Springer Proceedings in Mathematics & Statistics

L. Andries van der Ark
Daniel M. Bolt
Wen-Chung Wang
Jeffrey A. Douglas
Marie Wiberg *Editors*

# Quantitative Psychology Research

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

Springer

**Springer Proceedings in Mathematics & Statistics**

Volume 167

More information about this series at http://www.springer.com/series/10533

# Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of select contributions from workshops and conferences in all areas of current research in mathematics and statistics, including OR and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

L. Andries van der Ark • Daniel M. Bolt
Wen-Chung Wang • Jeffrey A. Douglas
Marie Wiberg
Editors

# Quantitative Psychology Research

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

Springer

*Editors*

L. Andries van der Ark
University of Amsterdam
Amsterdam, The Netherlands

Daniel M. Bolt
University of Wisconsin
Madison, Wisconsin, USA

Wen-Chung Wang
Education University of Hong Kong
Hong Kong, China

Jeffrey A. Douglas
University of Illinois
Champaign, Illinois, USA

Marie Wiberg
Umeå University
Umeå, Sweden

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-38757-4 ISBN 978-3-319-38759-8 (eBook)
DOI 10.1007/978-3-319-38759-8

Library of Congress Control Number: 2016944495

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

# Preface

This volume represents presentations given at the 80th annual meeting of the Psychometric Society, organized by the Beijing Normal University, during July 12–16, 2015. The meeting attracted 511 participants from 21 countries, with 254 papers being presented, along with 119 poster presentations, three pre-conference workshops, four keynote presentations, eight invited presentations, and six invited and five contributed symposia. This meeting was the first ever held in China, the birthplace of standardized testing, as was highlighted in the keynote address “the history in standardized testing” by Dr. Houcan Zhang. We thank the local organizers Tao Xin and Hongyun Liu and their staff and students for hosting this very successful conference.

Since the 77th meeting in Lincoln, Nebraska, Springer publishes the proceedings volume from the annual meeting of the Psychometric Society so as to allow presenters to quickly make their ideas available to the wider research community, while still undergoing a thorough review process. The first three volumes of the meetings in Lincoln, Arnhem, and Madison were received successfully, and we expect a successful reception of these proceedings too.

We asked authors to use their presentation at the meeting as the basis of their chapters, possibly extended with new ideas or additional information. The result is a selection of 29 state-of-the-art chapters addressing a diverse set of topics, including item response theory, factor analysis, structural equation modelling, time series analysis, mediation analysis, cognitive diagnostic models, and multi-level models.

| | |
|---|---|
| Amsterdam, The Netherlands | L. Andries van der Ark |
| Madison, WI | Daniel M. Bolt |
| Hong Kong, China | Wen-ChungWang |
| Urbana-Champaign, IL | Jeffrey A. Douglas |
| Umeå, Sweden | Marie Wiberg |

# Contents

# Continuation Ratio Model in Item Response Theory and Selection of Models for Polytomous Items

**Seock-Ho Kim**

**Abstract** In the continuation ratio model continuation ratio logits are used to model the probabilities of obtaining ordered categories in polytomously scored items. The continuation ratio model is an alternative to other models for ordered category items such as the graded response model, the generalized partial credit model, and the partial credit model. The theoretical development of the model, descriptions of special cases, maximum likelihood estimation of the item and ability parameters are presented. An illustration and comparisons of the models for ordered category items are presented using empirical data.

**Keywords** Bayesian estimation • Continuation ratio model • Item response theory • Maximum likelihood estimation • Multicategory logit model • Polytomous model

## 1 Introduction

When a free response item is scored in a dichotomous fashion, a single decision is performed in a sense that no further decisions will be made beyond the current decision to be taken. When a free response item is rated in a polytomous fashion, either a single decision is performed or multiple decisions in which dependent decisions are made in tandem are required.

Borrowing terms from the game theory (Luce & Raiffa, 1957), a particular alternative chosen by a rater at a given decision point is called a "choice," and the totality of choices available to a rater at the decision point constitutes a "move." A sequence of choices, one following another until the rating or scoring of an item is complete, can be called a "play." The play or the rating process for a given item can be depicted with a connected graph, called a decision tree, consists of a collection of nodes and branches between pairs of nodes. A decision tree with three decision points and four choices is presented in Fig. 1. The decision tree reflects

S.-H. Kim (✉)
Department of Educational Psychology, The University of Georgia, 325 Aderhold Hall, Athens, GA 30602-7143, USA
e-mail: shkim@uga.edu

L.A. van der Ark et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics 167, DOI 10.1007/978-3-319-38759-8_1

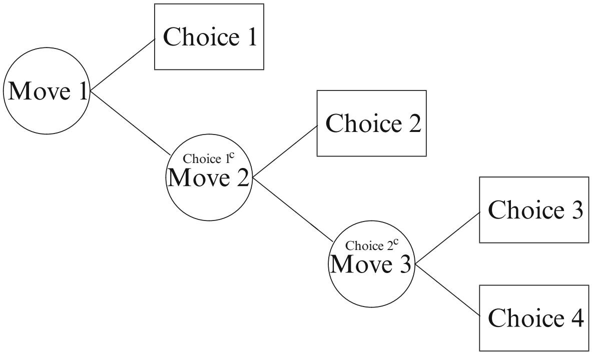

**Fig. 1** A decision tree with three decision points and four choices

the sequential nature of scoring. Each decision point is denoted as a circle and the chance events with respective but dependent probabilities are denoted as squares in Fig. 1. The superscript c of the choice number indicates the complement of the event.

The decision tree in Fig. 1 involves in a set of dependent events. The model for the ordered choices ought to reflect the joint probabilities and must take into account the conditional probabilities that characterize the dependence. The model for ordered category items to be described is called a continuation ratio model. Such a model that employs continuation-ratio logits with a manifested or directly-observed explanatory variable was originally developed to handle a multicategory response variable in logit models (Cox 1972). In the item response theory field, Mellenbergh (1995) presented conceptual notes on models for discrete polytomous item responses and indicated that the continuation ratio model could be considered as a special case of the Bock's (1972) model (cf. Tutz 1990; Hemker, van der Ark, & Sijtsma, 2001). The general discussion of the various item response theory models for polytomously scored items can be found in Hambleton, van der Linden, and Wells (2010).

## 2 The Continuation Ratio Model and Parameter Estimation

Let $Y_{ij}$ be a random variable that designates the rating or scored item response of individual $i$ to item $j$. The continuation ratio model assumes that the manifestation of $Y_{ij}$ or the probability of $Y_{ij}$ to be a specific value depends on a person's latent ability $\theta_i$ and a vector-valued item characteristic $\xi_j$ [i.e., $a_{jk}$'s and $b_{jk}$'s; see the definitions following Eq. (1)]. The probability that $y_{ij} = k$ given ability $\theta_i$ and item parameter $\xi_j$, Prob $\left(y_{ij} = k|\theta_i, \xi_j\right)$, is

$$P_{jk}(\theta_i) = \begin{cases} \dfrac{\exp\left[-a_{jk}(\theta_i - b_{jk})\right]}{\displaystyle\prod_{h=1}^{k}\left\{1 + \exp\left[-a_{jh}(\theta_i - b_{jh})\right]\right\}} & \text{for } k = 1, \ldots, K_j - 1 \\ \dfrac{1}{\displaystyle\prod_{h=1}^{K_j-1}\left\{1 + \exp\left[-a_{jh}(\theta_i - b_{jh})\right]\right\}} & \text{for } k = K_j, \end{cases} \tag{1}$$

where $a_{jk}$ is the slope parameter and $b_{jk}$ is the threshold parameter. The number of item parameters for item $j$ is $2(K_j - 1)$. When an item has two rating categories, that is, $K_j = 2$, the continuation ratio model becomes the two-parameter logistic model.

Under the assumption of conditional independence, the probability of a response vector $y_i = (y_{i1}, \ldots, y_{iJ})$, is given as $\text{Prob}\,(y_i|\theta_i, \xi) = p(y_i|\theta_i, \xi_1, \ldots, \xi_J) = \prod_{j=1}^{J} P_{jk}(\theta_i)$ and the joint probability of the response vectors of a sample of $I$ subjects is given as $\text{Prob}\,(y|\theta, \xi) = p(y_1, \ldots, y_I|\theta_1, \ldots, \theta_I, \xi) = \prod_{i=1}^{I}\prod_{j=1}^{J} P_{jk}(\theta_i)$. When the joint probability is considered as a function of unknown parameters $\xi$ and $\theta$, we call it the likelihood $L$. Inference of the values of unknown parameters from observed data can be accomplished by maximizing the likelihood or its modifications with respect to the unknown parameters.

Several estimation procedures are available to obtain parameter estimates in the continuation ratio model. Kim (2002) presented detailed estimation procedures including the marginal estimation of item parameters (Bock & Aitkin, 1981). Kim (2002) also presented model fit statistics, estimation of the latent criterion variable $\theta_i$ (i.e., methods of maximum likelihood, maximum a posteriori, and expected a posteriori), and information functions for the continuation ratio model.

It can be noted that the continuation ratio model treats a polytomously scored item as a set of dichotomously scored items (Kim 2013). For example, an item with four categories or choices can be converted into three dichotomously scored items with some dependency among the converted dichotomous items. It is possible, consequently, to obtain the parameter estimates under the continuation ratio model using computer programs that implemented the marginal maximum likelihood estimation of item parameters under the usual two-parameter logistic model and an ability estimation method. Kim (2013) presented means to obtain parameter estimates using several popular item response theory computer programs utilizing missing or not-presented options.

Note that other parameter estimation methods (e.g., Bayesian estimation, Markov chain Monte Carlo, Gibbs sampling; see Baker & Kim, 2004) implemented in item response theory computer programs can also be applied to obtain both item and ability parameter estimates under the continuation ratio model. Because of the relationship between the two-parameter logistic model and the continuation ratio model, priors of item parameters used in Bayesian estimation can be employed with minor changes (e.g., Swaminathan & Gifford 1985).

Although the continuation ratio model for the polytomous items with ordered categories has been available for some time, applications of the model to analyze polytomous data are not widely available. An illustration is presented next using empirical data with the Fortran implementation of the continuation ratio model and the computer program MULTILOG (Thjssen, Chen, & Bock, 2002). Subsequently, comparisons of the estimation results from several models for ordered category items are presented using MULTILOG.

## 3 An Illustration

The data from an experimental form of a French writing assessment were analyzed. The experimental form was a performance assessment rating instrument that consists of three polytomously scored items with four ordered rating categories. The participants were 120 college students who had complete data for the three item responses. Although there might be 64 different response patterns, 31 distinctive patterns were actually observed (see Table 2 for the response patterns and the number of examinees in each pattern).

The marginal maximum likelihood estimation of item parameters was carried out on the three French items from the experimental form using the Fortran computer program modified from the code written for Kim (2002). Ten quadrature fractile points were used for ability integration during calculations. After several cycles of the expected and maximization iterations, the item parameter estimates were stable to four significant figures. Goodness of fit for the model was assessed, and the resulting chi-square value of the $-2$ log likelihood was 38.81 with the degrees of freedom of 12 (i.e., the number of response patterns minus the number of parameters estimated minus one; see Bock & Aitkin, 1981). Although the solution showed reasonably good fit, the chi-square was relatively large (i.e., $p < .01$) due to the sparseness of data from the small frequencies of the 31 observed response patterns. Ability parameters were estimated with a method of expected a posteriori (Bock & Mislevy, 1982) using the Fortran program written for Kim (2002).

Item and ability parameter estimates of the continuation ratio model from MULTILOG were also obtained. The input files for the MULTILOG run are shown in the Appendix (i.e., FRENF.MLG and the data file without a name, e.g., FRENF.DAT). The exact interpretation of the keywords and command lines can be found in the manual of the computer program MULTILOG (see Thissen et al. 2002; du Toit 2003).

Item parameter estimates and standard errors of the continuation ratio model from the Fortran implementation of the marginal maximum likelihood estimation as well as those from MULTILOG are presented in Table 1. Because the source code of the proprietary program is not in general available, the estimation result from the Fortran implementation based on open source (i.e., the Fortran source code is available from the author) was used here as a reference purpose. All of the item parameter estimates for a given item between two computer programs are very similar. It should be noted that by changing the default settings of the program, it may be possible to obtain exactly the same estimation results.

**Table 1** The continuation ratio model item parameter estimates and standard errors (s.e.) from the Fortran program and MULTILOG

| Program | Item | Item parameter estimate | | | | | |
|---|---|---|---|---|---|---|---|
| | | $a_{j1}$ (s.e.) | $b_{j1}$ (s.e.) | $a_{j2}$ (s.e.) | $b_{j2}$ (s.e.) | $a_{j3}$ (s.e.) | $b_{j3}$ (s.e.) |
| Fortran | 1 | 2.22 (0.65) | −1.36 (0.25) | 2.60 (1.01) | −0.09 (0.26) | 3.89 (1.77) | 1.34 (0.17) |
| | 2 | 2.59 (1.32) | −1.85 (0.37) | 2.72 (1.25) | −0.31 (0.14) | 3.61 (1.11) | 1.12 (0.18) |
| | 3 | 2.14 (0.43) | −1.55 (0.25) | 1.79 (0.48) | −0.34 (0.19) | 3.91 (1.33) | 0.92 (0.17) |
| MULTILOG | 1 | 2.17 (0.61) | −1.38 (0.27) | 2.43 (0.67) | −0.11 (0.15) | 3.68 (1.24) | 1.32 (0.18) |
| | 2 | 2.68 (1.06) | −1.84 (0.32) | 2.96 (0.79) | −0.31 (0.13) | 3.82 (1.28) | 1.11 (0.15) |
| | 3 | 2.17 (0.52) | −1.55 (0.29) | 1.71 (0.47) | −0.37 (0.22) | 4.49 (1.54) | 0.93 (0.13) |

Plots of the category response functions of the three items under the continuation ratio model were obtained and presented in Fig. 2. For each of the items, the monotonic decreasing curve corresponds to the lowest category; the middle two curves correspond to the two middle categories; the monotonic increasing curve corresponds to the highest category. These indicate in each item that the examinees of indefinitely low ability will be assigned the lowest category and, conversely, that examinees of indefinitely high ability will be assigned the highest category. Considering the size of standard errors, these differences may be trivial. In sum, all category response functions from the programs are nearly the same, reflecting the similarity in the item parameter estimates.

Ability estimates from the method of expected a posteriori assuming that item parameter estimates under the continuation ratio model from the Fortran implementation to be true values were obtained and reported in Table 2. Ability estimates were also obtained from MULTILOG. A standard normal prior was used in ability estimation. Due to the similarity of the item parameter estimates, the ability estimates are very similar. One peculiar ability estimate was obtained for the response pattern of 443. The ability estimate was less than those obtained from the response patterns of 441 and 442. A procedure or constraint to prevent to yield illogical ability estimates may be applied in practice.

## 4 Comparisons of Polytomous Models

The same data from the experimental form of the French writing assessment were analyzed to compare models for ordered category items. Category response functions of the items under the graded response model (Samejima 1969), the generalized partial credit model (Muraki 1992), and the partial credit model (Masters 1982) were obtained using MULTILOG. Example input files for various polytomous models can be found in du Toit (2003).

Item parameter estimates under the graded response model, the generalized partial credit model, and the partial credit model are reported in Table 3. It should

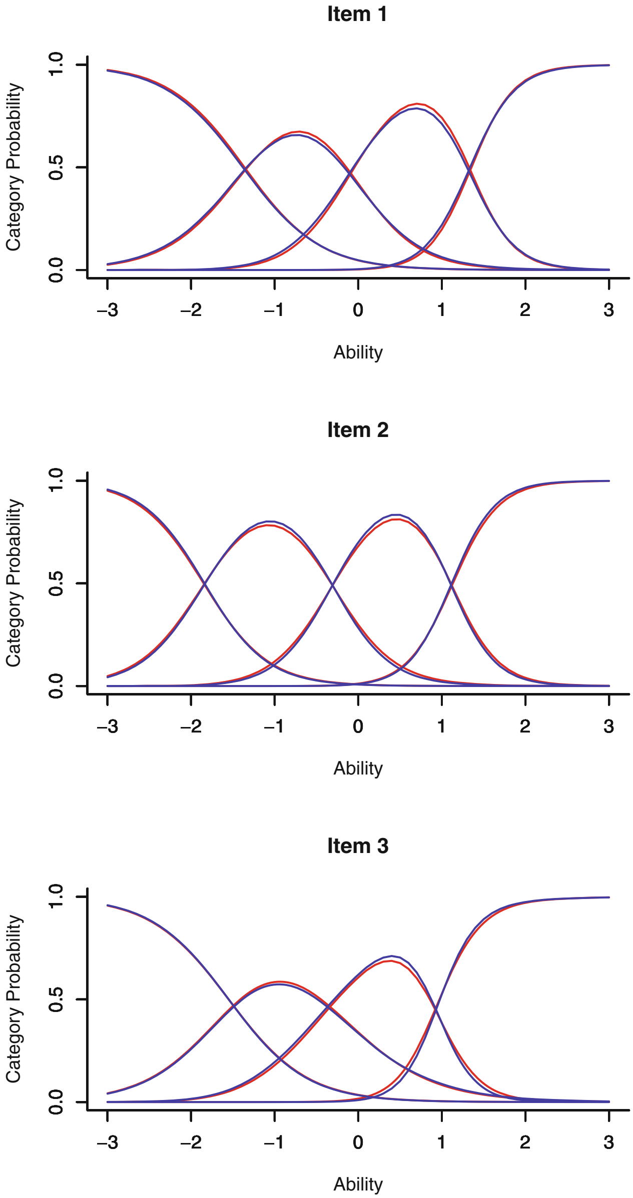

**Fig. 2** Category response functions for items 1–3 under the continuation ratio model from the Fortran program (*red*) and MULTILOG (*blue*)

be noted that the actual, unconstrained parameters estimated in the generalized partial credit model and the partial credit model from MULTILOG are those under the nominal response model. The output from MULTILOG contained both

**Table 2** Expected a posteriori (EAP) ability estimates and the posterior standard deviations (p.s.d.) from the Fortran program and MULTILOG

| | | Program | | | | Program | |
|---|---|---|---|---|---|---|---|
| | | Fortran | MULTILOG | | | Fortran | MULTILOG |
| Pattern | $n$ | EAP (p.s.d.) | EAP (p.s.d.) | Pattern | $n$ | EAP (p.s.d.) | EAP (p.s.d.) |
| 111 | 4 | −2.10 (0.57) | −2.10 (0.53) | 233 | 9 | 0.01 (0.49) | 0.02 (0.40) |
| 112 | 1 | −1.56 (0.44) | −1.61 (0.47) | 323 | 6 | −0.00 (0.49) | −0.05 (0.41) |
| 121 | 4 | −1.47 (0.41) | −1.47 (0.47) | 332 | 4 | 0.23 (0.45) | 0.18 (0.42) |
| 211 | 1 | −1.53 (0.44) | −1.58 (0.48) | 234 | 1 | 0.53 (0.28) | 0.69 (0.36) |
| 122 | 4 | −1.14 (0.49) | −1.09 (0.45) | 243 | 1 | 0.51 (0.27) | 0.60 (0.37) |
| 212 | 1 | −1.19 (0.48) | −1.17 (0.46) | 333 | 24 | 0.42 (0.27) | 0.37 (0.37) |
| 221 | 4 | −1.08 (0.50) | −1.07 (0.45) | 342 | 2 | 0.73 (0.45) | 0.83 (0.39) |
| 123 | 2 | −0.70 (0.47) | −0.75 (0.45) | 423 | 1 | 0.55 (0.31) | 0.54 (0.39) |
| 132 | 1 | −0.51 (0.43) | −0.48 (0.46) | 441 | 1 | 1.41 (0.34) | 1.31 (0.41) |
| 222 | 10 | −0.69 (0.43) | −0.76 (0.42) | 334 | 5 | 0.74 (0.43) | 0.89 (0.31) |
| 312 | 1 | −0.55 (0.46) | −0.63 (0.49) | 343 | 1 | 0.69 (0.40) | 0.82 (0.32) |
| 223 | 5 | −0.46 (0.33) | −0.46 (0.41) | 442 | 1 | 1.46 (0.32) | 1.40 (0.42) |
| 232 | 8 | −0.33 (0.40) | −0.24 (0.42) | 344 | 4 | 1.40 (0.27) | 1.25 (0.33) |
| 322 | 3 | −0.34 (0.39) | −0.33 (0.42) | 434 | 2 | 1.42 (0.25) | 1.24 (0.33) |
| 134 | 1 | 0.52 (0.33) | 0.66 (0.38) | 443 | 1 | 1.40 (0.27) | 1.17 (0.32) |
| Continued to the right-hand-side columns | | | | 444 | 7 | 1.74 (0.48) | 1.81 (0.48) |

unconstrained item parameter estimates as well as those transformed estimates with Bock's (1972) contrasts. The estimates reported under the generalized partial credit model and the partial credit model in Table 3 are the ones actually estimated by MULTILOG (see du Toit 2003 pp. 570–595).

Plots of category response functions obtained from the MULTILOG runs for the continuation ratio model, and the three other polytomous item response theory models are presented in Fig. 3. The third and fourth category response functions from the continuation ratio model seem different from those from the other polytomous item response theory models. The category response functions for item 2 from the graded response model and the generalized partial credit model look nearly the same.

The full-information fit statistics from MULTILOG were $G^2(12) = 40.4$ for the continuation ratio model, $G^2(22) = 45.5$ for the graded response model, $G^2(22) = 50.6$ for the generalized partial credit model, and $G^2(24) = 51.5$ for the partial credit model. All likelihood-ratio goodness-of-fit statistic values were statistically significant (i.e., $p < .01$) and relatively large due to the sparseness of data.

In addition, the Akaike's (1992) AIC (i.e., an information criterion) was obtained. The AIC values were 791.55 for the continuation ratio model, 784.66 for the graded response theory model, 789.75 for the generalized partial credit model, and 786.67 for the partial credit model (see Kang & Cohen, 2007). The graded response model seems to be the best fitting one for the current data. Thissen, Nelson, Rosa, and

**Table 3** Item parameter estimates and standard errors (s.e.) from the graded response (GR) model, the generalized partial credit (GPC) model, and the partial credit (PC) model

| GR model | Item | MULTILOG estimate | | | |
|---|---|---|---|---|---|
| | | $a_j$ (s.e.) | $b_{j1}$ (s.e.) | $b_{j2}$ (s.e.) | $b_{j3}$ (s.e.) |
| | 1 | 2.81 (0.45) | −1.26 (0.17) | −0.08 (0.12) | 1.46 (0.21) |
| | 2 | 3.00 (0.51) | −1.75 (0.22) | −0.31 (0.11) | 1.19 (0.17) |
| | 3 | 2.42 (0.35) | −1.47 (0.22) | −0.26 (0.14) | 1.15 (0.20) |

| GPC model | Item | MULTILOG estimate | | | |
|---|---|---|---|---|---|
| | | $\alpha_j$ (s.e.) | $\gamma_{j1}$ (s.e.) | $\gamma_{j2}$ (s.e.) | $\gamma_{j3}$ (s.e.) |
| | 1 | 2.31 (0.42) | −2.84 (0.64) | −0.23 (0.35) | 3.42 (0.73) |
| | 2 | 2.77 (0.54) | −4.81 (0.99) | −0.86 (0.38) | 3.27 (0.64) |
| | 3 | 1.87 (0.31) | −2.66 (0.54) | −0.54 (0.33) | 2.22 (0.54) |

| PC model | Item | MULTILOG estimate | | | |
|---|---|---|---|---|---|
| | | $\alpha_j$ (s.e.) | $\gamma_{j1}$ (s.e.) | $\gamma_{j2}$ (s.e.) | $\gamma_{j3}$ (s.e.) |
| | 1 | 2.27 (0.23) | −2.79 (0.52) | −0.23 (0.35) | 3.38 (0.59) |
| | 2 | 2.27 (0.23) | −4.11 (0.71) | −0.75 (0.36) | 2.81 (0.51) |
| | 3 | 2.27 (0.23) | −3.11 (0.58) | −0.60 (0.34) | 2.57 (0.52) |

McLeod (2001) reported that the graded response model might fit rating data better than the generalized partial credit model.

Based on the item parameter estimates from the various polytomous item response theory models, the ability parameters were estimated by the method of expected a posteriori using MULTILOG (see Table 4). Ability estimates from the continuation ratio model, the graded response model, the generalized partial credit model, and the partial credit models were very similar. As mentioned in the discussion of Table 2, one peculiar ability estimate was obtained for the response pattern of 443 under the continuation ratio model. Other models for the polytomous items didn’t exhibit such an illogical ability estimate.

## 5 Discussion

The purpose of the present paper was to provide information for the parameter estimation under the continuation ratio model using the Fortran implementation and MULTILOG. An illustration was provided with the performance assessment rating data. Marginal maximum likelihood estimation of item parameters was employed with the method of expected a posteriori for ability estimation. Item parameter estimates from the two programs under the continuation ratio model were very similar, and the ability estimates were also very much alike.

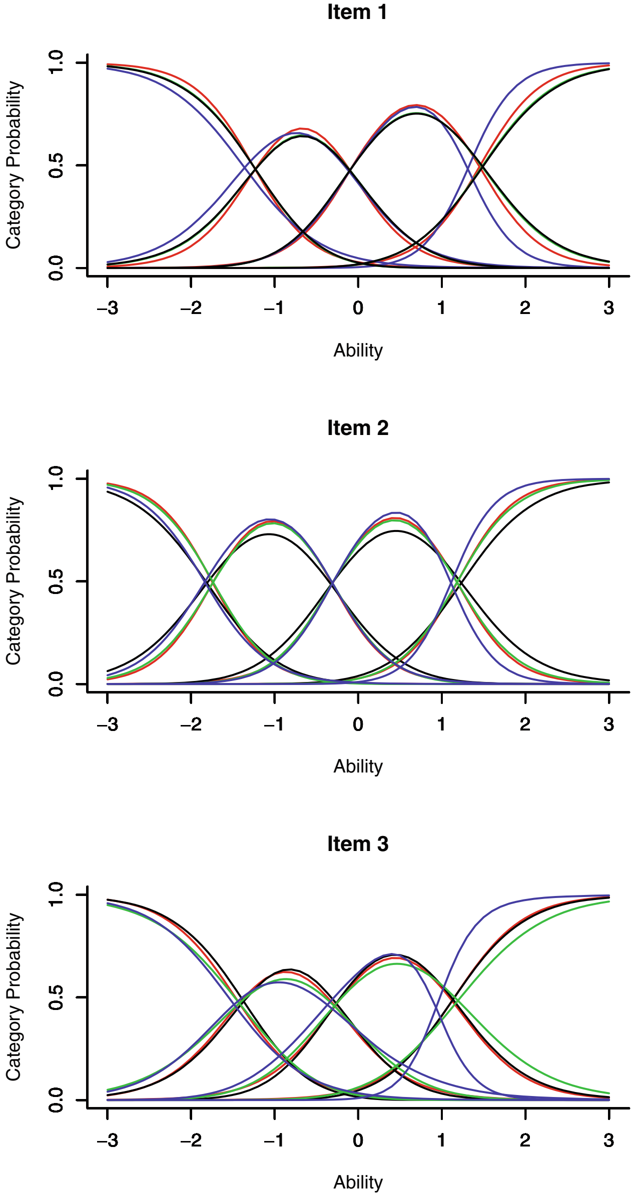

**Fig. 3** Category response functions for the continuation ratio model (*blue*), the graded response model (*red*), the generalized partial credit model (*green*), and the partial credit model (*black*)

In addition, the item and ability parameter estimates under the continuation ratio model were compared with those from the graded response model, the generalized partial credit model, and the partial credit model using MULTILOG. Although the

**Table 4** Expected a posteriori (EAP) ability estimates and the posterior standard deviations (p.s.d.) under the continuation ratio (CR) model, the graded response (GR) model, the generalized partial credit (GPC) model, and the partial credit (PC) model

| | | Model | | | |
|---|---|---|---|---|---|
| | | CR | GR | GPC | PC |
| Pattern | *n* | EAP (p.s.d.) | EAP (p.s.d.) | EAP (p.s.d.) | EAP (p.s.d.) |
| 111 | 4 | −2.10 (0.53) | −2.06 (0.51) | −2.05 (0.52) | −2.03 (0.53) |
| 112 | 1 | −1.61 (0.47) | −1.57 (0.43) | −1.61 (0.45) | −1.50 (0.45) |
| 121 | 4 | −1.47 (0.47) | −1.46 (0.42) | −1.44 (0.43) | −1.50 (0.45) |
| 211 | 1 | −1.58 (0.48) | −1.49 (0.43) | −1.52 (0.44) | −1.50 (0.45) |
| 122 | 4 | −1.09 (0.45) | −1.10 (0.40) | −1.10 (0.41) | −1.09 (0.41) |
| 212 | 1 | −1.17 (0.46) | −1.11 (0.41) | −1.18 (0.42) | −1.09 (0.41) |
| 221 | 4 | −1.07 (0.45) | −1.02 (0.41) | −1.03 (0.41) | −1.09 (0.41) |
| 123 | 2 | −0.75 (0.45) | −0.79 (0.44) | −0.79 (0.40) | −0.71 (0.40) |
| 132 | 1 | −0.48 (0.46) | −0.64 (0.43) | −0.64 (0.40) | −0.71 (0.40) |
| 222 | 10 | −0.76 (0.42) | −0.73 (0.48) | −0.72 (0.40) | −0.71 (0.40) |
| 312 | 1 | −0.63 (0.49) | −0.73 (0.48) | −0.79 (0.40) | −0.71 (0.40) |
| 223 | 5 | −0.46 (0.41) | −0.42 (0.39) | −0.42 (0.40) | −0.34 (0.40) |
| 232 | 8 | −0.24 (0.42) | −0.31 (0.39) | −0.27 (0.40) | −0.34 (0.40) |
| 322 | 3 | −0.33 (0.42) | −0.37 (0.40) | −0.35 (0.40) | −0.34 (0.40) |
| 134 | 1 | 0.66 (0.38) | 0.10 (0.52) | −0.04 (0.41) | 0.04 (0.41) |
| 233 | 9 | 0.02 (0.40) | 0.02 (0.40) | 0.04 (0.41) | 0.04 (0.41) |
| 323 | 6 | −0.05 (0.41) | −0.01 (0.41) | −0.04 (0.41) | 0.04 (0.41) |
| 332 | 4 | 0.18 (0.42) | 0.10 (0.41) | 0.11 (0.41) | 0.04 (0.41) |
| 234 | 1 | 0.69 (0.36) | 0.32 (0.45) | 0.36 (0.42) | 0.44 (0.43) |
| 243 | 1 | 0.60 (0.37) | 0.45 (0.46) | 0.52 (0.43) | 0.44 (0.43) |
| 333 | 24 | 0.37 (0.37) | 0.44 (0.40) | 0.44 (0.42) | 0.44 (0.43) |
| 342 | 2 | 0.83 (0.39) | 0.58 (0.47) | 0.60 (0.43) | 0.44 (0.43) |
| 423 | 1 | 0.54 (0.39) | 0.34 (0.49) | 0.36 (0.42) | 0.44 (0.43) |
| 441 | 1 | 1.31 (0.41) | 1.11 (0.51) | 0.68 (0.43) | 0.44 (0.43) |
| 334 | 5 | 0.89 (0.31) | 0.79 (0.42) | 0.78 (0.43) | 0.86 (0.44) |
| 343 | 1 | 0.82 (0.32) | 0.90 (0.41) | 0.95 (0.44) | 0.86 (0.44) |
| 442 | 1 | 1.40 (0.42) | 1.15 (0.49) | 1.04 (0.44) | 0.86 (0.44) |
| 344 | 4 | 1.25 (0.33) | 1.30 (0.42) | 1.32 (0.46) | 1.32 (0.46) |
| 434 | 2 | 1.24 (0.33) | 1.25 (0.43) | 1.23 (0.45) | 1.32 (0.46) |
| 443 | 1 | 1.17 (0.32) | 1.37 (0.43) | 1.41 (0.46) | 1.32 (0.46) |
| 444 | 7 | 1.81 (0.48) | 1.88 (0.52) | 1.88 (0.54) | 1.88 (0.54) |

overall patterns of the categorical response functions were similar in terms of plots, the continuation ratio model and the partial credit model yielded slightly different results from the graded response model and the generalized partial credit model. The model comparison using AIC indicated that the graded response model was the best fitting model to the data used in the illustration.

As long as the continuation ratio model yields similar item and ability parameters to other polytomous item response theory models as well as comparable information based goodness of fit measures, it can be viewed as an attractive alternative when polytomous items are analyzed. This study used a small data set for only a demonstration purpose. In order to understand the behavior of the item and ability parameter estimates under the continuation ratio model, a more extensive large scale simulation study should be performed.

It should be noted that in the continuation ratio model continuation ratio logits are sequentially used to model the probabilities of obtaining ordered categories in a polytomous item. In order to successfully apply the model to data, this sequential characteristic or nature of the assignment of ordered categories should be present in the construction of data. Inspecting the data if such a characteristic is present seems to be a prerequisite issue before applying logits to a multicategory variable.

In sum, the continuation ratio model considered in this paper can be applied to polytomous response items if they possess a special characteristic that the categories or ordered levels of the response are assigned in a forward, sequential manner. Note that not all polytomous, ordered responses have such a characteristic.

As long as the assumption is satisfied, the continuation ratio model is a unique model for the polytomous items due to the asymptotic independence of the categories within the item (cf. Fienberg 1980 pp. 110–111). Response categories of an item can be separately determined as if those were a set of dichotomous items. Hence, an application of the continuation ratio model in the context of differential item functioning may be promising because category response functions are rather independently obtained so that the category response functions from different groups can be directly compared (cf. Penfield, Gattamorta, & Childs, 2009). This model may also have a good potential use in metric linking and equating for polytomous items because the methods applicable to dichtomous items can be applied without any serious modifications (cf. Kim, Harris, & Kolen, 2010). The continuation ratio model may be a good choice for polytomous items when calibration is required for a test of items with mixed types (i.e., dichotomous and polytomous).

# Appendix

```
FRENF.MLG
L2
>PROBLEM RANDOM, PATTERNS, NITEMS=9, NGROUPS=1, NPATTERNS=31,
 DATA='FRENF.DAT';
>TEST ALL, L2;
>END;
3
019
111111111
Y
9
(4X,9A1,F3.0)

111 099099099  4
```

```
112 099099109  1
121 099109099  4
211 109099099  1
122 099109109  4
212 109099109  1
221 109109099  4
123 099109110  2
132 099110109  1
222 109109109 10
312 110099109  1
223 109109110  5
232 109110109  8
322 110109109  3
134 099110111  1
233 109110110  9
323 110109110  6
332 110110109  4
234 109110111  1
243 109111110  1
333 110110110 24
342 110111109  2
423 111109110  1
441 111111099  1
334 110110111  5
343 110111110  1
442 111111109  1
344 110111111  4
434 111110111  2
443 111111110  1
444 111111111  7
```

## References

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In S. Kotz, & N. L. Johnson (Eds.), *Breakthroughs in statistics: Vol. 1. Foundations and basic theory* (pp. 610–624). New York, NY: Springer. (Reprinted from *Second International Symposium on Information Theory,* pp. 267–281, by B. N. Petrov & F. Csaki, Eds., 1973, Budapest, Hungary: Akademiai Kiado).

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York, NY: Dekker.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29–51. doi:10.1007/BF02291411.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459. doi:10.1007/BF02293801.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6,* 431–444. doi:10.1177/014662168200600405.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, 34,* 187–220.

du Toit, M. (Ed.). (2003). *IRT from SSI.* Lincolnwood, IL: Scientific Software International.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: The MIT Press.

Hambleton, R. K., van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of model building advances. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 21–42). New York, NY: Routledge.

Hemker, B. T., van der Ark, L. A., & Sijtsma, K. (2001). On measurement properties of continuation ratio models. *Psychometrika, 66,* 487–506. doi:10.1007/BF02296191.

Kang, T.-H., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement, 31,* 331–358. doi:10.1177/0146621606292213.

Kim, S.-H. (2002, June). *A continuation ratio model for ordered category items.* Paper presented at the annual meeting of the Psychometric Society, Chapel Hill, NC. Retrieved from http://files.eric.ed.gov/fulltext/ED475828.pdf.

Kim, S.-H. (2013, April). *Parameter estimation of the continuation ratio model.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kim, S., Harris, D. H., & Kolen, J. J. (2010). Equating with polytomous item response models. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory* (pp. 257–292). New York, NY: Routledge.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York, NY: Wiley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174. doi:10.1007/BF02296272.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement, 19,* 91–100. doi: 10.1177/014662169501900110.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176. doi: 10.1177/014662169201600206.

Penfield, R. D, Gattamorta, K., & Childs, R. A. (2009). An NCME instructional module on using differential step functioning to refine the analysis of DIF in polytomous items. *Educational Measurement: Issues and Practice, 28*(1), 38–49. doi: 10.1111/j.1745-3992.2009.01135.x.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika, 50,* 349–364.

Thissen, D., Chen, W.-H., & Bock, R. D. (2002). *MULTILOG: Multiple, categorical item analysis and test scoring using item response theory [Computer software].* Lincolnwood, IL: Scientific Software International.

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen, & H. Wainer (Eds.), *Test scoring* (pp. 141–186). Mahwah, NJ: Lawrence Erlbaum Associates.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology, 43,* 39–55. doi:10.1111/j.2044-8317.1990.tb00925.x

# 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

**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; dmbolt@wisc.edu

L.A. van der Ark et al. (eds.), *Quantitative Psychology Research*, Springer
Proceedings in Mathematics & Statistics 167, DOI 10.1007/978-3-319-38759-8_2

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 disjuctive 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 conjuctively/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 multidimensional 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 $\sigma^2_{\varepsilon_i|\theta}$, 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 $\sigma^2_{\varepsilon_i|\theta}$. Molenaar proposed the following form of heteroscedasticity in the context of polytomously scored items (Molenaar, Dolan, & De Boeck, 2012):

$$\sigma^2_{\epsilon_i|\theta} = 2\delta_0[1 + \exp(-\delta_1\theta)]^{-1} \tag{1}$$

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 $\sigma^2_{\varepsilon_i|\theta} = \delta_0$; 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:

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

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).

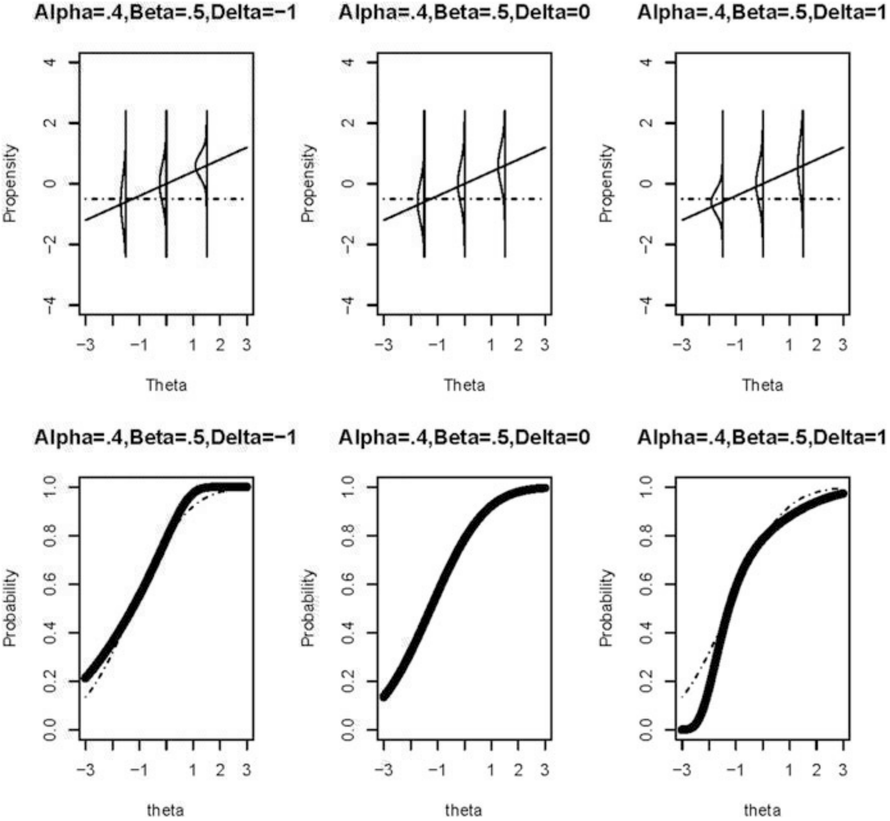

**Fig. 1** Residual heteroscedasticity and item characteristic curves

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:

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

$$\alpha_i \sim \text{Lognormal(0,2)}$$

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

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, $\gamma$:

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

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.,

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

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:

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

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:

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