

# Using the Asymmetry of Item Characteristic Curves (ICCs) to Learn About Underlying Item Response Processes

Sora Lee and Daniel M. Bolt

**Abstract** In this chapter, we examine how the nature and number of underlying response subprocesses for a dichotomously scored item may manifest in the form of asymmetric item characteristic curves. In a simulation study, binary item response datasets based on four different item types were generated. The item types vary according to the nature (conjunctively versus disjunctively interacting) and number (1–5) of subprocesses. Molenaar’s (2014) heteroscedastic latent trait model for dichotomously scored items was fit to the data. A separate set of simulation analyses considers also items generated with non-zero lower asymptotes. The simulation results illustrate that form of asymmetry has a meaningful relationship with the item response subprocesses. The relationship demonstrates how asymmetric models may provide a tool for learning more about the underlying response processes of test items. *online* at [www.SpringerLink.com](http://www.SpringerLink.com)

**Keywords** Item response theory • Asymmetric ICCs • Item complexity • Item validity

## 1 Introduction

The item characteristic curves (ICCs) of most traditional item response theory (IRT) models are symmetric. Specifically, the change in probability observed above the inflection point in the ICC is a mirror image of the change that occurs below the inflection point. IRT models such as the Rasch model, the two and three-parameter logistic and normal ogive models are well-known examples.

Recently, there has been a growing psychometric literature related to asymmetric ICCs, and models that can be used to represent and explain such asymmetry. There are good reasons to believe that the nature of the psychological response process underlying many educational test items will be better reflected by asymmetric models. As considered by Samejima (2000), items scored as binary can often

---

S. Lee (✉) • D.M. Bolt

Department of Educational Psychology, University of Wisconsin, Madison, WI, USA  
e-mail: [slee486@wisc.edu](mailto:slee486@wisc.edu); [dmbolt@wisc.edu](mailto:dmbolt@wisc.edu)

be viewed as representing outcomes of multiple conjunctively or disjunctively interacting subprocesses. An example is a complex math word problem, in which the final answer may be arrived at only following the correct execution of a series of steps (e.g., converting the stated problem into an algebraic equation, solving the algebraic equation, etc.), where failure at any one step would lead to an overall incorrect response on the item. Assuming the individual steps (i.e., subprocesses) each conform to a logistic model, the overall item score should yield an asymmetric curve. In the case of conjunctively interacting subprocesses, the result should be an asymmetric ICC that accelerates at a slower rate to the right of the inflection point than it accelerates to the left of the inflection point (Samejima 2000). The extent of the asymmetry will be affected by the number of conjunctively interacting subprocesses.

Alternatively, for many items, the item score might be the outcome of disjunctively interacting subprocesses. An example is ability-based guessing model of San Martín, Del Pino, and De Boeck (2006), a model designed for multiple-choice items. Under the ability-based guessing model, a separate problem-solving process and guessing process are applied in sequential fashion such that an incorrect outcome from the problem solving process (e.g., the answer arrived at is not among the available response options), can be overcome by the guessing process. The nature of the asymmetry created by these two disjunctive subprocesses at the item score level (assuming again that each subprocess follows a logistic/normal ogive form) is the opposite to that described for the complex math word problem example. Specifically, the ICC will accelerate at a faster rate to the right of the inflection point than it accelerates to the left of the inflection point (Samejima 2000).

Model-based approaches to representing asymmetric ICCs of these kinds can take different forms. Samejima (2000) presents a logistic positive exponent (LPE) model in which an exponent parameter (or “acceleration” parameter) is introduced to a standard logistic model. While estimation algorithms have been proposed for this model (e.g., Samejima 2000; Bolfarine & Bazan, 2010), a challenge is the confound between the exponent parameter and the difficulty parameter (Lee 2015; Bolt, Deng, & Lee, 2014).

An alternative approach is Molenaar’s (2014) normal ogive residual heteroscedasticity (RH) model. Molenaar (2014) illustrated how violation of the residual homoscedasticity assumption that underlies normal ogive models yields asymmetric ICCs for binary items. Such heteroscedasticity can be taken to reflect a greater variability in anticipated performances on an item conditional upon ability, and could conceivably reflect different underlying causes. In this chapter we consider the possibility that the heteroscedasticity reflects the nature and number of conjunctively/disjunctively interacting subprocesses described above, a feature that might often intuitively be expected to vary across items within a test. One of the advantages of the RH model is that the parameter associated with asymmetry is not confounded with difficulty, as in the LPE.

The purpose of this study is to examine whether the RH model can be used to inform about the underlying response processes associated with test items. Specifically, we examine how manipulation of both the nature and number of interacting subprocesses may be related to detectable asymmetries in the ICCs of

test items. Such an application, if successful, would support the RH model as item-level validation tool. From another perspective, it would suggest that the RH model may help in learning more about the underlying response process of a test item.

### ***1.1 Other Implications of Ignoring Asymmetry in ICCs***

The possibility that asymmetric ICCs can be used for item validation purposes represents just one additional reason for considering models such as the RH model.

The potential value of attending to asymmetry has already been considered from several different perspectives, suggesting that the implications of ignoring asymmetric ICCs, where they are present, can be significant. Woods and Harpole (2015), for example, have demonstrated the potential for inflated Type I error in DIF analyses when residual heteroscedasticity is present but ignored by the model testing for DIF. Molenaar (2014) illustrates how the estimated item information functions can be highly inaccurate when asymmetries are ignored. Such inaccuracies can not only influence how items are adaptively selected, but also the resulting estimated standard errors of ability estimates. With respect to person scoring, Samejima (2000) also notes an inconsistency in item weighting that emerges when using symmetric models, a problem that can be resolved using asymmetric models. Finally, ignoring asymmetry can also create problems related to the IRT metric. For example, Bolt et al. (2014) demonstrate how the presence of asymmetric ICCs may ultimately be responsible for the score deceleration problem seen when standardized tests are used to measure growth across grade levels.

### ***1.2 Item Response Processes and Asymmetric ICCs***

As indicated above, the purpose of this preliminary study was to examine whether the asymmetry of ICCs may also provide a way of learning about the nature and number of underlying item subprocesses, and whether the relationship is strong enough to allow asymmetric items to provide insight into the items. With multi-dimensional item response models, it has been common to attend to conjunctive or disjunctive response processes by considering different ways in which the latent traits, or more specifically, the processes associated with different latent traits, may interact. For example, cognitively diagnostic models emphasize skill attribute interactions as conjunctive versus disjunctive (e.g., Junker & Sijtsma, 2001; Maris 1995). Similarly, a distinction is often made between MIRT models that are compensatory versus noncompensatory (see e.g., Bolt & Lall, 2003). However, as emphasized in this paper, it can be useful to consider different forms of subprocess interaction in relation to collections of items that are statistically unidimensional. In Samejima's (2000) presentation of the LPE model, the number and nature of interacting subprocesses define the *complexity* of the item. We adopt the same terminology in this chapter, but use residual heteroscedasticity as a means of capturing such complexity as opposed to the exponent parameter used in the LPE.

## 2 Molenaar's Normal Ogive RH Model

The use of a normal ogive to represent an item response function for a binary item score follows from a model that assumes an underlying continuous latent response propensity that, conditional upon latent ability  $\theta$ , is normally distributed. The mean of the conditional distribution is assumed to be a linear function of  $\theta$ . The remaining variability in the response propensity conditional upon  $\theta$ , denoted  $\varepsilon_i|\theta$ , represents sources of random noise, and is assumed to have a constant variance across  $\theta$ , denoted  $\sigma_{\varepsilon_i|\theta}^2$ , referred to as the residual variance. In effect, scoring the item as binary can be viewed as defining a threshold with respect to  $\varepsilon_i$  that translates the continuous response propensity into a binary score. A normal ogive curve for the probability of correct response follows from the integration under the conditional normal distribution of the area above the threshold. Generalizations of this model to polytomous scores are straightforward, and simply require the consideration of multiple thresholds in relation to  $\varepsilon_i$  as opposed to just one (see e.g., Lord & Novick, 1968, pp. 370–371 for details).

The assumption of homoscedasticity of the response propensity variance across ability levels naturally plays an important role in how the probability of a correct response is defined. If heteroscedasticity of variance is present, it will alter the form of the probability curve assuming other features of the model are held constant. Generalization of the normal ogive model to accommodate heteroscedasticity of variance naturally requires specification of a suitable function for  $\sigma_{\varepsilon_i|\theta}^2$ . Molenaar proposed the following form of heteroscedasticity in the context of polytomously scored items (Molenaar, Dolan, & De Boeck, 2012):

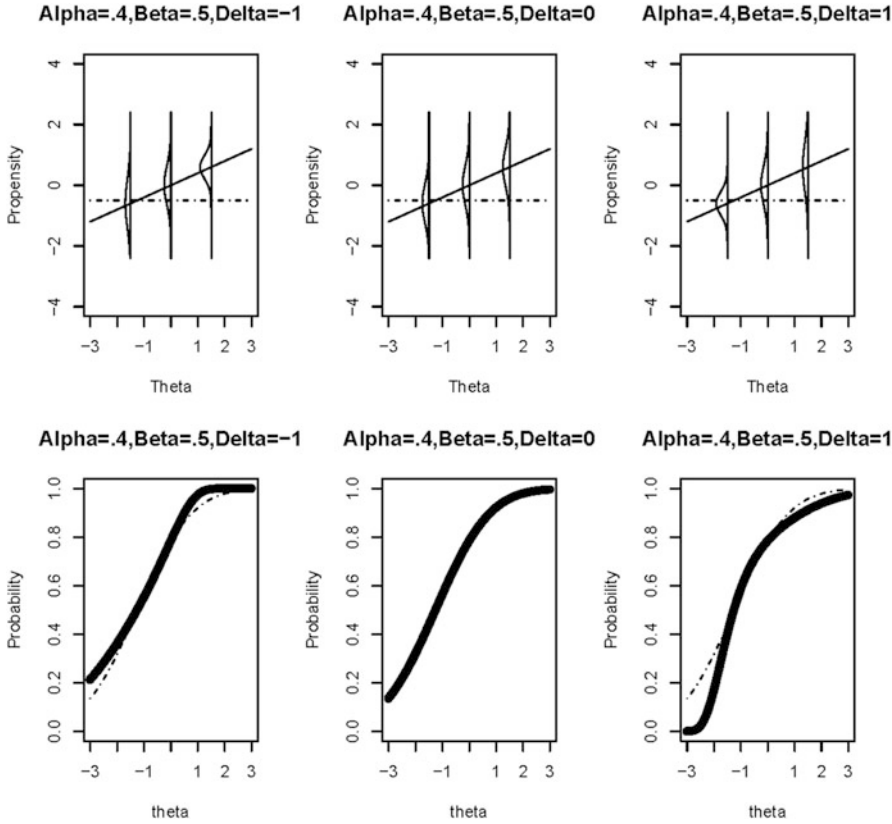
$$\sigma_{\varepsilon_i|\theta}^2 = 2\delta_0[1 + \exp(-\delta_1\theta)]^{-1} \quad (1)$$

where  $\delta_0$  is a baseline parameter, and  $\delta_1$  is heteroscedasticity parameter,  $\delta_0 \in (0, \infty)$  and  $\delta_1 \in (-\infty, \infty)$ . Note that if  $\delta_1 = 0$ , then the residual variances are homoscedastic with  $\sigma_{\varepsilon_i|\theta}^2 = \delta_0$ ; if  $\delta_1 > 0$ , then the residual variance is increasing with  $\theta$ ; if  $\delta_1 < 0$ , residual variances are decreasing with  $\theta$ .

Molenaar (2014) derived a corresponding model for dichotomously scored items based on the same model for heteroscedasticity. The resulting item response function is:

$$P(y_i = 1|\theta) = \Phi\left(\frac{\alpha_i\theta + \beta_i}{\sqrt{2}[1 + \exp(-\delta_{1i}\theta)]^{-1/2}}\right) \quad (2)$$

where  $\delta_{1i}$  is the item heteroscedastic parameter, and  $\alpha_i$  and  $\beta_i$  denote the slope (discrimination) and intercept (difficulty) parameters associated with the normal ogive model. As for the polytomous model, the model in (2) reduces to the standard normal ogive model in the case where  $\delta_{1i} = 0$ . Further details on this model are provided by Molenaar (2014).



**Fig. 1** Residual heteroscedasticity and item characteristic curves

Figure 1 provides an illustration of how manipulation of the  $\delta_{1i}$  parameter introduces ICC asymmetry. The plots at the top of the figure illustrate the heteroscedasticity associated with the RH model for three different hypothetical items that vary only with respect to  $\delta_{1i}$ . The middle figure corresponds to the condition of homoscedasticity, while the figures on the left and right correspond to examples where the residual variance decreases and increases, respectively, in relation to  $\theta$ . When translated into probability curves, the items yield different ICCs. In particular, a negative  $\delta_{1i}$  results in an ICC with a steeper slope to the right of the inflection point than the corresponding symmetric ICC, and a flatter slope to the left of the inflection point. Just the opposite is observed for the item with a positive delta value.

A primary goal of the current paper is to illustrate how Molenaar’s RH model can be used to capture differences in the underlying response processes associated with different items. To this end, we also attempt to illustrate how asymmetric ICCs can be a naturally expected outcome for educational test data. We also seek to clarify the potential of the RH model in recovering the nature of the asymmetry associated with these different response processes.

It is worth noting that estimation procedures also exist for other models that can flexibly account for asymmetric ICCs. For example, Bolfarine and Bazan (2010) considered the use of Bayesian estimation techniques with Samejima's LPE model. Preliminary work (Lee 2015), however suggests that the RH model of Molenaar may be slightly better in terms of recovery, perhaps in large part due to the greater separation of parameters associated with the asymmetry and item difficulty. We therefore focus on Molenaar's RH model in the current paper.

## 2.1 *Bayesian Estimation of Heteroscedastic Two-Parameter and Three-Parameter Normal Ogive Models*

Molenaar (2014) presents a marginal maximum likelihood algorithm for the RH model. In this paper we consider the model in a Bayesian estimation framework, as well as a three-parameter version that introduces a lower asymptote parameter.

Under the two-parameter Residual Heteroscedasticity (2P-RH) model, we assume the following priors for the item parameters:

$$\begin{aligned}\beta_i &\sim \text{Normal}(0,1) \\ \alpha_i &\sim \text{Lognormal}(0,2) \\ \delta_{1i} &\sim \text{Normal}(0,1)\end{aligned}$$

and for the person parameter:

$$\theta \sim \text{Normal}(0,1)$$

For the three-parameter Residual Heteroscedasticity (3P-RH) model, we consider use of the same parameters, but add a fixed lower asymptote parameter,  $\gamma$ :

$$P(y_i = 1|\theta) = \gamma + (1 - \gamma)\Phi\left(\frac{\alpha_i\theta + \beta_i}{\sqrt{2}[1 + \exp(-\delta_{1i}\theta)]^{-1/2}}\right) \quad (3)$$

In the current study,  $\gamma = 0.2$  when generating the data, and we also fix  $\gamma = 0.2$  when estimating the model, as might reflect a multiple-choice test with five options per item. Thus the three-parameter simulation evaluates how well the model functions in the presence of known guessing effects. Our preliminary analyses did consider a 3P-RH model with an estimated lower asymptote, although the model resulted in simulated chains with poor convergence.

### 3 Simulation Study

To evaluate the effectiveness of the RH model in informing about underlying response process, we simulated item response data to conform to different types of response processes. In effect, we assumed each binary item was the outcome of one of four possible types, ordered from the least to most complex: (1) a disjunctive two-subprocess item; (2) a single subprocess item; (3) a conjunctive two-subprocess item; and (4) a conjunctive five-subprocess item. In all cases, data were simulated as unidimensional. It is worth noting that unlike models such as in Whitely (1980), the presence of distinct subprocesses is not associated with multidimensionality, reflecting the fact that as a statistical dimension, a single underlying latent trait can often reflect what is in reality a complex constellation of skills. Regardless of the item type, each subprocess was simulated from a normal ogive model, i.e.,

$$P_{ik}(\theta) = P(u_{ik} = 1|\theta) = \Phi(\alpha_{ik}\theta + \beta_{ik}), \quad (4)$$

where  $P_{ik}(\theta)$  denotes the probability of successfully executing subprocess  $k$  on item  $i$  (i.e.,  $u_{ik} = 1$ ), and  $\alpha_{ik}$ ,  $\beta_{ik}$  denote item subprocess discrimination and difficulty (threshold) parameters, respectively. The distinguishing characteristics of the items relate to the number of subprocesses as well as the nature of their interaction.

#### 3.1 Low Complexity Disjunctive Items: A Two Subprocess Model

The first item type simulated assumes two subprocesses with a disjunctive interaction:

$$P(y_i = 1|\theta) = P_{i1}(\theta) + (1 - P_{i1}(\theta))P_{i2}(\theta) \quad (5)$$

As noted earlier, such a model could reflect an ability-based guessing context (San Martín et al. 2006), whereby a student can solve the item in one of two ways: (1) ordinary problem solving behavior, where the solution may be arrived at using the intended approach, while if not attained is followed by (2) guessing behavior, where the various response options are evaluated apart from the intended problem-solving process, and the most sensible option is chosen.

#### 3.2 Moderate Complexity Items: One Subprocess Model

For comparison purposes, we consider also a one subprocess item:

$$P(y_i = 1|\theta) = P_{i1}(\theta) \quad (6)$$

Such items reflect a condition of ICC symmetry, and might correspond to items that reflect direct testing of particular components of knowledge, such as the definition of a concept, for example. With respect to the RH model, they should yield items for which the estimated  $\delta_{1i}$  is near 0.

### 3.3 *Moderately High Complexity Conjunctive Items: A Two Subprocess Model*

A third item type assumes two subprocesses, but with a conjunctive interaction:

$$P(y_i = 1|\theta) = P_{i1}(\theta)P_{i2}(\theta) \quad (7)$$

From the section above, it is anticipated that conjunctive items will yield positive delta estimates. Such items would represent an item that involves two steps, where a correct answer is only achieved when both steps are successfully executed.

### 3.4 *High Complexity Conjunctive Items: A Five Subprocess Model*

The fourth item type is similar to the third, but involves five, as opposed to two, subprocesses:

$$P(y_i = 1|\theta) = P_{i1}(\theta)P_{i2}(\theta)P_{i3}(\theta)P_{i4}(\theta)P_{i5}(\theta) \quad (8)$$

Such items could be viewed as items involving five steps, where a correct answer is only attained when all five steps are executed correctly. Relative to the previous category, these items should return the most positive estimates of  $\delta_{1i}$ .

In order to simulate items that varied primarily in the number and nature of interacting subprocesses, we simulated subprocess parameters using distributions within each item type that would render items that were comparable in terms of overall item discrimination and difficulty. It was our intent that the primary psychometric feature distinguishing these four categories of item types from each other would be the asymmetry of their ICCs, not characteristics such as difficulty or discrimination. For the five subprocess conjunctive items, we generated  $\alpha_{ik} \sim \text{lognorm}(-0.3, 0.4)$  and  $\beta_{ik} \sim \text{unif}(1, 2.5)$ ; for the two subprocess conjunctive items,  $\alpha_{ik} \sim \text{lognorm}(-0.1, 0.4)$  and  $\beta_{ik} \sim \text{unif}(0, 1.5)$ ; for the two subprocess disjunctive model,  $\alpha_{ik} \sim \text{lognorm}(0, 0.4)$  and  $\beta_{ik} \sim \text{unif}(-1.5, 1)$ ; for the one subprocess model,  $\alpha_{i1} \sim \text{lognorm}(0, 0.4)$  and  $\beta_{i1} \sim \text{unif}(-2, 2)$ .

In all cases, we also simulated examinee proficiency  $\theta$  as normal, with a mean of 0 and variance of 1.



Finally, in order to consider a situation in which a non-zero lower asymptote was also present, in a separate set of simulation analyses, we generated items from the same four item type categories but now using a simulation model that introduced a nonzero lower asymptote. Specifically, for the low complexity disjunctive items we simulate:

$$P(y_i = 1|\theta) = \gamma + (1 - \gamma)[P_{i1}(\theta) + (1 - P_{i1}(\theta))P_{i2}(\theta)],$$

while for the moderate complexity items:

$$P(y_i = 1|\theta) = \gamma + (1 - \gamma)P_{i1}(\theta),$$

for the moderately high complexity conjunctive items:

$$P(y_i = 1|\theta) = \gamma + (1 - \gamma)P_{i1}(\theta)P_{i2}(\theta),$$

and for the high complexity conjunctive items:

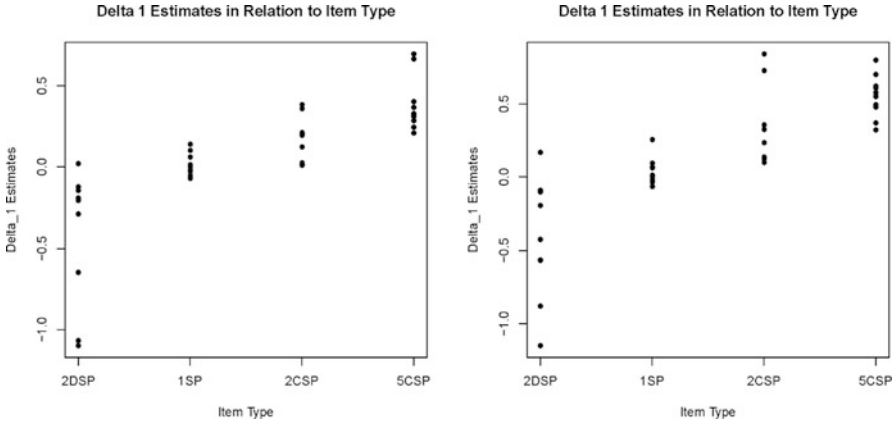
$$P(y_i = 1|\theta) = \gamma + (1 - \gamma)P_{i1}(\theta)P_{i2}(\theta)P_{i3}(\theta)P_{i4}(\theta)P_{i5}(\theta),$$

where in all cases,  $\gamma = .2$ . As described above, we also fixed the  $\gamma$  at .2 when estimating the model.

Each simulated dataset included ten items from each category, so 40 items total per simulated dataset, and simulated responses for 25,000 examinees. All MCMC runs were run out to 10,000 iterations, and  $\delta_{li}$  estimates were obtained for each item. We carried out 20 replications for each of the two-parameter and three-parameter simulation models. In each case the appropriate model (two-parameter or three-parameter RH model) was used as corresponded to the simulation condition.

## 4 Simulation Results

Figure 2 provides a graphical illustration of the  $\delta_{li}$  estimates for a single simulation run in each of the two-parameter and three-parameter conditions against the item type category. The item type categories are ordered from least to most complex, such that the increase in  $\delta_{li}$  estimates across categories is as expected. Tables 1 and 2 provide a tabulation of the results across 20 replications in each condition. Also apparent from the table is the tendency for the  $\delta_{li}$  estimates to increase as item complexity increases. Nevertheless, there remains a fair amount of variability within each category, variability that can be attributed to the imprecision in estimating  $\delta_{li}$  as well as the potential sensitivity of the  $\delta_{li}$  estimates to other characteristics of items (e.g., the difficulty and discrimination of the individual subprocesses within item) that varied within the simulation and may have an effect on these estimates. It is, however, noteworthy that the vast majority of items in the low complexity



**Fig. 2**  $\delta_{1i}$  estimates against the item type category in 2P (*left*) and 3P (*right*) condition, respectively

**Table 1**  $\delta_{1i}$  estimates against the item type in 2P condition (ICC = 0.65)

Item type	$\hat{\delta}_1$ Mean	$\hat{\delta}_1$ Std dev
2DSP	−0.39	0.41
1SP	0.01	0.07
2CSP	0.17	0.13
5CSP	0.38	0.17

**Table 2**  $\delta_{1i}$  estimates against the item type in 3P condition (ICC = 0.64)

Item type	$\hat{\delta}_1$ Mean	$\hat{\delta}_1$ Std dev
2DSP	−0.39	0.41
1SP	0.04	0.09
2CSP	0.31	0.27
5CSP	0.55	0.14

category return  $\delta_{1i}$  estimates less than 0, while those in the moderate complexity category are centered right around 0, and the vast majority of those in the moderate or high complexity category return  $\delta_{1i}$  estimates greater than 0. Intraclass correlation estimates, which are from variance component estimation using the ANOVA method to determine within and between item type variance, were 0.65 and 0.64 for the two-parameter and three-parameter analyses, respectively, suggesting that the presence of a nonzero lower asymptote (corresponding to the effects of random guessing) does not have a deleterious effect on the  $\delta_{1i}$  estimates. It is also worth noting, however, that the category of low item complexity seemed to yield the highest variability in  $\delta_{1i}$  estimates. Such a result may reflect the metric of the  $\delta_{1i}$  parameter.

## 5 Discussion

There are several limitations to our study. First, it is only a simulation, and should be replicated with real data. Identifying example items where the underlying response process is known or highly suspected, and seeing  $\delta_{li}$  estimates from real data analyses that are consistent with such knowledge, would provide strong evidence in support of the approach. Second, our simulation used a proficiency distribution that matched that assumed by the estimation algorithm (in both cases normal). The possibility of non-normal trait distributions, and the implications this has for representing asymmetries and how they vary across items, should be further examined. The shape of any ICC is to a large extent arbitrary when considering arbitrary nonlinear alterations of the proficiency metric. Alternative approaches have considered retaining the symmetric model, but allowing for nonnormal trait distributions (see e.g., Woods & Thissen, 2006). The possibility of altering the ICC shape versus altering the proficiency metric is often unclear when analyzing real data (Molenaar 2014). The presence of items that vary in the number and nature of subprocesses is important in generating meaningful variability in delta. Third, the nature of the response processes for the different item type categories are simplistic. It is of course conceivable that an item may contain a mix of conjunctively and disjunctively interacting subprocesses, and that many items may also be solved using multiple different strategies. Fourth, our simulation study used large samples, as may often be available for large-scale assessments. It remains to be seen how well the model performs with smaller samples.

There are also additional extensions to the method and its application that could be considered. As noted earlier, the possibility of estimating a lower asymptote parameter for the RH model could be considered. In addition, other forms of heteroscedasticity in relation to the proficiency could be developed, some of which may be more appropriate than the current approach for the types of items being simulated. In general, beyond seeing relationships between the  $\delta_{li}$  parameter and item type category, more work is needed in evaluating how well the RH model actually fits items of the type simulated in this chapter. Finally, the possibility of using the RH model as a basis for IRT applications, such as CAT or vertical scaling, and comparisons against traditional approaches using symmetric models, would be useful.

## References

- Bolfarine, H., & Bazan, J. L. (2010). Bayesian estimation of the logistic positive exponent IRT model. *Journal of Educational and Behavioral Statistics*, 35, 693–713.
- Bolt, D. M., Deng, S., & Lee, S. (2014). IRT model misspecification and measurement of growth in vertical scaling. *Journal of Educational Measurement*, 51(2), 141–162.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional IRT models using Markov chain Monte Carlo. *Applied Psychological Measurement*, 27, 395–414.

- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272.
- Lee, S. (2015). *A comparison of methods for recovery of asymmetric item characteristic curves in item response theory* (Unpublished master's thesis). Madison: University of Wisconsin.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60, 523–547.
- Molenaar, D. (2014). Heteroscedastic latent trait models for dichotomous data. *Psychometrika*, 80(3), 625–644.
- Molenaar, D., Dolan, C. V., & De Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika*, 77, 455–478.
- Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60(4), 549–572.
- Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, 65, 319–335.
- San Martín, E., Del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, 30(3), 183–203.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4), 479–494.
- Woods, C. M., & Harpole, J. K. (2015). How item residual heterogeneity affects tests for differential item functioning. *Applied Psychological Measurement*, 39, 251–263.
- Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, 71, 281–301.

Quantitative Psychology Research

The 80th Annual Meeting of the Psychometric Society,  
Beijing, 2015

van der Ark, L.A.; Bolt, D.M.; Wang, W.C.; Douglas, J.A.;  
Wiberg, M. (Eds.)

2016, IX, 398 p. 50 illus., 29 illus. in color., Hardcover  
ISBN: 978-3-319-38757-4