

A comparison of two estimation algorithms for Samejima's continuous IRT model

Cengiz Zopluoglu

Published online: 26 June 2012
© Psychonomic Society, Inc. 2012

Abstract This study compares two algorithms, as implemented in two different computer softwares, that have appeared in the literature for estimating item parameters of Samejima's continuous response model (CRM) in a simulation environment. In addition to the simulation study, a real-data illustration is provided, and CRM is used as a potential psychometric tool for analyzing measurement outcomes in the context of curriculum-based measurement (CBM) in the field of education. The results indicate that a simplified expectation-maximization (EM) algorithm is as effective and efficient as the traditional EM algorithm for estimating the CRM item parameters. The results also show promise for using this psychometric model to analyze CBM outcomes, although more research is needed in order to recommend CRM as a standard practice in the CBM context.

Keywords Continuous response model · Continuous IRT model · Item response theory · Item parameter estimation · Item parameter recovery · Simulation · Curriculum-based measurement

Introduction

A latent trait model for continuous item scores has been suggested as a limiting form of the graded response model (Samejima, 1973); however, the continuous response model (CRM) has not received much attention in the fields of education and psychology, although there have been a few applications (Bejar, 1977; Ferrando, 2002; Wang & Zeng, 1998). This lack of popularity may be attributable to the preferences

of applied researchers for using a simpler linear factor model in modeling continuous measurement outcomes, instead of using a more complex nonlinear model, and to the availability and accessibility of software for estimating linear factor model parameters using a limited-information approach.

There are two approaches that have been used in practical applications for estimating the CRM item parameters: limited-information and full-information approaches. In the limited-information framework (also known as heuristic estimation), the observed scores are first rescaled between 0 and 1, and the rescaled scores are transformed to logit. Then a linear factor model is fitted to the covariance/correlation matrix obtained from the transformed data (Bejar, 1977; Ferrando 2002). This approach can be implemented using any well-known popular software (e.g., MPlus, LISREL, SAS, SPSS). Once the linear factor model parameters are estimated, researchers can put them into the IRT metric, if there is any interest, using the linear transformations presented by Ferrando (2002, Equation 5).

Within the full-information framework, there are two approaches presented in the literature for estimating the CRM item parameters (Shojima, 2005; Wang & Zeng, 1998). Both approaches use the marginal maximum likelihood (MML) method and the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), with a slightly different variation. In the first approach, Wang and Zeng obtained an expected log-likelihood function in the E-step by approximating the integration over the posterior ability distribution with Gaussian quadrature points or equally spaced quadrature points, as a standard procedure in most other psychometric software (e.g., BILOG, MULTI-LOG). In the M-step, they iteratively solved for the item parameters that maximized the log-likelihood function, using a numerical approach, the Newton–Raphson method. The EM2 program was developed in C language and released as an executable file to implement the algorithm above, and it is available for free from its authors.

C. Zopluoglu (✉)
Department of Educational Psychology, University of Minnesota,
140 Education Sciences Building, 56 East River Road,
Minneapolis, MN 55455-0364, USA
e-mail: zoplu001@umn.edu

In another approach, Shojima (2005) developed a simplified version of the EM algorithm for estimating the CRM item parameters. In this simplified algorithm, Shojima (2005) showed that the expected log-likelihood function can be explicitly obtained in the E-step without using any numerical approximation for the integration over the posterior ability distribution. In addition, Shojima (2005) derived the closed formulas for estimating the item parameters by assuming a uniform prior distribution for each item parameter. Therefore, the item parameters that maximize the log-likelihood function can be obtained in a noniterative fashion without using the Newton–Raphson method in the M step. An R package, EstCRM (Zopluglu, 2012), that implements Shojima’s (2005) simplified algorithm for estimating the CRM item parameters is available for free.

A practical concern is how well these different approaches recover the CRM item parameters. So far, the limited-information approach has been compared with the full-information approach, as implemented in the EM2 program, in a simulation setting (Ferrando, 2002), and the performances of two full-information approaches have been examined independently in separate experimental settings (Shojima, 2005; Wang & Zeng, 1998).

Research purpose

There are two main purposes of this study. The first purpose is to compare two different full-information approaches proposed for estimating Samejima’s CRM item parameters in the same setting, using simulated data, and to examine whether the simplified EM algorithm (Shojima, 2005), as implemented in the EstCRM package, outperforms a more traditional EM algorithm (Wang & Zeng, 1998), as implemented in the EM2 program. The second purpose of the study is to compare two full-information approaches, using real data, and to provide an application of CRM in the field of education in the context of curriculum-based measurement (Deno, 1985; Deno & Mirkin, 1977).

Theoretical background

A response model for continuous measurement outcomes was originally proposed by Samejima (1973) as a limiting case of the graded response model. Yet Wang and Zeng (1998) proposed a different parameterization of CRM. The reason for their reparameterization was to create an item difficulty parameter that is on the θ scale and an item discrimination parameter counterpart in discrete IRT models to make the model more understandable for practitioners. In this model, the probability of the i th examinee obtaining a

score of x or higher on the j th item is equal to

$$P(X_{ij} \geq x | \theta_i, a_j, b_j, \alpha_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v e^{-\frac{t^2}{2}} dt, \\ v = a_j(\theta_i - b_j - \frac{1}{\alpha_j} \ln \frac{x_{ij}}{k_j - x_{ij}})$$

where x_{ij} is the observed score between 0 and k_j for examinee i on item j ; θ_i is the ability level of examinee i ; a_j , b_j , and α_j are the discrimination, difficulty, and scaling parameters, respectively, for item j ; and k_j is the possible maximum score on item i . In this model, a and b parameters are interpreted similarly as in binary and polytomous IRT models. The model includes an additional scaling parameter α with no practical interpretation.

The conditional probability distribution function is also derived as

$$P(X_{ij} = z | \theta_i, a_j, b_j, \alpha_j) = \frac{a_j}{\sqrt{2\pi\alpha_j}} \\ \times \exp \left\{ - \frac{[a_j(\theta_i - b_j - \frac{z}{\alpha_j})]^2}{2} \right\},$$

where $z = \ln \frac{x}{k-x}$ (Wang & Zeng, 1998). The observed scores (x) are transformed to the z scores as defined above for simplicity in the estimation process.

Estimating item parameters

As was briefly discussed above, there are two approaches for estimating the CRM item parameters: limited-information and full-information approaches. A detailed technical discussion for estimating the CRM item parameters in the limited-information framework was given by Ferrando (2002). Since the focus of the present article is to compare two full-information approaches, it will not be covered here.

Similar to the other well-known psychometric software, the MML-EM approach is used to estimate the CRM item parameters within the full-information framework. Two different implementations of the MML-EM approach have appeared in the literature (Shojima, 2005; Wang & Zeng, 1998). Wang and Zeng’s implementation is more traditional, while Shojima’s implementation is a simplified procedure with additional assumptions on item parameter distributions. In both implementations, the likelihood of the observed response vector for person i is defined as

$$P_i(z_i | \theta_i, a_j, b_j, \alpha_j) = \prod_j P(z_{ij} | \theta_i, a_j, b_j, \alpha_j)$$

on the basis of the local independence assumption among items. Then the log-likelihood function to be maximized

in the E-step is derived as

$$\left(\sum_{i=1}^N \int_{\theta_i} \ln P_i(z_i|\theta_i, a_j, b_j, \alpha_j) P_i(\theta_i|a_j, b_j, \alpha_j, z_i) d\theta_i \right) + \ln P(a_j, b_j, \alpha_j) + const.,$$

where $P(a_j, b_j, \alpha_j)$ is a prior distribution on the item parameters.

Wang and Zeng (1998) approximated the integration over the ability distribution using Gaussian quadrature

$$\left(\left[N \sum_{j=1}^n \left(\ln a_j + \ln \frac{1}{\alpha_j} \right) \right] - \left[\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^n a_j^2 \left\{ \left(b_j + \frac{z_{ij}}{\alpha_j} - \mu_i^{(k)} \right)^2 + \sigma^{(k)2} \right\} \right] \right) + \ln P(a_j, b_j, \alpha_j) + const.$$

where $\mu_i^{(k)}$ and $\sigma^{(k)}$ are the estimated mean and standard deviation of the posterior ability distribution in the k th EM cycle. The mean and variance of the posterior ability distribution is updated in each EM cycle. In addition, Shojima (2005) concluded that the item parameters can be solved explicitly in the M-step without using the Newton–Raphson method if uniform prior distributions are assumed on the item parameters (interested readers are encouraged to review the original sources for additional equations and more technical information).

Two software are available to practitioners for estimating the CRM item parameters within the full-information framework. The first one, the EM2 program (Wang & Zeng, 1998), is developed in C language and is available as an executable file from its authors. EM2 uses the following starting parameters for item j in the first EM cycle:

$$a_j = \frac{1}{\sqrt{\text{var}(z_j) - 1}}, \quad b_j = -\text{mean}(z_j), \quad \alpha_j = 1.$$

The user can specify the maximum number of EM cycles, criteria to stop the EM cycles, the maximum number of Newton–Raphson iterations in the M-step, and criteria to stop the iterations. The syntax file to prepare for the program is very similar to the ones prepared for other well-known psychometric software (e.g., BILOG, MULTILOG).

The second option for practitioners is the EstCRM package (Zopluoglu, 2012) developed in R language (R Development Core Team, 2011). In contrast to the EM2 program, the simplified EM algorithm (Shojima, 2005) is used in the EstCRM package to estimate the CRM item parameters. The package uses 1, $-\text{mean}(z_j)$, and 1 as the starting parameters for a , b , and α parameters, respectively, for item j in the first EM cycle. The reason for using a different starting value for parameter a is that the variance of transformed scores for item j may not be bigger than 1, so $\sqrt{\text{var}(z_j) - 1}$ is not

points or equally spaced quadrature points to obtain the log-likelihood in the E-step. In the following step, they obtained the item parameter estimates by simultaneously solving the first derivatives of the log-likelihood function with respect to item parameters, using a Newton–Raphson method in the M-step.

In a slightly different implementation, Shojima (2005) showed that the log-likelihood function in the E-step can be explicitly computed without approximating the integration. Shojima (2005) derived the log-likelihood function in the E-step as equal to

always a real number. The R code for analyzing sample data sets is available in the package manual.

Recovery of item parameters for CRM

Wang and Zeng (1998) evaluated their algorithm in a simulation study in 12 different conditions (3 [sample size] \times 2 [number of items] \times 2 [type of ability distribution]) and reported item recovery statistics for the EM2 program. The simulation results indicated that EM2 recovered item parameters well enough, especially when the ability distribution was normal and the sample size was bigger than 500. The root mean square error (RMSE) values were 0.06, 0.04, and 0.05 for parameters a , b , and α , respectively, when sample size was 2,000, number of items was 6, and the ability distribution was normal. Similarly, the RMSE values were 0.04, 0.05, and 0.03 for parameters a , b , and α , respectively, when sample size was 2,000, number of items was 12, and ability distribution was normal. When the ability distribution was skewed, the RMSE values were slightly higher, especially for parameter b . However, this study was limited, because only one fixed set of parameters was used across all replications.

In another simulation study, Shojima (2005) evaluated the item recovery of his simplified algorithm in nine different conditions (3 [sample size] \times 3 [number of items]). The parameter a and parameter α were generated from a log-normal distribution with a mean of zero and standard deviation of .09, while the parameter b was drawn from a standard normal distribution across all conditions and replications. The ability levels were drawn from a standard normal distribution across all conditions. The results were almost identical to those in Wang and Zeng’s (1998) study for similar conditions. Shojima (2005) reported that the RMSE values were 0.04, 0.05, and 0.03 for parameters a , b , and α , respectively, when the sample size was 2,000 and the number of items was 10.

Both studies reported that the item recovery was better as the sample size and number of items increased.

The two different approaches differed in terms of estimation bias. Shojima's (2005) simplified algorithm seemed to produce less biased parameter estimates. The estimation bias for the simplified algorithm was not larger than .009 in any experimental condition and was negligible. On the other hand, Wang and Zeng (1998) reported that the estimation bias for the EM2 program ranged from .024 to .048 for parameter a , from .001 to .010 for parameter b , and from .01 to .048 for parameter α when the distribution of ability was standard normal.

Although the two previous studies provide some information regarding the utility of two full-information approaches in estimating the CRM item parameters, they are limited in the conditions used to simulate data when assessing item recovery. In addition, both studies reported the item recovery statistics under independent experimental settings. The goal of the present study is to provide an empirical comparison between two MML-EM algorithms in estimating the CRM item parameters under the same experimental setting, using both simulated and real data. The study provides some insight regarding the utility of two computer programs available for practitioners who may want to use Samejima's (1973) IRT model for their continuous measurement outcomes.

Method

Simulated data

Five independent variables were manipulated in the study: number of items (10 and 20), sample size (500 and 1,000), the distribution of parameter a (Lognormal $\sim [0,.3]$, Lognormal $\sim [0,.1]$, Uniform $\sim [.4,2]$), the distribution of parameter b (Normal $\sim [0,1]$, Normal $[0,.5]$, Uniform $[-3,3]$), and the distribution of parameter α (Lognormal $\sim [0,.3]$, Lognormal $\sim [0,.1]$, Uniform $\sim [.4,2]$). Independent variables were fully crossed, for a total of 108 conditions. All simulation conditions were based on a common ability distribution. The ability levels of the examinees were drawn from a standard normal distribution across all conditions.

In CRM, the conditional probability distribution function of transformed score for person i on item j (z_{ij}) follows a normal distribution with a mean of $\alpha_j(\theta_i - b_j)$ and a variance of α_j^2/a_j^2 (Shojima, 2005; Wang & Zeng, 1998). Given the generated ability level and item parameters, each z_{ij} was drawn from a normal distribution with a mean of $\alpha_j(\theta_i - b_j)$ and a variance of α_j^2/a_j^2 . Each item was assumed to have a score scale between 0 and 50, so the generated z_{ij} was transformed to observed score (x_{ij}), using the equation

$$z_{ij} = \ln \frac{x_{ij}}{k_j - x_{ij}}.$$

One hundred data sets that follow CRM were simulated within each experimental condition, given the generated ability levels and item parameters.

After simulating the data sets, item parameters were calibrated in two different computer programs, EM2 and EstCRM, implementing two different approaches as described above. The number of maximum EM cycles was set to 500, and the EM cycles were terminated when the difference in log-likelihood values between two successive EM cycles was smaller than .01. The same settings were used in both programs. For the EM2 program, the maximum number of Newton–Raphson iterations in the M-step was set to 50, and the criterion to stop the iterations was set to .0001. Since the EstCRM package directly computes the item parameter estimates in the M-step, there was no need for a similar setting.

Two dependent variables, $RMSE$ and mean error (ME), were computed to assess the accuracy and bias. The $RMSE$ is the square root of the averaged squared deviation between a parameter and its estimate. For all the simulated data within an experimental condition, the $RMSE$ values for corresponding parameters a , b , and α were computed across items:

$$RMSE_r(\lambda) = \sqrt{\frac{\sum_{k=1}^j (\lambda_k - \hat{\lambda}_k)^2}{j}},$$

where $RMSE_r$ is the $RMSE$ value in the r th replication and λ_k and $\hat{\lambda}_k$ are the corresponding parameter and parameter estimate for item k , respectively. Then the $RMSE$ values were averaged across all replications within an experimental condition. Similarly, the ME s for the simulated data within an experimental condition were computed for the corresponding parameters a , b , and α as the following:

$$ME_R(\lambda) = \frac{\sum_{k=1}^j \lambda_k - \hat{\lambda}}{j}.$$

Then the ME values were averaged across all replications for an experimental condition

Real data

Data, provided by a large, urban, Midwestern school district, were collected from 2,810 students in the first grade at 49 different schools. Three of the Early Mathematics Curriculum Based Measurement (EM-CBM) probes were administered at the beginning of Fall 2008 by the school district. These probes were the number identification (NI), quantity discrimination (QD), and quantity array (QA), used to measure the early numeracy skills at the kindergarten and first-grade level (Clarke & Shinn, 2004; Lembke & Foegen, 2009). Each probe

was individually administered to students for 1 min. The NI probe consisted of numbers from 1 to 90, and students were required to name the numbers orally. The count of correctly named numbers was recorded as the observed score for the students. The QD probe required each student to compare two numbers and to identify the larger one. The QD probe had 54 comparisons, and the number of correct identifications was recorded as the observed score. The QA probe required identifying the number of dots ranging from 1 to 10 in a box. The quantity array probe had 38 boxes, and the number of correctly identified boxes was the observed scores for the probe.

In addition to the EM-CBM probes, three reading passages were presented to the students to measure oral reading fluency (Hintze, Christ, & Methe, 2006). The students were asked to start reading the passage and to stop reading after 1 min. The number of words read correctly per minute was recorded as the observed score for each student.

Samejima's (1973) continuous IRT model was fitted separately to the EM-CBM and curriculum-based measurement reading (CBM-R) passage data. The item parameters were estimated using both software, and the standard errors of item parameter estimates were calculated using a nonparametric bootstrap sampling approach with 500 bootstrap samples.

Results

Simulation study

The results of the simulation study for all 108 conditions are presented in Tables 1, 2, 3, and 4. Tables 1 and 2 present the results regarding the estimation bias. The sample size and number of items did not seem to affect the bias in estimation regardless of the type of computer program used to estimate item parameters. However, the type of parameter distribution had some effect. In almost all conditions, there was no bias in the estimation of b parameters. Across all 108 conditions, only 3 conditions were flagged as problematic in estimating b parameters for the EM2 program in terms of bias. The EstCRM package did not show any bias in the estimation of b parameters in any condition. Both software showed a slightly positive bias in estimating parameter a in some conditions. In most of these conditions, either one or two of the parameters a , b , and α had uniform distribution. The conditions where the parameters had normal or lognormal distributions, the bias for the parameter a was ignorable. In terms of parameter α , the EM2 program showed a positive bias in a few conditions regardless of sample size when the number of items was 10; however, there were only two cases in which the bias occurred when the number of items was 20. The conditions in which the bias in estimating parameter α occurred were mostly the conditions where

the parameter α had uniform distribution. In contrast to the EM2 program, the EstCRM package showed remarkable negative bias in estimating parameter α in most of the conditions. The only conditions in which the bias did not occur in estimating parameter α were the conditions where all the parameters had either normal or lognormal distributions.

The results regarding the precision of item parameter estimates are presented in Tables 3 and 4. Both computer programs estimated parameter a with almost identical precision in most conditions. In a few cases, EstCRM seemed to give more precise estimates for parameter a , but the difference was ignorable. In terms of parameter b , the EstCRM package produced more precise estimates than did the EM2 program. Overall, the *RMSE* value of parameter b was about .23 for the EM2 program, while it was about .13 for the EstCRM package. A closer look at the tables revealed that the EstCRM package outperformed the EM2 program, especially in conditions where the distribution of parameter b was uniform. In contrast to the b parameter, the EM2 program outperformed the EstCRM package in estimating parameter α . The overall *RMSE* values for the parameter α were .11 and .17 for the EM2 program and EstCRM package, respectively.

Real data

The descriptive statistics for the EM-CBM probes (number identification, quantity discrimination, and quantity array) and CBM reading passages, as well as the correlations among the measures, are presented in Table 5. The data are assumed to be unidimensional to be able to fit Samejima's (1973) continuous IRT model, since it is a standard assumption in other IRT models. The eigenvalues were extracted from the correlations among the probes to examine the plausibility of this assumption. The eigenvalues of 2.23, 0.47, and 0.30 and the eigenvalues of 2.94, 0.04, and 0.02 are obtained, respectively, from the correlations among the three EM-CBM probes and from the correlations among the three CBM reading passages. Using a parallel analysis approach (Horn, 1965), the eigenvalues of random multivariate data with the same sample size and number of items were also computed. The 99th percentiles of the first, second, and third eigenvalues from random data were obtained as 1.07, 1.02, and 0.99, respectively. Since the first sample eigenvalues were much higher than the first eigenvalue of random data and the second sample eigenvalues were much lower than the second eigenvalue of random data, it was concluded that the unidimensionality assumption is plausible for the three EM-CBM probes, as well as the three CBM reading passages.

Samejima's (1973) continuous IRT model was fit to the EM-CBM and CBM reading passage data separately, and the parameters including the standard errors, as presented in

Table 1 Mean error between item parameters and estimates ($N = 500$)

Program			ME						ME					
			Number of items = 20						Number of items = 10					
Type of distribution			EM2			EstCRM			EM2			EstCRM		
			<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α
Par. <i>a</i>	Par. <i>b</i>	Par. α												
LN1	N1	LN1	0.01	0.00	0.01	0.02	0.00	-0.01	0.01	0.01	0.01	0.01	0.01	-0.02
LN1	N1	LN2	0.01	0.00	0.01	0.02	0.00	-0.05	0.03	-0.01	0.03	0.02	0.00	-0.06
LN1	N1	U1	0.04	0.00	0.04	0.03	0.00	-0.13	0.07	0.00	0.08	0.03	0.00	-0.12
LN1	N2	LN1	0.01	-0.01	0.00	0.03	-0.01	-0.05	0.01	0.00	0.00	0.03	-0.01	-0.05
LN1	N2	LN2	0.02	-0.01	0.02	0.03	-0.01	-0.08	0.04	0.00	0.04	0.03	0.00	-0.08
LN1	N2	U1	0.04	-0.01	0.05	0.04	0.00	-0.15	0.08	-0.02	0.08	0.04	-0.01	-0.14
LN1	U2	LN1	0.03	0.00	0.00	0.07	0.01	-0.13	0.04	0.00	0.00	0.07	0.00	-0.13
LN1	U2	LN2	0.04	-0.09	0.00	0.07	0.00	-0.15	0.05	0.01	0.02	0.06	0.00	-0.15
LN1	U2	U1	0.06	-0.01	0.05	0.07	0.01	-0.20	0.08	0.03	0.09	0.07	0.00	-0.19
LN2	N1	LN1	0.02	0.00	0.02	0.01	0.00	-0.03	0.01	0.00	0.02	0.01	0.00	-0.03
LN2	N1	LN2	0.02	0.01	0.02	0.02	0.00	-0.05	0.03	0.00	0.03	0.01	0.00	-0.07
LN2	N1	U1	0.06	0.00	0.06	0.03	0.00	-0.12	0.10	-0.01	0.11	0.03	0.00	-0.12
LN2	N2	LN1	0.03	0.01	0.02	0.03	0.01	-0.05	0.02	0.00	0.01	0.03	0.00	-0.05
LN2	N2	LN2	0.03	-0.01	0.02	0.03	0.01	-0.09	0.05	-0.01	0.04	0.03	-0.01	-0.09
LN2	N2	U1	0.06	-0.01	0.06	0.05	0.00	-0.14	0.10	0.01	0.09	0.05	0.00	-0.15
LN2	U2	LN1	0.05	0.00	0.01	0.08	-0.01	-0.14	0.05	0.00	0.01	0.08	0.00	-0.14
LN2	U2	LN2	0.05	-0.01	0.02	0.06	0.00	-0.14	0.06	0.06	0.03	0.07	0.00	-0.15
LN2	U2	U1	0.07	0.02	0.05	0.08	0.01	-0.20	0.10	-0.02	0.10	0.08	-0.01	-0.19
U1	N1	LN1	0.04	-0.01	0.03	0.01	-0.01	-0.03	0.04	0.00	0.03	0.02	0.00	-0.03
U1	N1	LN2	0.05	-0.01	0.04	0.02	-0.01	-0.07	0.07	-0.01	0.06	0.03	-0.01	-0.06
U1	N1	U1	0.08	0.00	0.07	0.04	0.00	-0.12	0.13	0.00	0.13	0.05	0.00	-0.11
U1	N2	LN1	0.05	0.00	0.03	0.04	0.01	-0.06	0.04	0.00	0.03	0.03	0.00	-0.06
U1	N2	LN2	0.07	0.00	0.04	0.04	0.00	-0.09	0.06	0.00	0.05	0.05	0.00	-0.08
U1	N2	U1	0.10	-0.01	0.08	0.06	0.00	-0.15	0.13	0.00	0.13	0.07	0.00	-0.13
U1	U2	LN1	0.07	0.00	0.02	0.09	-0.01	-0.13	0.07	-0.01	0.02	0.10	-0.01	-0.13
U1	U2	LN2	0.07	0.08	0.03	0.10	0.01	-0.13	0.11	0.07	0.04	0.09	0.00	-0.15
U1	U2	U1	0.12	0.01	0.06	0.12	-0.01	-0.18	0.16	0.01	0.13	0.12	0.00	-0.19
		Mean	0.05	0.00	0.03	0.05	0.00	-0.11	0.06	0.00	0.05	0.05	0.00	-0.11
		SD	0.03	0.02	0.02	0.03	0.01	0.05	0.04	0.02	0.04	0.03	0.00	0.05

Note. Each cell statistic is the average over 100 replications. LN1, Lognormal $\sim (0, .1)$; LN2, Lognormal $\sim (0, .3)$; N1, Normal $\sim (0, .5)$; N2, Normal $\sim (0, 1)$; U1, Uniform $\sim (.4, 2)$; U2, Uniform $\sim (-3, 3)$

Table 6, were estimated for each probe. Since these measures were not regular items, the parameters were actually interpreted as *probe parameters*. Therefore, discrimination and difficulty parameters were interpreted as *probe difficulty* and *probe discrimination*. For instance, fitting Samejima's continuous IRT model helps to assign a difficulty and discrimination level for each reading passage administered.

The probe parameters were calibrated by using two different approaches as implemented in two different software. The probe parameters calibrated from two software were

very close to each other for the EM-CBM probes. The probe discrimination and probe difficulty parameters were estimated slightly higher by the EM2 program. The empirical standard errors, based on the nonparametric bootstrap sampling, were lower in the EM2 program, except for the difficulty parameter for the quantity array probe. EM-CBM probes differed in difficulty levels. The quantity array probe was the most difficult probe, while the number identification probe was second and the quantity discrimination probe was last in difficulty. The number identification and

Table 2 Mean error between item parameters and estimates ($N = 1,000$)

Program			ME						ME					
			Number of items = 20						Number of items = 10					
Type of distribution			EM2			EstCRM			EM2			EstCRM		
			<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α
Par. <i>a</i>	Par. <i>b</i>	Par. α												
LN1	N1	LN1	0.00	0.00	0.01	0.01	0.00	-0.02	0.00	0.00	0.01	0.01	0.00	-0.02
LN1	N1	LN2	0.02	-0.01	0.02	0.02	-0.01	-0.05	0.03	0.01	0.03	0.02	0.01	-0.05
LN1	N1	U1	0.04	0.00	0.04	0.03	0.00	-0.11	0.07	0.00	0.08	0.03	0.00	-0.12
LN1	N2	LN1	0.02	0.00	0.01	0.03	0.00	-0.05	0.02	0.00	0.01	0.04	-0.01	-0.04
LN1	N2	LN2	0.02	0.01	0.02	0.03	0.00	-0.08	0.03	0.01	0.03	0.03	0.00	-0.08
LN1	N2	U1	0.04	0.00	0.03	0.04	0.01	-0.15	0.08	0.00	0.08	0.04	-0.01	-0.15
LN1	U2	LN1	0.03	-0.01	-0.01	0.07	-0.01	-0.13	0.04	-0.01	0.00	0.07	0.00	-0.14
LN1	U2	LN2	0.04	-0.01	0.02	0.06	0.00	-0.15	0.06	0.04	0.03	0.06	0.00	-0.16
LN1	U2	U1	0.06	0.00	0.04	0.07	-0.01	-0.19	0.09	0.02	0.09	0.07	0.01	-0.20
LN2	N1	LN1	0.02	-0.01	0.02	0.02	-0.01	-0.03	0.02	0.00	0.02	0.02	0.00	-0.03
LN2	N1	LN2	0.03	0.00	0.03	0.02	0.01	-0.07	0.05	0.00	0.05	0.02	0.00	-0.06
LN2	N1	U1	0.05	0.00	0.05	0.03	0.00	-0.13	0.10	0.00	0.10	0.03	0.00	-0.13
LN2	N2	LN1	0.03	0.00	0.02	0.04	0.00	-0.05	0.03	0.00	0.02	0.03	0.00	-0.06
LN2	N2	LN2	0.04	0.00	0.03	0.04	0.00	-0.09	0.05	-0.01	0.05	0.04	-0.01	-0.08
LN2	N2	U1	0.06	0.01	0.05	0.05	0.00	-0.13	0.10	0.02	0.10	0.05	0.01	-0.14
LN2	U2	LN1	0.04	-0.01	0.00	0.08	-0.02	-0.13	0.05	0.00	0.01	0.08	0.00	-0.13
LN2	U2	LN2	0.04	-0.02	0.02	0.07	0.01	-0.15	0.06	-0.01	0.03	0.08	-0.01	-0.15
LN2	U2	U1	0.09	-0.01	0.04	0.08	0.00	-0.19	0.11	-0.05	0.10	0.09	-0.01	-0.19
U1	N1	LN1	0.04	0.00	0.03	0.01	0.00	-0.03	0.03	0.00	0.03	0.02	0.00	-0.03
U1	N1	LN2	0.05	0.01	0.04	0.02	0.01	-0.06	0.06	0.00	0.05	0.02	0.00	-0.06
U1	N1	U1	0.09	0.01	0.08	0.04	0.01	-0.12	0.14	-0.01	0.13	0.04	0.00	-0.11
U1	N2	LN1	0.04	0.01	0.03	0.04	0.00	-0.05	0.04	0.01	0.03	0.04	0.01	-0.05
U1	N2	LN2	0.05	-0.02	0.03	0.05	-0.01	-0.08	0.08	0.00	0.06	0.05	0.00	-0.08
U1	N2	U1	0.10	0.01	0.06	0.07	-0.01	-0.13	0.14	0.01	0.13	0.07	0.00	-0.14
U1	U2	LN1	0.07	0.00	0.02	0.10	-0.01	-0.13	0.08	0.00	0.02	0.10	0.00	-0.13
U1	U2	LN2	0.06	0.00	0.03	0.10	0.00	-0.14	0.10	-0.03	0.05	0.10	0.00	-0.14
U1	U2	U1	0.10	-0.03	0.07	0.12	0.00	-0.20	0.16	-0.01	0.13	0.11	0.00	-0.19
		Mean	0.05	0.00	0.03	0.05	0.00	-0.10	0.07	0.00	0.05	0.05	0.00	-0.11
		SD	0.03	0.01	0.02	0.03	0.01	0.05	0.04	0.02	0.04	0.03	0.00	0.05

Note. Each cell statistic is the average over 100 replications. LN1, Lognormal $\sim (0, .1)$; LN2, Lognormal $\sim (0, .3)$; N1, Normal $\sim (0, .5)$; N2, Normal $\sim (0, 1)$; U1, Uniform $\sim (.4, 2)$; U2, Uniform $\sim (-3, 3)$

the quantity discrimination probes were equally discriminating among the students with low number sense and high number sense, while the quantity array probe had slightly lower discriminating power.

The *passage difficulty* and *passage discrimination* parameters were estimated for each CBM reading passage administered to the students. Similar to the EM-CBM probes, EM2 estimated the difficulty and discrimination parameters for reading passages slightly higher in magnitude with smaller empirical standard error (higher precision). All three reading

passages were very close to each other in difficulty level, and the results can be interpreted such that the passages were equivalent to each other in difficulty. However, the discriminating parameters for the passages were different in magnitude. The reading passages were highly discriminative among the low- and high-ability students in reading, and the discrimination value for the third reading passage is questionable. It is unexpectedly high in terms of the usual range observed for the discriminating parameter in IRT applications. In addition, the empirical standard error of the estimate is quite large enough

Table 3 Root mean square error between item parameters and estimates ($N = 500$)

Program			RMSE						RMSE					
			Number of items = 20						Number of items = 10					
Type of distribution			EM2			EstCRM			EM2			EstCRM		
			<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α
Par. <i>a</i>	Par. <i>b</i>	Par. α												
LN1	N1	LN1	0.06	0.06	0.05	0.07	0.07	0.07	0.06	0.06	0.05	0.07	0.07	0.07
LN1	N1	LN2	0.07	0.07	0.07	0.08	0.07	0.11	0.08	0.08	0.08	0.07	0.07	0.12
LN1	N1	U1	0.08	0.07	0.08	0.08	0.07	0.19	0.09	0.09	0.10	0.07	0.07	0.19
LN1	N2	LN1	0.06	0.08	0.05	0.08	0.10	0.10	0.06	0.08	0.05	0.08	0.11	0.10
LN1	N2	LN2	0.08	0.14	0.09	0.09	0.10	0.15	0.09	0.21	0.12	0.08	0.10	0.15
LN1	N2	U1	0.08	0.16	0.11	0.09	0.12	0.22	0.10	0.16	0.13	0.09	0.10	0.22
LN1	U2	LN1	0.07	0.11	0.06	0.11	0.24	0.20	0.07	0.11	0.06	0.11	0.24	0.20
LN1	U2	LN2	0.11	0.73	0.18	0.11	0.22	0.23	0.10	0.59	0.17	0.11	0.22	0.23
LN1	U2	U1	0.11	0.52	0.20	0.11	0.19	0.27	0.12	0.71	0.26	0.12	0.21	0.28
LN2	N1	LN1	0.07	0.06	0.05	0.07	0.07	0.08	0.07	0.07	0.06	0.07	0.07	0.08
LN2	N1	LN2	0.08	0.08	0.07	0.08	0.07	0.12	0.08	0.08	0.08	0.08	0.07	0.13
LN2	N1	U1	0.09	0.09	0.10	0.08	0.08	0.19	0.12	0.09	0.14	0.08	0.07	0.19
LN2	N2	LN1	0.07	0.08	0.06	0.09	0.11	0.11	0.07	0.08	0.05	0.08	0.11	0.11
LN2	N2	LN2	0.11	0.18	0.10	0.09	0.10	0.16	0.11	0.21	0.11	0.09	0.11	0.16
LN2	N2	U1	0.11	0.17	0.13	0.10	0.10	0.22	0.14	0.23	0.16	0.10	0.11	0.23
LN2	U2	LN1	0.09	0.12	0.06	0.14	0.24	0.20	0.09	0.12	0.06	0.13	0.24	0.21
LN2	U2	LN2	0.15	0.48	0.13	0.12	0.21	0.22	0.20	0.76	0.19	0.13	0.22	0.24
LN2	U2	U1	0.17	0.71	0.26	0.14	0.22	0.28	0.19	0.71	0.27	0.14	0.21	0.28
U1	N1	LN1	0.08	0.07	0.06	0.08	0.07	0.09	0.08	0.07	0.06	0.08	0.07	0.09
U1	N1	LN2	0.10	0.07	0.08	0.09	0.07	0.13	0.10	0.08	0.09	0.08	0.07	0.13
U1	N1	U1	0.11	0.08	0.10	0.10	0.07	0.18	0.16	0.11	0.15	0.10	0.08	0.18
U1	N2	LN1	0.09	0.09	0.06	0.10	0.11	0.12	0.08	0.09	0.06	0.10	0.11	0.12
U1	N2	LN2	0.13	0.15	0.10	0.11	0.11	0.16	0.15	0.23	0.13	0.11	0.11	0.15
U1	N2	U1	0.16	0.21	0.16	0.13	0.11	0.23	0.21	0.33	0.22	0.13	0.12	0.22
U1	U2	LN1	0.11	0.13	0.07	0.16	0.23	0.20	0.11	0.12	0.06	0.16	0.23	0.20
U1	U2	LN2	0.23	0.61	0.16	0.17	0.22	0.23	0.29	0.92	0.21	0.18	0.23	0.25
U1	U2	U1	0.25	0.74	0.26	0.19	0.23	0.27	0.28	0.79	0.29	0.19	0.23	0.29
		Mean	0.11	0.22	0.11	0.11	0.13	0.18	0.12	0.27	0.13	0.10	0.14	0.18
		SD	0.05	0.23	0.06	0.03	0.07	0.06	0.06	0.27	0.07	0.03	0.07	0.06

Note. Each cell statistic is the average over 100 replications. LN1, Lognormal $\sim (0, .1)$; LN2, Lognormal $\sim (0, .3)$; N1, Normal $\sim (0, .5)$; N2, Normal $\sim (0, 1)$; U1, Uniform $\sim (.4, 2)$; U2, Uniform $\sim (-3, 3)$

to be suspicious. This might suggest some item–model fit problem for the third reading passage.

Discussion

On the basis of the conditions investigated in this study, the results showed that Shojima's (2005) simplified EM algorithm, as implemented in the EstCRM package, is as effective and efficient as the traditional EM algorithm, as implemented

in the EM2 program, in most conditions. Both programs produced similar results regarding the estimation bias and precision; however, they had some advantages and disadvantages in a few conditions. The bias and precision in estimating the parameter a were very similar for both programs. Similarly, both programs showed almost no bias in estimating the b parameter in almost every condition. But the EstCRM package was more successful in terms of precision when the parameter b was estimated, especially in conditions where the distribution of parameter b was uniform. On the other

Table 4 Root mean square error between item parameters and estimates ($N = 1,000$)

Program			RMSE						RMSE					
			Number of items = 20						Number of items = 10					
Type of distribution			EM2			EstCRM			EM2			EstCRM		
			<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α	<i>a</i>	<i>b</i>	α
Par. <i>a</i>	Par. <i>b</i>	Par. α												
LN1	N1	LN1	0.05	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05
LN1	N1	LN2	0.06	0.05	0.05	0.05	0.05	0.10	0.06	0.05	0.06	0.05	0.05	0.10
LN1	N1	U1	0.07	0.06	0.07	0.06	0.05	0.17	0.09	0.06	0.09	0.06	0.05	0.18
LN1	N2	LN1	0.05	0.06	0.04	0.07	0.09	0.09	0.05	0.06	0.04	0.07	0.17	0.09
LN1	N2	LN2	0.06	0.11	0.06	0.07	0.10	0.14	0.07	0.23	0.10	0.07	0.09	0.15
LN1	N2	U1	0.07	0.12	0.09	0.07	0.08	0.22	0.09	0.15	0.12	0.07	0.08	0.22
LN1	U2	LN1	0.06	0.09	0.05	0.10	0.23	0.19	0.06	0.08	0.04	0.10	0.24	0.20
LN1	U2	LN2	0.08	0.33	0.11	0.10	0.19	0.23	0.10	0.70	0.17	0.10	0.21	0.23
LN1	U2	U1	0.10	0.56	0.21	0.11	0.19	0.26	0.12	0.54	0.21	0.11	0.19	0.27
LN2	N1	LN1	0.06	0.05	0.04	0.06	0.05	0.07	0.05	0.05	0.04	0.05	0.05	0.06
LN2	N1	LN2	0.07	0.05	0.06	0.06	0.05	0.12	0.08	0.06	0.08	0.06	0.05	0.12
LN2	N1	U1	0.08	0.06	0.08	0.07	0.05	0.19	0.11	0.07	0.12	0.07	0.05	0.19
LN2	N2	LN1	0.06	0.06	0.04	0.07	0.09	0.09	0.06	0.07	0.05	0.07	0.09	0.10
LN2	N2	LN2	0.08	0.10	0.08	0.08	0.08	0.14	0.11	0.19	0.10	0.08	0.09	0.14
LN2	N2	U1	0.11	0.18	0.12	0.09	0.08	0.20	0.13	0.21	0.15	0.09	0.10	0.21
LN2	U2	LN1	0.08	0.11	0.05	0.12	0.22	0.19	0.08	0.09	0.05	0.13	0.22	0.19
LN2	U2	LN2	0.14	0.44	0.13	0.12	0.19	0.22	0.22	1.03	0.22	0.13	0.20	0.23
LN2	U2	U1	0.17	0.65	0.23	0.13	0.18	0.27	0.19	0.76	0.27	0.13	0.19	0.27
U1	N1	LN1	0.06	0.05	0.05	0.06	0.05	0.07	0.06	0.05	0.05	0.06	0.05	0.07
U1	N1	LN2	0.08	0.06	0.07	0.07	0.05	0.11	0.09	0.06	0.07	0.07	0.05	0.11
U1	N1	U1	0.12	0.07	0.11	0.09	0.05	0.18	0.15	0.08	0.15	0.09	0.05	0.17
U1	N2	LN1	0.08	0.09	0.05	0.08	0.08	0.09	0.07	0.07	0.05	0.09	0.09	0.10
U1	N2	LN2	0.11	0.13	0.07	0.10	0.09	0.13	0.15	0.21	0.11	0.10	0.09	0.15
U1	N2	U1	0.16	0.21	0.15	0.12	0.10	0.21	0.18	0.24	0.19	0.12	0.09	0.21
U1	U2	LN1	0.10	0.09	0.05	0.17	0.22	0.19	0.11	0.10	0.05	0.17	0.22	0.19
U1	U2	LN2	0.20	0.47	0.13	0.16	0.19	0.21	0.28	0.87	0.20	0.17	0.20	0.22
U1	U2	U1	0.26	0.72	0.25	0.19	0.20	0.29	0.28	0.75	0.29	0.19	0.20	0.28
		Mean	0.10	0.19	0.09	0.09	0.11	0.16	0.11	0.25	0.12	0.09	0.12	0.17
		SD	0.05	0.20	0.06	0.04	0.07	0.07	0.07	0.30	0.07	0.04	0.07	0.07

Note. Each cell statistic is the average over 100 replications. LN1, Lognormal $\sim (0, .1)$; LN2, Lognormal $\sim (0, .3)$; N1, Normal $\sim (0, .5)$; N2, Normal $\sim (0, 1)$; U1, Uniform $\sim (.4, 2)$; U2, Uniform $\sim (-3, 3)$

hand, the EM2 program was more effective in estimating the parameter α in terms of both precision and bias. The negative bias was remarkable in the EstCRM package when the parameter α was estimated.

Both programs are available for practitioners at no charge. One limitation of the EM2 program is that it works only with 32-bit computers. It was compiled as an executable file in C language about 15 years ago, and there has not been an update since then. The program is available for free from its authors, but the practical use

of the program is limited to 32-bit computers. On the other hand, the EstCRM package is compiled in R language and works in an R environment in any Windows, Linux, and Mac OS platforms. The manual including the sample R code for analyzing the sample data sets are available online for the EstCRM package.

In addition to the full-information approaches examined in this study, the reader should also be aware that the CRM item parameters can also be estimated by fitting a linear factor model using a limited-information approach. From a

Table 5 The descriptive statistics for EM-CBM and CBM-R probes

	N	Mean	SD	Skew.	Kurt.	Min	Max	Correlation					
								NI	QD	QA	P1	P2	P3
NI	2,808	27.51	14.75	0.57	0.07	0	90	1	.70	.56	.64	.64	.62
QD	2,810	20.61	9.17	0.13	-0.17	0	54		1	.57	.50	.49	.48
QA	2,808	11.60	4.30	0.77	2.94	0	38			1	.45	.45	.45
P1	2,808	20.18	26.31	2.34	5.87	0	180				1	.97	.96
P2	2,810	22.67	33.16	2.71	8.29	0	251					1	.98
P3	2,810	21.92	32.46	2.71	8.06	0	228						1

Note. NI, number identification; QD, quantity discrimination; QA, quantity array; P1, reading passage 1; P2, reading passage 2; P3, reading passage 3

theoretical point of view, CRM is a more complex nonlinear counterpart to the linear factor model and is more appropriate for measurement outcomes. From a practical point of view, however, whether a nonlinear model fitted with the full-information approach outperforms a linear model fitted with the limited-information approach is an open question. A previous simulation study suggested that a linear model fitted with the limited information approach would perform as well as the nonlinear model fitted with the full-information approach, unless the items were at extreme difficulty levels and highly discriminating (Ferrando, 2002). Also, the full-information approach may be more efficient than the limited-information approach in case of a large amount of missing data that does not permit computing a satisfactory covariance matrix. More simulation studies are needed to investigate the conditions where a full-information approach or a limited-information approach would be preferred over the other.

A real data illustration was also provided in the study. The CRM was used as a potential psychometric tool for analyzing the data from EM-CBM probes and CBM reading passages. There are many advantages to using the CRM in the context of curriculum-based measurement. First, it provides an analytical way to examine *probe or passage equivalency* in a psychometric context when several reading passages or probes are

administered to the same sample of students in a single administration. Second, the analytical procedure for linking the scores under the CRM is already developed for common-examinee and common-item designs (Shojima, 2003). Using the recommended approach, the oral fluency reading scores can be linked across samples using the CRM as an alternative method to the observed score equating procedures (Albano & Rodriguez, 2012). Third, the closed formula for estimating the ability level from the calibrated item parameters is derived for the CRM (Samejima, 1973). The estimation of ability under the CRM takes both *probe or passage difficulty* and *probe or passage discrimination* into account, which is not a standard practice in the current application.

The CRM has the potential to help CBM practitioners with some technical difficulties faced in practice. The real data illustration in this study aimed to introduce CRM for practitioners and to open the doors for its applicability in education, especially for CBM outcomes. Future research should analyze more data for different samples of students by using different sets of probes or reading passages and assess the model fit at the item and person level.

Lastly, one aspect of this study should be noted as a limitation. In the present simulation study, it is assumed that there are no missing data. Certainly, the presence and nature of missing data will deteriorate the precision and bias of

Table 6 Parameter estimates for the EM-CBM probes and CBM-R passages

	EstCRM package			EM2 program		
	a	b	α	a	b	α
NI	1.265 (.145)	1.126 (.073)	.888 (.059)	1.383 (.079)	1.159 (.044)	0.847 (.031)
QD	1.261 (.162)	.707 (.046)	.870 (.061)	1.385 (.096)	0.726 (.031)	0.821 (.036)
QA	0.932 (.079)	1.789 (.139)	.504 (.042)	0.966 (.051)	1.867 (.210)	0.465 (.021)
P1	2.112 (.125)	1.669 (.042)	1.734 (.050)	2.490 (.106)	1.913 (.041)	1.454 (.029)
P2	1.357 (.051)	1.773 (.038)	1.994 (.047)	1.856 (.060)	1.969 (.036)	1.678 (.028)
P3	4.822 (.772)	1.673 (.034)	1.965 (.047)	5.142 (.645)	1.944 (.034)	1.628 (.026)

Note. The standard errors, reported in parenthesis, were obtained using a nonparametric bootstrap sampling based on 500 resamples with replacement

item parameter estimates, and future simulation studies should manipulate the presence and nature of missing data. It is hypothesized that the presence of missing data will have more effect on the item parameter estimates in the limited-information approach than in the full-information approach, because either list-wise or pair-wise deletion will lead to more information loss when the covariance matrix is computed, as compared with the full-information approach that uses all item responses available. To conclude, as one of the reviewers pointed out, the continuous response formats will be more frequently used with the increasing popularity of computer administrations. This study aimed to address a small piece regarding the item parameter estimation. More studies are needed in the future to examine item parameter estimation, DIF analysis, linking, model fit, and item and person fit in the context of CRM, to provide the best tools for practitioners.

Author Note The author acknowledges Dr. David Heistad, Dr. Chi-Keung (Alex) Chan, and Mary Pickart at the Minneapolis Public Schools for their tremendous support, as well as willingness to share the data for this study. The author also greatly appreciates Dr. Chi-Keung (Alex) Chan for serving as the mentor for his internship.

References

- Albano, A. D., & Rodriguez, M. C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology, 50*, 43–59.
- Bejar, I. I. (1977). An application of the continuous response level model to personality measurement. *Applied Psychological Measurement, 1*, 509–521.
- Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234–248.
- Dempster, A. P., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B, 39*, 1–38.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232.
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Arlington: Council for Exceptional Children.
- Ferrando, P. J. (2002). Theoretical and empirical comparisons between two models for continuous item responses. *Multivariate Behavioral Research, 37*, 521–542.
- Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*, 45–56.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.
- Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research and Practice, 24*, 12–20.
- R Development Core Team. (2011). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika, 38*, 203–219.
- Shojima, K. (2003). Linking tests under the continuous response model. *Behaviormetrika, 30*, 155–171.
- Shojima, K. (2005). A noniterative item parameter solution in each EM cycle of the continuous response model. *Educational Technology Research, 28*, 11–22.
- Wang, T., & Zeng, L. (1998). Item parameter estimation for a continuous response model using an EM algorithm. *Applied Psychological Measurement, 22*, 333–344.
- Zopluglu, C. (2012). EstCRM: An R package for Samejima's continuous IRT model. *Applied Psychological Measurement, 36*, 149–150.