

UCLA

UCLA Previously Published Works

Title

Estimating the Nominal Response Model Under Nonnormal Conditions

Permalink

<https://escholarship.org/uc/item/2hj536kk>

Journal

Educational and Psychological Measurement, 74(3)

ISSN

0013-1644

Authors

Preston, Kathleen Suzanne Johnson
Reise, Steven Paul

Publication Date

2014-06-01

DOI

10.1177/0013164413507063

Peer reviewed

Estimating the Nominal Response Model Under Nonnormal Conditions

Educational and Psychological
Measurement

2014, Vol. 74(3) 377–399

© The Author(s) 2013

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164413507063

epm.sagepub.com



Kathleen Suzanne Johnson Preston¹ and
Steven Paul Reise²

Abstract

The nominal response model (NRM), a much understudied polytomous item response theory (IRT) model, provides researchers the unique opportunity to evaluate within-item category distinctions. Polytomous IRT models, such as the NRM, are frequently applied to psychological assessments representing constructs that are unlikely to be normally distributed in the population. Unfortunately, models estimated using estimation software with the MML/EM algorithm frequently employs a set of normal quadrature points, effectively ignoring the true shape of the latent trait distribution. To address this problem, the current research implements an alternative estimation approach, Ramsay Curve Item Response Theory (RC-IRT), to provide more accurate item parameter estimates modeled under the NRM under normal, skewed, and bimodal latent trait distributions for ordered polytomous items. Based on the results of improved item parameter recovery under RC-IRT, it is recommended that RC-IRT estimation be implemented whenever a researcher considers the construct being measured has the potential of being nonnormally distributed.

Keywords

item response theory, nonnormality, nominal response model, parameter recovery

Unidimensional item response theory (IRT) models are most frequently fitted with marginal maximum likelihood (MML; Bock & Lieberman, 1970) estimation, which

¹California State University, Fullerton, CA, USA

²University of California, Los Angeles, CA, USA

Corresponding Author:

Kathleen Suzanne Johnson Preston, Department of Psychology, CSUF, PO Box 6846, Fullerton, CA 92834-6846, USA.

Email: kpreston@fullerton.edu

assumes that (a) the examinees are independent; (b) item responses are independent, conditional on the latent trait, θ ; and (c) the probability distribution of the population of examinees must be specified prior to estimation of the item parameters (Bock & Aitkin, 1981; Bock & Lieberman, 1970). For this third assumption, any shape of the latent trait distribution can be specified; however, current computer software overwhelmingly implements a normally distributed set of quadrature points, thus creating a normal latent trait distribution. Unfortunately, the true latent trait distribution in the population is unobservable, and frequently studied psychological constructs such as depression and pain may be particularly unlikely to have a normal latent trait distribution.

Furthermore, many psychological constructs fall along a continuum, which are arguably better assessed by Likert-type response formats over dichotomous scoring, with response options such as *strongly disagree to strongly agree*, *not at all like me to very much like me*, or *never occurs to always occurs*. Polytomous IRT models allow estimation of multipoint Likert-type response formats and allow scale constructors to conduct a more informed analysis of polytomous items. Additionally, scale constructors have the opportunity to evaluate the functioning of each within-item category distinction by using a largely understudied polytomous IRT model, the nominal response model (NRM; Bock, 1972, 1997).

Fortunately, recent developments have been made in IRT estimation procedures, such as Ramsay curve item response theory (RC-IRT; Woods & Thissen, 2006), promising accurate item parameter recovery. This estimation methodology is implemented in RCLOG (Woods & Thissen, 2006) and EQSIRT (Multivariate Software, 2010); however, only EQSIRT allows RC-IRT estimation under the NRM. Further advancements in user-accessible IRT software, such as EQSIRT, permit easy and accurate IRT model estimation of nonnormal data. However, it remains undetermined whether analyzing nonnormal data under the NRM with EQSIRT using RC-IRT estimation actually produces more accurate item parameter estimates. It is, therefore, the purpose of this study to evaluate the recovery of category boundary discrimination (CBD) parameters using RC-IRT as estimated using EQSIRT for ordered categorical data under the NRM.

Effects of Nonnormality

Frequently, substantive researchers assessing psychological constructs fit IRT models to data sets without assessing the shape of the latent trait distribution and posit the parameter estimates produced will be unbiased (e.g., Cooper, Balsis, & Zimmerman, 2010; Kim, Kim, & Kamphaus, 2010; Meade, 2010; Purpura, Wilson, & Lonigan, 2010; Samuel, Simms, Clark, Livesley, & Wigider, 2010; Thomas & Locke, 2010). Contrary to the beliefs held by many substantive researchers, simulation research indicates that results from IRT models can be nontrivially biased when the true population distribution is nonnormal (Abdel-fattah, 1994; Boulet, 1996; de Ayala & Sava-Bolesta, 1999; DeMars, 2003; Kirisci, Hsu, & Yu, 2001; Seong, 1990; Stone, 1992;

van den Oord, 2005; Zwinderman & van den Wollenberg, 1990). Specifically, MML estimates of item parameters increase in bias as the distribution deviates further from normality (Boulet, 1996; Stone, 1992; Woods, 2006, 2007a, 2007b, 2008; Woods & Lin, 2009; Woods & Thissen, 2006).

Seong (1990) found parameter estimation accuracy can be improved by increasing the test length, number of quadrature points, and specifying a prior distribution that matches the true shape of the latent trait distribution. Unfortunately, the shape of the latent trait distribution in the population is never known making the implementation of these recommendations unrealistic. Additionally, most studies assessing item parameter accuracy are conducted on dichotomous IRT models, specifically the 2- or 3-parameter logistic models (Abdel-fattah, 1994; Boulet, 1996; Kirisci et al., 2001; Stone, 1992; van den Oord, 2005; Zwinderman & van den Wollenberg, 1990). In fact, to our knowledge, the only research that has investigated item parameter recovery in the NRM was conducted by de Ayala and Sava-Bolesta (1999) and DeMars (2003).

De Ayala and Sava-Bolesta (1999) manipulated a number of factors to investigate the relationship between the ratio of the sample size to the number of parameters to be estimated, latent trait distribution, maximum item information, and item parameter estimation with the NRM. Results of their simulations showed that manipulating the information and the sample size to parameter ratio had similar, but inverse, effects on both the CBD and intersection parameter recovery. Specifically, bias in item parameters increased as the maximum item information increased and as the sample size to parameter ratio decreased. Manipulating the distribution of the latent trait (i.e., skew and kurtosis) also affected the estimation accuracy of CBD parameters. They recommended increasing the sample size as one means of increasing the number examinees responding to an unattractive option as determined by the degree of nonnormality, and therefore increasing the accuracy of the item parameter estimates.

DeMars (2003) expanded on the research conducted de Ayala and Sava-Bolesta (1999) by evaluating item parameter recovery in the NRM under variations in ratio of total number of item parameters to sample size as determined by test length, number of response options, and sample size. They also manipulated the shape of the latent trait distribution and average within-item category boundary discrimination. Error variance of the item parameter estimates increased with increases in the number of response options and average within-item category discrimination. Additionally, biases in item parameters were identified when the shape of the latent trait distribution was skewed. DeMars (2003) suggests that the ratio of sample size to the number of response categories, as opposed to the ratio of sample size to total number of item parameters, is the greatest determinant of item parameter estimation accuracy.

Nominal Response Model

The NRM has the unique ability to evaluate the functioning of each within-item category distinction, termed *CBD parameters*. The size of a CBD indicates the amount

of relative information provided by adjacent response categories (e.g., the degree to which a response in Category 3 vs. Category 2 differentiates among people on the latent trait). Also estimated under the NRM are intersection parameters, which identify the point on the latent trait where an individual is likely to respond in adjacent categories. It follows that the NRM is considered the most general divide-by-total polytomous model, because this model allows for the most flexibility in parameter estimation, and all other divide-by-total models are constrained versions of the NRM (de Ayala, 2009; Ostini & Nering, 2006; Thissen & Steinberg, 1986). Preston and Reise (2013) illustrate several potentially useful applications of the NRM including exploring whether (a) CBD parameters vary within an item, (b) an item contains too many response options, and (c) response options are ordered.

The NRM can be used to model completely nominal items, multiple choice items, partially ordered items, and completely ordered items. In the NRM, the conditional probability of an individual with trait level θ responding in category x ($x = 0, \dots, m_i$) on item i can be written as

$$P_{ix}(\theta) = \frac{\exp(a_{ix}\theta + c_{ix})}{\sum_{x=1}^m \exp(a_{ix}\theta + c_{ix})}, \quad (1)$$

where, for identification, $\sum a_{ix} = \sum c_{ix} = 0$ (or, in some cases, that the parameters for the lowest response category $a_{i1} = c_{i1} = 0$). This constraint forces one response option to have a monotonically increasing CRC (the one with the most positive a), and one response option to have a monotonically decreasing CRC (the one with the lowest a). The category slope, a_{ix} , should not be confused with the CBD parameter, a^*_{ix} , which determines the amount of relative information provided by a response in category x versus responding in adjacent category $x - 1$. The category intercept parameters, c_{ix} , in Equation 1 are not intersection parameters,

$$c_j^* = \frac{c_{(x-1)} - c_x}{a_x - a_{(x-1)}}; \quad (2)$$

rather, they reflect the relative popularity of the response option. Specifically, the intersection parameters, c_j^* , identify the point on the latent trait where an individual is equally likely to respond in adjacent categories.

Ramsay Curve IRT

RC-IRT (Woods & Thissen, 2006) is a new method for fitting IRT models for which the latent trait distribution is assumed differentiable and strictly positive, but not necessarily normal. This approach combines MML/EM item parameter estimation (Bock & Aitkin, 1981; Bock & Lieberman, 1970) with a spline-based density approximation procedure described by Ramsay (2000). Specifically, in RC-IRT the distribution of the latent trait is estimated, instead of the default normal distribution as in traditional in MML. In the E step of the EM algorithm (Bock & Aitkin, 1981), the

total number of people at each quadrature point is estimated using the current characterization of the latent trait distribution. The normal distribution is used to start. In the M step, after the likelihoods for each item are maximized, the likelihood of the current characterization of latent trait distribution is maximized as approximated as a Ramsay curve using basis spline (B-spline) functions (de Boor, 2001). The shape of the Ramsay curve is determined by a combination of B-splines through the specification of the order of each B-spline polynomial function joined together at a specified number of knots. As implemented in EQSIRT, the order of the polynomial, termed *degree*, and the number of knots must be chosen by the computer user, and they determine the number of B-splines, m , used during latent trait estimation

$$m = (d + 1) + \text{number of knots} - 2, \quad (3)$$

where d is the degree of the polynomial function producing each spline (e.g., $d = 2$ produces splines with one bend; $d = 3$ produces splines with two bends; and $d = 4$ produces splines with three bends), and a knot is the point on the latent trait where separate B-splines are joined together. Woods and Thissen (2006) determined that larger order and/or more than two knots produce increasingly nonnormal Ramsay curves.

Demonstration

Here, a brief example is provided to demonstrate the severity of NRM parameter estimation bias caused by nonnormality when analyzed with MML, implementing a normal prior distribution, and the improvement in item parameter recovery when RC-IRT is implemented. For this demonstration, responses to ten 4-category items were simulated for 1,000 simulees. The CBD parameters were simulated as highly discriminating, 1.5, or moderately discriminating, 0.75, and were constrained to equality within an item. The intersection parameters were simulated as symmetric and fixed to -1.0 , 0 , and 1.0 across conditions. Figure 1 displays the true shape of the latent trait distribution that was simulated as normal, skewed (skew = 1.75, kurtosis = 6.75), and bimodal (described in detail below). Each of the six conditions was replicated 1,000 times. As seen in Table 1, the CBD parameters, highly discriminating and moderately discriminating, are recovered with excellent accuracy when the true shape of the latent trait distribution is normal. However, as the shape of the distribution deviates further from normality, the CBD parameter estimates become increasingly downwardly biased. This bias is more pronounced in items with more highly discriminating CBD parameters.

The estimated highly discriminating CBD parameters under the normal, skew, and bimodal conditions are also depicted graphically in Figure 2A, B, and C, respectively. Each plot in Figure 2 displays the 95% confidence interval around the CBD estimates for each estimated CBD parameter per item. Specifically, three CBD parameters were estimated for each 4-category item, so each item is represented by three tick-marks along the x -axis. A horizontal line is drawn across each plot to represent

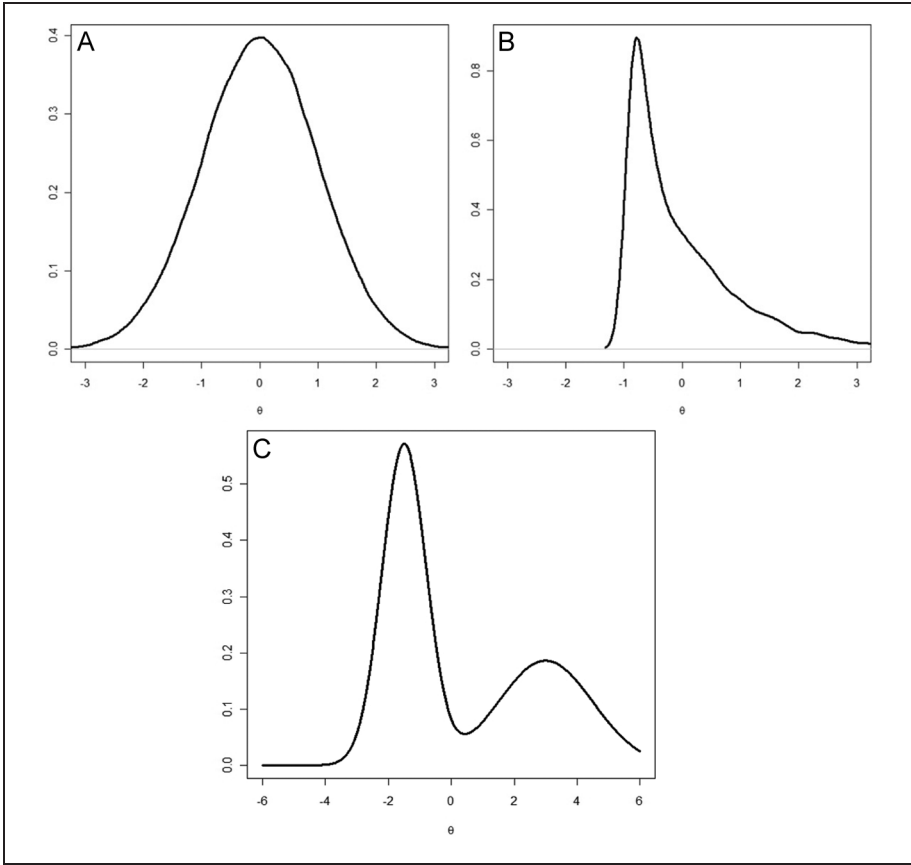


Figure 1. True shape of the latent trait: (A) standard normal distribution, (B) skewed distribution, and (C) bimodal distribution.

Table 1. Demonstrating Effects of Nonnormality Under the NRM for a 10-Item Test With 4 Response Options (1,000 replications).

Latent trait distribution	α^*	Demonstration results		
		α^* (SE)	Bias	RMSE
Normal	1.50	1.512 (0.187)	0.012	0.188
Normal	0.75	0.755 (0.138)	0.005	0.138
Skew	1.50	1.322 (0.187)	-0.178	0.595
Skew	0.75	0.679 (0.140)	0.071	0.281
Bimodal	1.50	1.465 (0.192)	0.121	0.260
Bimodal	0.75	0.755 (0.138)	0.050	0.112

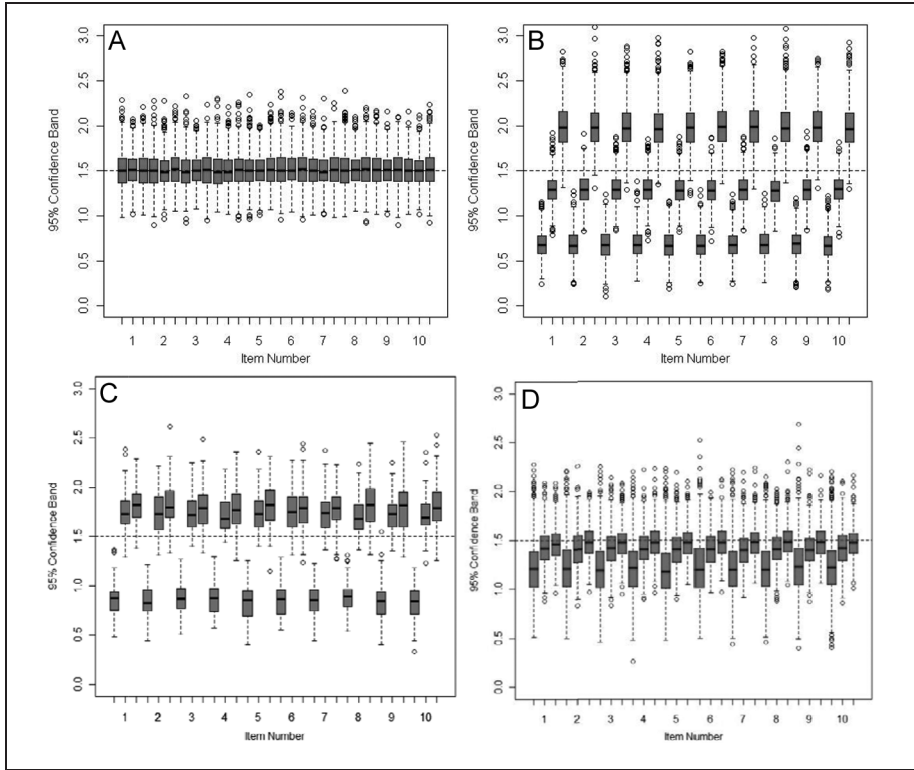


Figure 2. Average CBD parameter estimates and 95% confidence bands under (A) normal, (B) skewed, and (C) bimodal distributions. (D) The average CBD parameter estimates and 95% confidence bands under the high skew distribution estimated with RC-IRT (1,000 replications).

the true CBD parameter. Within each plot, the sampling distribution of the estimated CBD parameters for each item is displayed as a vertical 95% confidence band with the average estimated CBD parameter for the item plotted as a dot along the confidence bar. Visual inspection of this plot makes it obvious that nonnormality greatly affects item parameter recovery under the NRM.

In Figure 2A, the average estimated CBD parameters are plotted very close to the true CBD parameter and the sampling distribution is quite narrow. However, as the distribution of the latent trait increases in skew and kurtosis, the sampling distribution of the recovered CBD parameters become more variable and the average estimated CBD parameters drift further from the simulated value. Specifically, a pattern manifested under both the skew and bimodal distribution conditions such that the first CBD parameter for each item is underestimated, the second CBD parameter is fairly accurately estimated, and the third CBD parameter is consistently

overestimated. Unfortunately, since the CBD parameters were constrained to equality within the item, when averaged the estimates were averaged; the parameters appear as though they are accurately estimated. This pattern, which was more pronounced in the skewed distribution condition, is expected considering that when the distribution of the latent trait is incorrect, the expectation of the probability distribution for each of these categories is wrong.

In an effort to estimate CBD parameters more accurately, the high skew condition with highly discriminating CBD parameters was estimated under the NRM again, but using the RC-IRT estimation algorithm with the EQSIRT program defaults, order of 3 and 4 knots. The recovered CBD parameters and 95% confidence band, displayed in Figure 2D, are more accurately estimated with the average estimated CBD parameter of 1.373 (.209), average bias of the CBD parameters only $-.127$, and $RMSE = .273$. Using RC-IRT to account for and model the nonnormality of the distribution provides a marked improvement over imposing a normal prior distribution with MML estimation. Specifically, the overestimation of the third CBD parameter is greatly reduced, but the first CBD parameter is still considerably underestimated. The results of this demonstration support the importance of evaluating the recovery of CBD parameters using RC-IRT as estimated using EQSIRT for ordered categorical data under the NRM.

Method

Design

Manipulated variables were determined by prior research suggesting factors that influence item parameter recovery, producing 144 simulation conditions, which include the following: (a) within-item CBD variation; (b) sample size; (c) distribution of the latent trait, θ ; (d) order of the polynomial; and (e) number of knots. All conditions were simulated with 10 items containing 4 response options each, and CBD parameters were manipulated consistently across all conditions. The distances between intersection parameters were manipulated according to the distribution of θ , as described below. The conditions with a normally distributed latent trait, θ , estimated with order of 2 and 2 knots (Woods, 2006) are considered the null conditions because it is the same as assuming a normal distribution and all results will be compared with those conditions.

Factors Influencing Item Parameter Recovery

CBD Parameters. Adapting Preston and Reise (2013), the average size of the CBD parameter was set to 1.00, indicating moderate discrimination. To create items with variation in the CBD parameters, values were drawn randomly from a uniform distribution. The minimum and maximum values of the distribution originated at 1.00, a moderate discrimination, which constrained the within-item CBD parameters to equality creating data under the generalized partial credit model. The maximum

value of the distribution increased by .25 and the minimum values of the distribution decreased by .25, creating variation in within item CBD parameter that increased in increments of .5 for each condition. The range of the distribution increased by increments of .5 until the minimum value of the distribution reached .25.

Intersection parameters. DeMars (2003) showed the distance between the intersection parameters influences item parameter recovery under the NRM. Taking into consideration how the shape of the latent trait distribution partly determines the response distribution, the category intersection parameters were generated based on the shape of the latent trait. Specifically, under the normal and bimodal distributions intersections were determined by the 20th, 50th, and 70th percentiles of the simulated distribution. However, it is reasonable to assume that, under a skewed distribution, the responses would follow the nonnormality of the latent trait distribution. Therefore, for the skewed distribution conditions, the intersections were determined by the 50th, 80th, and 95th percentiles of the simulated distribution.

Sample size. De Ayala and Sava-Bolesta (1999) and DeMars (2003) consistently found that the sample size to response option ratio is an important factor in the accuracy of item parameter recovery in the nominal response model. Therefore, two sample size conditions, $N = 500$ and $N = 2,000$, were considered, because sample size is well known to affect the accuracy of but not necessarily bias item parameter estimation (de Ayala & Sava-Bolesta, 1999; DeMars, 2003). The smaller sample size ($N = 500$) represents the general number of participants frequently included in a psychological study, and the larger sample size ($N = 2,000$) is representative of larger sets of individuals responding to educational tests.

Latent trait distribution. Three latent trait distributions were considered because the focus of the present research is to determine the degree to which the latent trait distribution affects item parameter recovery in the NRM (see Figure 1). The shape of the normal and skewed distributions was determined by Fleishman's (1978) power method weights. In this article, Fleishman presents a table of power method weights, (a , b , c , and d) that can correspond to particular skew and kurtosis of the desired distribution to be used in simulating nonnormal distributions. These values are applied to the polynomial transformation equation $Y = a + bX + cX^2 + dX^3$. Figure 1A and B displays the shape of the true latent trait distribution, θ , as standard normal (skew = 0.0, kurtosis = 3.0, $a = 0.0$, $b = 1.0$, $c = 0.0$, $d = 0.0$), and skewed (skew = 1.75, kurtosis = 6.75, $a = -0.39949667453766$, $b = 0.92966052480111$, $c = 0.39949667453766$, $d = -0.03646699281275$). The highly skewed condition uses coefficients that are within the range of the latent distribution estimated by Woods and Thissen (2006) for dichotomous items related to panic disorder with skew of 1.04 and kurtosis of 6.53.

Fleishman's (1978) power method weights does not address bimodal distributions, so the bimodal distribution was determined by a mixture of normal distributions (where μ = mean, σ = standard deviation, and mp = mixing proportion). Figure 1C

displays the shape of the true latent trait distribution, θ , as bimodal ($\mu_1 = -1.5$, $\mu_2 = 3.0$, $\sigma_1 = 0.7$, $\sigma_2 = 1.5$, $mp_1 = 1.0$, $mp_2 = 0.7$). The bimodal latent trait distribution represents a scenario typically seen in health outcomes research (e.g., Hanger, Fogarty, Wilkinson, & Sainsbury, 2000), where there is a large group of individuals who are considered “normal” and score low, and there is a small group of “patients” who score high on the latent trait but are normally distributed within that group. The distribution was standardized ($\mu = 0.0$, $\sigma = 1.0$) prior to data simulation per direction from Sam He (personal communication, November 17, 2010).

Order of the polynomial and knots. Since the implementation of the NRM with EQSIRT using RC-IRT has never been evaluated, this research will manipulate pre-determined combinations of polynomial order and knots to evaluate the consistency and accuracy of the CBD parameter estimates. In EQSIRT, RC-IRT is adapted as a type of MML/EM parameter estimation method. This implementation borrows the idea of de Boor’s B-spline curve fitting algorithm, which requires two pieces of information to normalize the curve to the population density: (a) order of the polynomial and (b) number of knots. The degree, as termed in EQSIRT, is the order of the polynomial of the B-splines curve that is computed during the parameter estimation. The knots are the number of joints of the B-splines curve. These joints are presented in the form of parameters that are estimated simultaneously during parameter estimation. As implemented in EQSIRT, the user may choose either order 2 or 3 for the order of the polynomial and between 2 and 10 knots. Because the order of the polynomial and knots are related to the shape of the latent trait distribution, which influences the recovery of item parameters, data were estimated with order 2 and 3, exhausting all EQSIRT implemented options. Data were further estimated with 2, the minimum option in EQSIRT; 4, the default option in EQSIRT; or 10 knots, the maximum option in EQSIRT.

Data Generation

For each condition, 1,000 data sets were generated. All data sets were generated under the NRM and item parameters were estimated under the NRM via RC-IRT as implemented in EQSIRT (Multivariate Software, 2010).

Model Fitting

The newly developed EQSIRT (Multivariate Software, 2010) was used for all conditions estimated under RC-IRT. Each data set was fitted with the NRM using mostly the program defaults, except the maximum number of EM cycles was set to 999. The output of EQSIRT contains a and c parameters as calculated in Equation 1; therefore, the a and c parameters are converted into the CBD parameters and category intersections using the formulas described above. As mentioned previously, estimation of RC-IRT in EQSIRT requires the user to specify the order of the polynomial and the

number of knots or joints for the b-splines curve. The default number of order = 3 (range = 2-3), and knots = 4 (range = 2-10).

Outcome Measure

This study focuses exclusively on the evaluation of CBD parameter recovery, which is considered the more useful parameter. Additionally, the intersection parameters are a function of the CBD parameters, so the recovery of the intersections was not explicitly evaluated. CBD parameter recovery was evaluated in several ways (de Ayala & Sava-Bolesta, 1999; Woods, 2006): (a) mean and standard deviation of the recovered parameters, (b) average absolute bias, and (c) a computed index of the absolute difference between the true and estimated test characteristic curves (TCCs). First, each item's characteristic curve (ICC) was computed based on CRC and ICC relationships discussed in Embretson and Reise (2000), the CRC computed for each of the four categories per item for Equation 1 is multiplied by 0, 1, 2, and 3, respectively, creating an ICC. Finally, the ICCs within a test were summed over to create a TCC for each test. The expected value of the true and estimated TCC was evaluated at 60 points evenly spaced between -3 and 3 . The difference index was computed by averaging over the sum of the absolute differences between the true and estimated expected values at each of the 60 points.

Results

The results of the simulation study are presented in Tables 2 through 6, where each table is devoted to presenting results for all conditions as summarized by one of the outcome measures. The first three columns of each of these tables list order of the polynomial, number of knots, and range of CBD, respectively. The remaining columns list the outcome measure under the normal distribution, skewed distribution, and bimodal distribution for each sample size condition (500, 2,000), respectively. Each outcome measure is discussed separately. The first four rows of each table provide results for conditions estimated using order 2 and 2 knots, specifying a normal distribution under RC-IRT estimation, which produces results equivalent to those obtained without implementing RC-IRT estimation. Therefore, under the normal distribution conditions, these first four rows provide information about the accuracy of item parameter recovery when the estimated latent trait distribution matches the true latent trait distribution, the null condition, producing ideal outcome measures. The remaining rows of each table under the normal distribution demonstrate the consequence to CBD parameter estimation when using RC-IRT estimation unnecessarily. However, under the skewed and bimodal distributions, these first four rows produce results when the shape of the latent trait distribution is assumed normal, demonstrating the effect of ignoring nonnormality on the outcome measures. The remaining rows of each table under the skewed and bimodal latent trait distributions

Table 2. Average CBD Parameter Recovery and CBD Parameter Standard Deviation.

Order	Knots	CBD range	CBD mean (SD)								
			Normal distribution			Skewed distribution			Bimodal distribution		
			N = 500	2,000	500	2,000	500	2,000			
2	2	1.00-1.00	1.110 (.134)	1.008 (.121)	1.127 (.285)	1.100 (.143)	1.076 (.262)	1.060 (.132)			
2	2	0.75-1.25	1.110 (.215)	1.008 (.193)	1.122 (.335)	1.093 (.221)	1.076 (.307)	1.057 (.201)			
2	2	0.50-1.50	1.117 (.359)	0.999 (.313)	1.119 (.448)	1.109 (.365)	1.068 (.406)	1.059 (.330)			
2	2	0.25-1.75	1.135 (.528)	1.000 (.457)	1.130 (.598)	1.109 (.538)	1.067 (.536)	1.048 (.472)			
2	4	1.00-1.00	1.008 (.115)	1.001 (.126)	1.011 (.249)	0.997 (.121)	1.083 (.368)	1.018 (.135)			
2	4	0.75-1.25	1.006 (.192)	1.000 (.192)	1.013 (.294)	0.995 (.185)	1.083 (.443)	1.023 (.206)			
2	4	0.50-1.50	1.004 (.318)	1.000 (.316)	1.021 (.420)	1.005 (.320)	1.111 (.639)	1.040 (.355)			
2	4	0.25-1.75	1.005 (.494)	1.011 (.462)	1.024 (.646)	0.984 (.445)	1.112 (.795)	1.042 (.501)			
2	10	1.00-1.00	0.980 (.117)	0.939 (.126)	0.974 (.442)	0.919 (.126)	1.055 (.405)	1.003 (.128)			
2	10	0.75-1.25	0.991 (.191)	0.940 (.187)	0.979 (.434)	0.926 (.186)	1.049 (.410)	1.005 (.202)			
2	10	0.50-1.50	0.986 (.317)	0.948 (.301)	0.979 (.631)	0.924 (.296)	1.058 (.583)	1.000 (.324)			
2	10	0.25-1.75	0.981 (.440)	0.953 (.423)	1.032 (.056)	0.923 (.413)	1.083 (.807)	1.027 (.482)			
3	2	1.00-1.00	1.009 (.117)	0.996 (.124)	1.022 (.253)	1.000 (.118)	1.088 (.423)	1.018 (.131)			
3	2	0.75-1.25	1.010 (.185)	0.994 (.190)	1.015 (.329)	0.991 (.185)	1.077 (.429)	1.021 (.225)			
3	2	0.50-1.50	1.009 (.319)	1.000 (.317)	1.016 (.410)	0.990 (.314)	1.089 (.583)	1.035 (.354)			
3	2	0.25-1.75	1.019 (.461)	0.993 (.460)	1.033 (.777)	0.983 (.447)	1.119 (.783)	1.056 (.525)			
3	4	1.00-1.00	1.018 (.120)	1.002 (.125)	1.018 (.253)	0.993 (.114)	1.076 (.323)	1.028 (.131)			
3	4	0.75-1.25	1.018 (.188)	1.003 (.196)	1.009 (.294)	0.986 (.177)	1.078 (.399)	1.028 (.211)			
3	4	0.50-1.50	1.014 (.326)	0.996 (.312)	1.022 (.404)	0.996 (.319)	1.083 (.496)	1.039 (.349)			
3	4	0.25-1.75	1.015 (.467)	0.988 (.455)	1.046 (.680)	0.990 (.452)	1.121 (.735)	1.031 (.501)			
3	10	1.00-1.00	0.980 (.120)	0.930 (.130)	0.979 (.368)	0.919 (.123)	1.044 (.337)	1.002 (.124)			
3	10	0.75-1.25	0.981 (.189)	0.941 (.186)	0.970 (.431)	0.929 (.179)	1.046 (.422)	1.003 (.201)			
3	10	0.50-1.50	0.979 (.312)	0.941 (.299)	0.997 (.617)	0.916 (.295)	1.067 (.625)	0.991 (.331)			
3	10	0.25-1.75	0.986 (.440)	0.957 (.418)	1.036 (.926)	0.915 (.446)	1.099 (.867)	1.001 (.469)			

Table 3. Average Proportion of Converged Replications.

Order	Knots	CBD range	Proportion converged					
			Normal distribution		Skewed distribution		Bimodal distribution	
			N = 500	2,000	500	2,000	500	2,000
2	2	1.00-1.00	1.000	1.000	0.996	1.000	1.000	1.000
2	2	0.75-1.25	1.000	1.000	1.000	1.000	1.000	1.000
2	2	0.50-1.50	1.000	1.000	0.996	1.000	0.992	1.000
2	2	0.25-1.75	0.980	1.000	0.988	1.000	0.992	1.000
2	4	1.00-1.00	1.000	1.000	1.000	1.000	1.000	1.000
2	4	0.75-1.25	1.000	1.000	1.000	1.000	1.000	1.000
2	4	0.50-1.50	1.000	1.000	1.000	1.000	0.996	1.000
2	4	0.25-1.75	0.992	0.990	0.988	1.000	0.980	1.000
2	10	1.00-1.00	0.980	1.000	0.944	0.940	1.000	1.000
2	10	0.75-1.25	0.992	0.990	0.980	0.920	1.000	0.980
2	10	0.50-1.50	0.988	1.000	0.952	0.900	0.984	0.980
2	10	0.25-1.75	0.988	1.000	0.988	0.970	0.980	0.990
3	2	1.00-1.00	1.000	1.000	1.000	1.000	1.000	1.000
3	2	0.75-1.25	1.000	1.000	1.000	1.000	1.000	1.000
3	2	0.50-1.50	1.000	1.000	1.000	1.000	1.000	1.000
3	2	0.25-1.75	0.976	1.000	0.976	1.000	0.988	1.000
3	4	1.00-1.00	1.000	1.000	1.000	1.000	1.000	1.000
3	4	0.75-1.25	1.000	1.000	1.000	1.000	1.000	1.000
3	4	0.50-1.50	1.000	1.000	1.000	1.000	0.992	1.000
3	4	0.25-1.75	0.992	1.000	0.980	1.000	0.988	0.990
3	10	1.00-1.00	0.996	0.990	0.952	0.980	0.996	0.980
3	10	0.75-1.25	0.996	1.000	0.968	0.960	0.992	1.000
3	10	0.50-1.50	0.988	1.000	0.952	0.940	0.988	1.000
3	10	0.25-1.75	0.964	0.990	0.956	0.940	0.984	1.000

demonstrate the improvement in CBD parameter and θ score recovery when the shape of the latent trait distribution is estimated using RC-IRT.

Descriptive Statistics

Table 2 contains the means and standard deviations of the recovered CBD parameters. As expected, the standard deviations of the CBD parameters increased as the range of the CBD parameters increased. Specifically, when CBD parameters were all constrained to 1.0, producing the generalized partial credit model (Muraki, 1992), the standard deviation averaged over all distribution, sample size, order of the polynomial, and number of knots conditions was 0.170, which increased as the CBD parameter variation increased to $SD = 0.231$ with CBD parameters ranging from 0.75 to 1.25, to $SD = 0.360$ with CBD parameters ranging from 0.50 to 1.50, and to $SD = 0.539$ with CBD parameters ranging from 0.25 to 1.75. The standard deviations of

Table 4. CBD Absolute Bias.

Order	Knots	CBD range	CBD absolute bias					
			Normal distribution		Skewed distribution		Bimodal distribution	
			N = 500	2,000	500	2,000	500	2,000
2	2	1.00-1.00	0.022	0.008	0.127	0.100	0.076	0.060
2	2	0.75-1.25	0.019	0.007	0.123	0.092	0.077	0.060
2	2	0.50-1.50	0.021	0.001	0.119	0.107	0.068	0.056
2	2	0.25-1.75	0.021	0.008	0.129	0.118	0.069	0.047
2	4	1.00-1.00	0.025	0.001	0.011	-0.003	0.083	0.018
2	4	0.75-1.25	0.023	0.000	0.013	-0.008	0.085	0.021
2	4	0.50-1.50	0.024	0.000	0.025	-0.004	0.116	0.035
2	4	0.25-1.75	0.034	0.004	0.031	-0.013	0.123	0.041
2	10	1.00-1.00	-0.030	-0.061	-0.026	-0.081	0.055	0.003
2	10	0.75-1.25	-0.031	-0.060	-0.023	-0.072	0.048	0.005
2	10	0.50-1.50	-0.026	-0.052	-0.016	-0.070	0.059	0.006
2	10	0.25-1.75	-0.020	-0.062	0.030	-0.078	0.082	0.016
3	2	1.00-1.00	0.023	-0.004	0.022	0.000	0.088	0.018
3	2	0.75-1.25	0.009	-0.006	0.015	-0.006	0.076	0.025
3	2	0.50-1.50	0.012	-0.002	0.019	-0.012	0.090	0.037
3	2	0.25-1.75	0.017	-0.003	0.034	-0.009	0.117	0.054
3	4	1.00-1.00	0.023	0.002	0.018	-0.007	0.076	0.028
3	4	0.75-1.25	0.032	0.003	0.010	-0.014	0.077	0.027
3	4	0.50-1.50	0.037	0.002	0.021	-0.002	0.085	0.036
3	4	0.25-1.75	0.041	0.002	0.039	-0.008	0.123	0.040
3	10	1.00-1.00	-0.030	-0.070	-0.021	-0.081	0.044	0.002
3	10	0.75-1.25	-0.036	-0.061	-0.029	-0.077	0.047	0.005
3	10	0.50-1.50	-0.033	-0.058	-0.005	-0.073	0.063	0.001
3	10	0.25-1.75	-0.018	-0.058	0.031	-0.076	0.101	0.005

the CBD parameters remained fairly consistent under the large sample size condition ($N = 2,000$) regardless of distribution condition (Mean $SD = 0.268$ under normal distribution; Mean $SD = 0.272$ under skewed distribution; Mean $SD = 0.293$ under bimodal distribution), whereas the standard deviation of the CBD parameters were considerably more variable under the small sample size condition ($N = 500$; Mean $SD = 0.278$ under normal distribution; Mean $SD = 0.483$ under skewed distribution; Mean $SD = 0.516$ under bimodal distribution). The standard deviations of the CBD parameters remained consistent for all order of the polynomial and knots.

Table 3 displays the proportion of converged replications for each condition, which were generally excellent. Proportion of replications converged was high over all conditions with 0.90 being the minimum number of converged replications occurring under the skewed distribution estimated using order of 2 and 10 knots with CBD parameters varying from 0.50 to 1.50. Overall, the conditions estimated with 10 knots consistently resulted in the lowest proportion converged, which was magnified under the small sample size condition ($N = 500$).

Table 5. CBD Absolute Bias for Low, Moderate, and High CBD Parameters.

CBD range	Absolute bias					
	Normal distribution		Skewed distribution		Bimodal distribution	
	N = 500	2,000	500	2,000	500	2,000
0.25-0.75	0.009	0.002	0.051	0.042	0.046	0.034
0.75-1.25	0.013	0.010	0.119	0.123	0.077	0.041
1.25-1.75	0.041	0.011	0.217	0.191	0.085	0.065
0.25-0.75	0.012	0.000	0.010	-0.005	0.024	0.010
0.75-1.25	0.031	-0.002	0.007	-0.014	0.081	0.025
1.25-1.75	0.058	0.014	0.077	-0.022	0.027	0.089
0.25-0.75	-0.010	-0.023	-0.022	-0.031	0.011	0.005
0.75-1.25	-0.028	-0.059	-0.010	-0.077	0.053	0.003
1.25-1.75	-0.023	-0.100	0.120	-0.126	0.182	0.039
0.25-0.75	0.007	-0.010	-0.002	-0.002	0.015	0.004
0.75-1.25	0.014	0.000	0.013	-0.015	0.088	0.040
1.25-1.75	0.029	0.003	0.092	-0.011	0.248	0.116
0.25-0.75	0.012	0.000	0.006	-0.001	0.026	0.008
0.75-1.25	0.028	0.000	0.012	-0.015	0.084	0.038
1.25-1.75	0.082	0.008	0.096	-0.007	0.260	0.076
0.25-0.75	-0.021	-0.019	-0.021	-0.034	0.019	-0.003
0.75-1.25	-0.019	-0.052	-0.026	-0.075	0.073	0.004
1.25-1.75	-0.013	-0.100	0.135	-0.121	0.214	0.014

CBD Parameter Recovery

Absolute bias. Table 4 displays the absolute bias in CBD parameter recovery for each condition, with values near zero indicating no bias. As expected, parameters are less biased under the larger sample size condition (mean bias = -0.001) than the small sample size condition (mean bias = 0.039). The CBD parameters were recovered accurately under the normal distribution, the null condition, with order of 2 and 2 knots with an average bias of 0.013; however, the bias is reduced even further to 0.006 when estimated with order of 3 and 2 knots. Ignoring nonnormality under the bimodal distribution produces considerable positive bias in the CBD estimates (mean bias = 0.064), which is even more pronounced under the skewed distribution (mean bias = 0.114).

Focusing on the skewed distribution, the average bias in CBD parameter estimates is greatly reduced to 0.007 when estimated using order of 3 and 4 knots, and even further reduced to 0.006 with order of 2 and 4 knots averaging over sample size and CBD variation. RC-IRT estimation using order of 2 and 3 with 10 knots reduced the bias to -.042 and -.041, respectively, but the parameter estimates became downwardly biased. The bias in the CBD parameter estimates is smallest for the bimodal distribution condition when estimated using order of 2 and 4 knots. RC-IRT estimation using order of 2 and 3 with 10 knots reduced the bias when compared with

Table 6. Difference Index of True and Estimated TCCs.

Order	Knots	CBD range	TCC difference index					
			Normal distribution		Skewed distribution		Bimodal distribution	
			N = 500	2,000	500	2,000	500	2,000
2	2	1.00-1.00	0.286	0.152	1.034	0.989	0.469	0.437
2	2	0.75-1.25	0.307	0.141	1.003	0.978	0.474	0.444
2	2	0.50-1.50	0.300	0.145	1.005	0.965	0.476	0.429
2	2	0.25-1.75	0.299	0.139	0.993	0.950	0.492	0.419
2	4	1.00-1.00	0.342	0.170	0.464	0.281	0.336	0.167
2	4	0.75-1.25	0.328	0.164	0.440	0.270	0.346	0.141
2	4	0.50-1.50	0.322	0.152	0.412	0.278	0.337	0.193
2	4	0.25-1.75	0.360	0.168	0.427	0.292	0.336	0.196
2	10	1.00-1.00	0.400	0.338	0.532	0.469	0.311	0.160
2	10	0.75-1.25	0.409	0.339	0.525	0.430	0.332	0.169
2	10	0.50-1.50	0.387	0.312	0.540	0.440	0.319	0.154
2	10	0.25-1.75	0.388	0.321	0.545	0.451	0.320	0.160
3	2	1.00-1.00	0.287	0.162	0.401	0.197	0.352	0.160
3	2	0.75-1.25	0.299	0.155	0.390	0.213	0.326	0.166
3	2	0.50-1.50	0.303	0.162	0.388	0.227	0.342	0.189
3	2	0.25-1.75	0.304	0.153	0.433	0.223	0.325	0.191
3	4	1.00-1.00	0.338	0.163	0.400	0.257	0.343	0.181
3	4	0.75-1.25	0.333	0.165	0.420	0.270	0.326	0.171
3	4	0.50-1.50	0.346	0.157	0.400	0.255	0.331	0.192
3	4	0.25-1.75	0.349	0.176	0.429	0.245	0.344	0.186
3	10	1.00-1.00	0.397	0.369	0.542	0.468	0.322	0.160
3	10	0.75-1.25	0.407	0.336	0.528	0.447	0.321	0.166
3	10	0.50-1.50	0.402	0.317	0.531	0.437	0.319	0.180
3	10	0.25-1.75	0.385	0.317	0.561	0.450	0.327	0.171

ignoring the shape of the distribution to -0.042 and -0.041 , but as with the normal distribution, the parameter estimates became downwardly biased.

Across nonnormal distribution conditions, RC-IRT estimation with order of 2 or 3 and 4 knots produced the most accurate CBD parameter estimates in terms of absolute bias. Overall, estimation with either order of 2 or 3 and 10 knots produced CBD parameter estimates that were slightly downwardly biased (overall mean bias = -0.018), indicating that estimating distributions with 10 knots may overcharacterize the distribution, focusing on unimportant nuances, when the distributions are a more conventional, yet nonnormal, shape. The size of the absolute bias did not change as a function of CBD range.

In evaluating the absolute bias values averaged over all CBD values, some misestimation may be missed for the most extreme CBD values. Therefore, to further examine potential misestimation in extreme CBD values, the true CBD parameters and the corresponding estimates from the condition with the most widely varying CBD values

were divided into low, moderate, and high at CBD values ranging from 0.25 to 0.75, 0.75 to 1.25, and 1.25 to 1.75, respectively. Table 5 presents these low, moderate, and high absolute bias values. Overall, the estimation of the CBD parameters is best for the low range at 0.004, and worsens as the category becomes more discriminating to 0.016 for moderate, and 0.050 for high CBD parameters. These findings are consistent with those from the demonstration study, which compared moderate CBD values of 0.75 to high CBD values of 1.5, and found inaccuracies in estimation were more pronounced for the high CBD values. As mentioned above, RC-IRT estimation using order of 2 and 4 knots produced the most accurate estimates as measured by absolute bias for the skewed. The optimal RC-IRT combination of polynomial order and knots appears to be mainly determined by the accuracy of the estimation of the high CBD values because the corresponding absolute bias value was at the minimum of 0.077 the small sample size condition, and -0.022 for the large sample size condition. These findings suggest that the more highly discriminating category distinctions are more important in determining optimal RC-IRT polynomial order and knots combination for estimation.

TCC difference index. Table 6 presents the absolute difference between the true and estimated TCCs. For this difference index, values close to zero indicate the estimated and true TCCs overlap entirely.

Overall, CBD variation did not affect the magnitude of the difference index. However, the differences in sample size condition were quite pronounced, with the small sample size condition producing an average difference index value of .0419 and the large sample size condition producing a value of 0.289.

Figure 3A, 3B, and 3C shows the true TCC for the normal, skewed, and bimodal distribution conditions, respectively, and the TCC estimated under RC-IRT using order of 2 and 2 knots, effectively ignoring the true shape of the latent trait distribution. Figure 3B and C also displays the TCC estimated under RC-IRT using the optimal order and knots as determined by the TCC difference index, which is order of 3 and 2 knots for Figure 3B and order of 3 and 10 knots for Figure 3C. The solid line represents the true TCC, the dashed line represents the estimated TCC under order of 2 and 2 knots, and the dotted line represents the estimated TCC under the optimal order and knots. As can be seen in Figure 3A, the true and estimated TCCs overlap almost entirely, which indicates excellent parameter recovery at the test level. Consequently, under the normal distribution condition, the TCC difference index was at the minimum value for all conditions of 0.221 averaged over sample size conditions, indicating that true item parameters and estimated item parameters did not differ in difficulty when estimated using a normal distribution. The TCC difference index remained consistently small, but inflated slightly when estimated using any other combination of order and knots, indicating that overparameterizing the distribution worsens the accuracy of the model at the test level.

As mentioned above, Figure 3B displays the true, estimated ignoring nonnormality, and optimal RC-IRT recovered TCCs under the skewed distribution. As can be

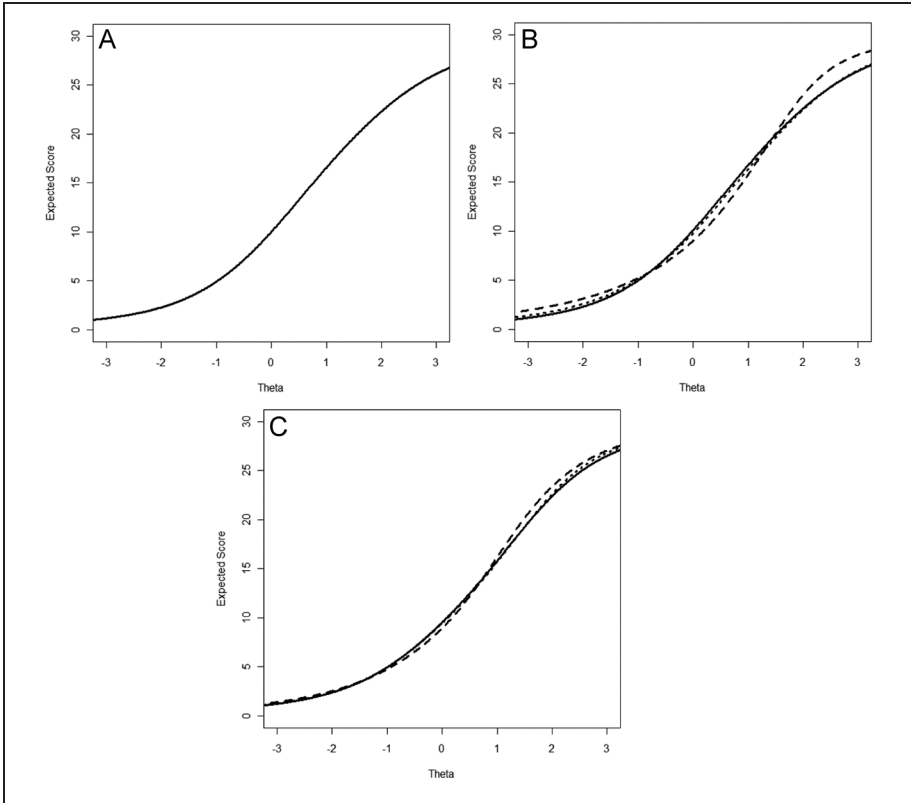


Figure 3. True, estimated, and recovered test characteristic curves (TCCs) averaged over sample size and CBD variation conditions under the (A) normal distribution condition, (B) skewed distribution condition, and (C) bimodal distribution condition. True TCC is represented by the solid line, the estimated TCC using order of 2 and 2 knots is represented by the dashed line, and the optimum polynomial order and knot combination recovered TCC is represented by the dotted line.

seen, the TCC was underestimated at the negative extreme end of θ and was overestimated at the positive extreme end of θ . Accordingly, the TCC difference index was extremely large at 0.990, averaged over sample size and CBD variation conditions. Using RC-IRT with any combination of order and knots to account for the nonnormality of the distribution improved the accuracy of the estimated TCC, but RC-IRT estimation with order of 3 and 2 knots reduced the TCC difference index the most to 0.309. Visual inspection of the Figure 3B shows that the true TCC and the TCC estimated under RC-IRT using order of 3 and 2 knots are nearly indistinguishable.

Figure 3C displays the true, estimated ignoring nonnormality, and optimal RC-IRT recovered TCCs under the bimodal distribution. Similar to the skewed

distribution condition, the TCC was underestimated at the negative extreme end of θ and was overestimated at the positive extreme end of θ . However, in the bimodal distribution condition, the over- and underestimation was noticeably smaller, resulting in moderate TCC difference index values of 0.455. As in the skewed distribution condition, estimation with any combination of order and knots under RC-IRT reduced the TCC difference index, but estimation with order of 2 and 4 knots produced the most accurate TCCs with a TCC difference index value of 0.241.

Figure 4A to F displays the true versus recovered TCCs as estimated under RC-IRT using all combinations of order and knots for the skewed distribution condition. The true TCC is represented by the solid line and the recovered TCC is represented by the dashed line. Comparing Figure 4A to Figure 4B through F illustrates that merely estimating under RC-IRT is more critical to accurate item parameters than determining the appropriate number of order and knots. As can be seen in Figure 4A, estimation of a test, while ignoring the nonnormality, results in under- and overestimation of the item parameters, especially at the extremes of the trait level. In Figure 4B through F, the TCCs overlap almost entirely, regardless of combination of order and knots, showing how important it is to use RC-IRT estimation when the measured construct is nonnormally distributed in the population. However, careful visual inspection of Figure 4E and F supports the general finding that estimation using 10 knots decreases the accuracy of the item parameters slightly, but are still considerably more accurate than ignoring that nonnormality in the distribution.

Discussion

Polytomous IRT models are becoming increasingly popular as a method of constructing and evaluating educational and psychological measurement instruments. In its increasing popularity, polytomous IRT models are being used to study psychological and educational constructs that are likely not normally distributed in the population. Moreover, these assessments are being used to make high-stakes decisions about individuals, such as a clinical diagnosis or whether or not a child qualifies for special education. Unfortunately, the inherent nonnormality of these constructs is frequently ignored during estimation using the MML/EM estimation algorithm, potentially leading to an incorrect diagnosis or incorrectly placing a child in a special education classroom.

This research provided a comprehensive demonstration of the effects of ignoring nonnormality in the distribution of the latent trait under the NRM. The NRM was chosen for this demonstration because it has the unique ability to evaluate response category functioning, allowing researchers to conduct in-depth item analyses to identify items containing too many response options, unordered response options, or poorly functioning categories that do not provide useful psychometric information. To determine the effects of nonnormality on CBD parameter estimation and latent trait distribution recovery, data were generated under normal, skewed, and bimodal distributions. These data were estimated under RC-IRT in EQSIRT using order of 2

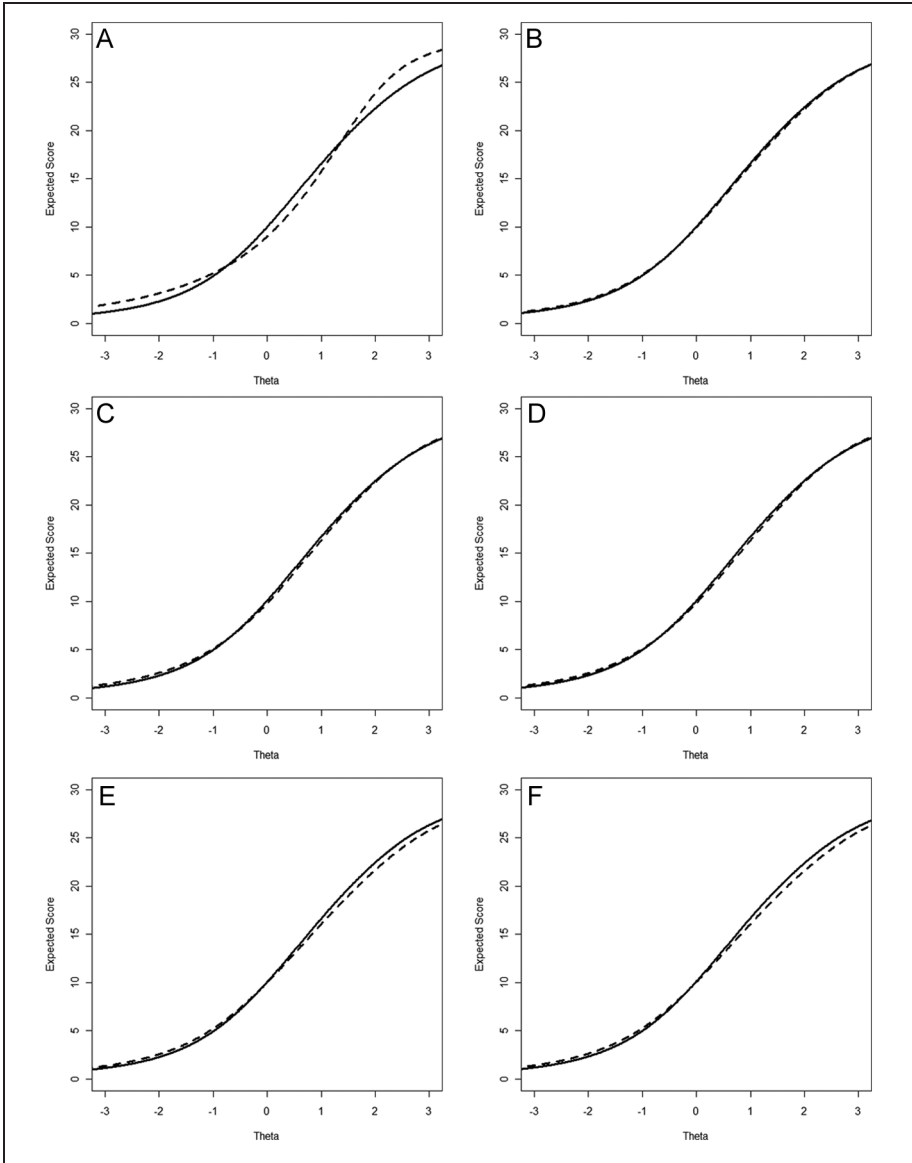


Figure 4. True versus estimated TCCs under the skewed condition for all combinations of polynomial order and knots: (A) 2–2, (B) 3–2, (C) 2–4, (D) 3–4, (E) 2–10, and (F) 3–10. True TCC is represented by the solid line and the estimated TCC by the dashed line.

and 2 knots, effectively ignoring the shape of the true latent trait distribution. Results showed that CBD parameters were grossly inaccurate and upwardly biased leading

to large differences between estimated and true item parameters. Generally, consequences of ignoring the nonnormality of the latent trait distribution were greater under a skewed distribution, and slightly less so under a bimodal distribution. Inaccuracies of CBD parameter estimates were magnified under a small sample size. Overall, the consequences of ignoring nonnormality in the CBD parameter estimates were severe.

Fortunately, recent developments in IRT estimation procedures, such as RC-IRT (Woods & Thissen, 2006), and user-friendly software, such as EQSIRT (Multivariate Software, 2010), improves the estimation of CBD parameters. The primary focus of this research was to conduct an extensive evaluation of item parameter recovery using EQSIRT's implementation of RC-IRT for ordered categorical data under the NRM. To study the improvement of estimation when accounting for the shape of the latent trait distribution, the normal, skewed, and bimodal distributions were estimated using a variety of combinations of polynomial order and knots. Generally, CBD parameter recovery was greatly improved when the nonnormality of the distributions was estimated using any of the studied combinations of order and knots. Specifically, using either 2 or 4 knots improved CBD recovery the most when compared with estimation using 10 knots. These results indicate that, under RC-IRT, estimating more parameters than necessary to account for the nonnormality in the latent trait distribution actually worsens the accuracy of the estimation. Under a skewed distribution, generally the most accurate CBD parameter estimates were obtained using order of 2 and 4 knots, whereas CBD parameters under the bimodal distribution were most accurately recovered when estimated using order of 3 and 4 knots. These findings are consistent with Woods (2006), that knowledge about the shape of the distribution can enhance estimation precision in the recovery of the CBD parameter estimates. The current results demonstrate consistently improved CBD parameter accuracy under the NRM while using predetermined combinations of polynomial order and knots; therefore, future research will use a model selection algorithm to determine the optimal combination of polynomial order and knots for each generated dataset and to improve the external validity of the research. Additionally, future research will evaluate the influence of complimentary factors known to affect the accuracy of item parameter recovery such as the number of items, the number of response options, and additional nonnormal latent trait distributions.

Based on these results, it is recommended that RC-IRT estimation be implemented whenever a researcher considers the construct being measured has the potential of being nonnormally distributed in the studied population. As demonstrated, ignoring the true shape of the latent trait distribution creates gross inaccuracies in the estimation of the item parameters. Moreover, the benefits to implementing RC-IRT estimation in the accuracy of item parameters are great, and there is very little consequence to implementing RC-IRT estimation when the latent trait is actually normally distributed. Finally, it is recommended that unless theory suggests otherwise, researchers should use combinations of order of 2 or 3 and 2 or 4 knots in applications because these combinations resulted in the most consistently accurate estimates.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institutes of Health, the NIH Roadmap for Medical Research Grant (AR052177; PI: David Cella), and a predoctoral advanced quantitative methodology training grant (#R305B080016) awarded to UCLA by the Institute of Education Sciences of the U.S. Department of Education.

References

- Abdel-fattah, A. A. (1994, April). *Comparing BILOG and LOGIST estimates for normal, truncated normal, and beta ability distributions*. Paper presented at the annual meeting of the American Education Research Association, New Orleans, GA.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29-51.
- Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33-49). New York, NY: Springer.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443-459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored item. *Psychometrika*, *35*, 179-197.
- Boulet, J. R. (1996). *The effect of nonnormal ability distributions on IRT parameter estimation using full-information and limited-information methods (item response theory, nonlinear factor analysis)*. Dissertation abstracts online. University of Ottawa, Canada.
- Cooper, L. D., Balsis, S., & Zimmerman, M. (2010). Challenges associated with a polythetic diagnostic system: Criteria combinations in the personality disorders. *Journal of Abnormal Psychology*, *119*, 886-895.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- de Ayala, R. J., & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, *23*, 3-19.
- de Boor, C. (2001). *A practical guide to splines*. New York, NY: Springer-Verlag.
- DeMars, C. E. (2003). Sample size and the recovery of the nominal response model item parameters. *Applied Psychological Measurement*, *27*, 275-288.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.
- Hanger, H. C., Fogarty, B., Wilkinson, T. J., & Sainsbury, R. (2000). Stroke patients' views on stroke outcomes: death versus disability. *Clinical Rehabilitation*, *14*, 417-424.
- Kim, S., Kim, S., & Kamphaus, R.W. (2010). Is aggression the same for boys and girls? Assessing measurement invariance with confirmatory factor analysis and item response theory. *School Psychology Quarterly*, *25*, 45-61.

- Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*, 146-162.
- Meade, A. M. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology, 95*, 728-743.
- Multivariate Software. (2010). *EQSIRT: Item response theory software*. Encino, CA: Author.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.
- Preston, K. S. J., & Reise, S. P. (2013). Diagnosing problematic items with the NRM. In S. P. Reise & D. Revicki (Eds.), *Handbook of item response theory models: With an emphasis on patient reported outcomes*. Manuscript under preparation.
- Purpura, D. J., Wilson, S. B., & Lonigan, C. J. (2010). Attention-deficit/hyperactivity disorder symptoms in preschool children: Examining psychometric properties using item response theory. *Psychological Assessment, 22*, 546-558.
- Ramsay, J. O. (2000). Differential equation models for statistical functions. *Canadian Journal of Statistics, 28*, 225-240.
- Samuel, D. B., Simms, L. J., Clark, L. A., Livesley, W. J., & Wigider, T. A. (2010). An item response theory integration of normal and abnormal personality scales. *Personality Disorders: Theory, Research, and Treatment, 1*, 5-21.
- Seong, T. J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement, 14*, 299-311.
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement, 16*, 1-16.
- Thomas, M. L., & Locke, D. E. C. (2010). Psychometric properties of the MMPI-2-RF somatic complaints (RC1) scale. *Psychological Assessment, 22*, 492-503.
- van den Oord, E. J. C. G. (2005). Estimating Johnson curve population distributions in MULTILOG. *Applied Psychological Measurement, 29*, 23-30.
- Woods, C. M. (2006). Ramsay-curve item response theory to detect and correct for non-normal latent variables. *Psychological Methods, 11*, 253-270.
- Woods, C. M. (2007a). Ramsay curve IRT for Likert-type data. *Applied Psychological Measurement, 31*, 195-212.
- Woods, C. M. (2007b). Empirical histograms in IRT with ordinal data. *Educational and Psychological Measurement, 67*, 73-87.
- Woods, C. M. (2008). Consequences of ignoring guessing when estimating the latent density in item response theory. *Applied Psychological Measurement, 32*, 371-384.
- Woods, C. M., & Lin, N. (2009). Item response theory with estimation of the latent density using Davidian curves. *Applied Psychological Measurement, 33*, 102-117.
- Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika, 71*, 1-22.
- Zwinderman, A. H., & van den Wollenberg, A. L. (1990). Robustness of marginal maximum likelihood estimation in the Rasch model. *Applied Psychological Measurement, 14*, 73-81.