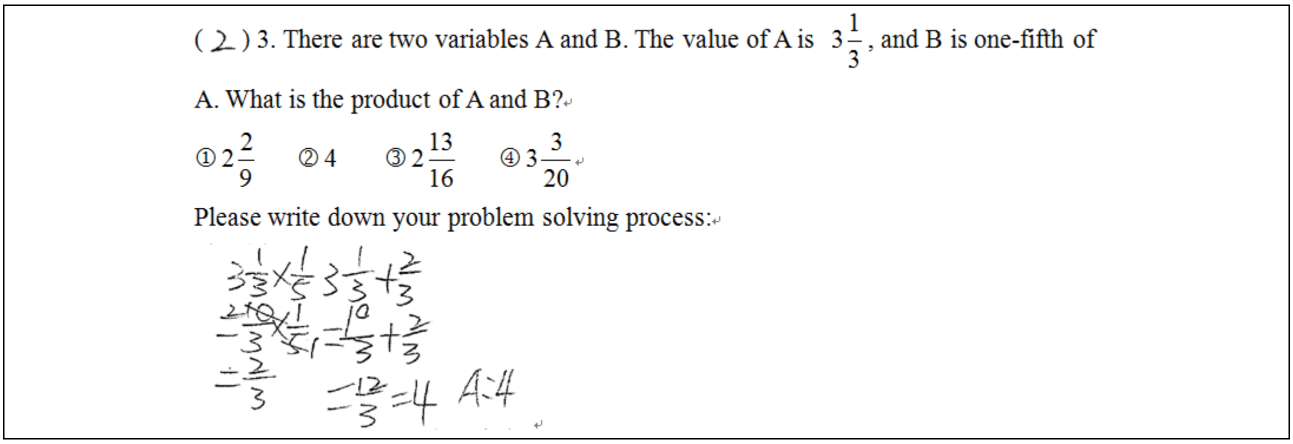
Table B2. The Q-matrix for the Fraction Multiplication Data.

Item	Attribute						
	Skill				Misconception		
	S1	S2	S3	S4	B1	B2	B3
1	1	1	0	1	1	1	0
2	0	1	1	0	1	0	0
3	0	1	0	1	0	1	1
4	1	1	0	1	1	1	0
5	1	1	1	1	1	1	0
6	1	1	0	1	1	1	0
7	0	1	1	1	0	1	0

Table B3. Q-matrices for the third and fourth real data-based simulation studies (modifications highlighted)

	Q3								Q4						
Item	S1	S2	S3	S4	B1	B2	B3	Items	S1	S2	S3	S4	B1	B2	B3
1	1	1	0	1	1	1	0	1	0	1	0	0	0	0	0
2	0	1	1	0	1	0	0	2	1	0	0	0	0	0	0
3	0	1	0	1	0	1	1	3	0	0	1	0	0	0	0
4	1	1	0	1	1	1	0	4	0	0	0	1	0	0	0
5	1	1	1	1	1	1	0	5	0	0	0	0	1	0	0
6	1	1	0	1	1	1	0	6	0	0	0	0	0	1	0
7	0	1	1	1	0	1	0	7	0	0	0	0	0	0	1
8	1	1	0	1	1	1	0	8	1	1	0	1	1	1	0
9	0	1	1	0	1	0	0	9	0	1	1	0	1	0	0
10	0	1	0	1	0	1	1	10	0	1	0	1	0	1	1
11	1	1	0	1	1	1	0	11	1	1	0	1	1	1	0
12	1	1	1	1	1	1	0	12	1	1	1	1	1	1	0
13	1	1	0	1	1	1	0	13	1	0	0	0	1	0	0
14	0	1	1	1	0	1	0	14	0	1	1	1	0	1	0

Appendix C



(2) 3. There are two variables A and B. The value of A is $3\frac{1}{3}$, and B is one-fifth of A. What is the product of A and B?

① $2\frac{2}{9}$ ② 4 ③ $2\frac{13}{16}$ ④ $3\frac{3}{20}$

Please write down your problem solving process:

$3\frac{1}{3} \times \frac{1}{5}$ $3\frac{1}{3} + \frac{2}{3}$

$= \frac{\cancel{10}}{3} \times \frac{1}{\cancel{5}} = \frac{10}{3} + \frac{2}{3}$

$= \frac{2}{3}$ $= \frac{12}{3} = 4$ A:4

Figure C1. An example of Item 3.