

# Response Styles in the Partial Credit Model

Applied Psychological Measurement  
2018, Vol. 42(6) 407–427  
© The Author(s) 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0146621617748322  
journals.sagepub.com/home/apm



Gerhard Tutz<sup>1</sup>, Gunther Schauberger<sup>1</sup>, and Moritz Berger<sup>2</sup>

## Abstract

In the modeling of ordinal responses in psychological measurement and survey-based research, response styles that represent specific answering patterns of respondents are typically ignored. One consequence is that estimates of item parameters can be poor and considerably biased. The focus here is on the modeling of a tendency to extreme or middle categories. An extension of the partial credit model is proposed that explicitly accounts for this specific response style. In contrast to existing approaches, which are based on finite mixtures, explicit person-specific response style parameters are introduced. The resulting model can be estimated within the framework of generalized mixed linear models. It is shown that estimates can be seriously biased if the response style is ignored. In applications, it is demonstrated that a tendency to extreme or middle categories is not uncommon. A software tool is developed that makes the model easy to apply.

## Keywords

partial credit model, Likert-type scales, rating scales, response styles, ordinal data, generalized linear models

## Introduction

Response styles are an important problem in psychological measurement and survey data. Various response styles have been identified (for an overview see, for example, Baumgartner & Steenkamp, 2001; Messick, 1991). In Likert-type scales, which represent the level of agreement in the form *strongly disagree*, *moderately disagree*, . . . , *moderately agree*, *strongly agree*, a particularly interesting response style is the extreme response style (ERS) and its counterpart, the tendency to favor middle categories. The problem with response styles is that they can affect the validity of scale scores because estimates of the substantive trait may be biased if the response style is ignored. Models that explicitly account for response styles are able to reduce the bias. They account for additional heterogeneity in the population, which in some applications is itself of interest, in particular, if it is linked to explanatory variables.

Various methods for investigating response styles have been proposed in survey research with a focus on the dependence of the response styles on covariates, for an overview, see Van

<sup>1</sup>Ludwig-Maximilians-Universität München, Germany

<sup>2</sup>Institut für Medizinische Biometrie, Informatik und Epidemiologie, Universitätsklinikum Bonn, München, Germany

## Corresponding Author:

Gerhard Tutz, Ludwig-Maximilians-Universität München, Akademiestraße 1, München 80799, Germany.

Email: [tutz@stat.uni-muenchen.de](mailto:tutz@stat.uni-muenchen.de)

Vaerenbergh and Thomas (2013). Here, the focus is on the modeling of response styles in item response models. In item response data, respondents rate their level of agreement on a series of items and response styles can be considered a consistent pattern of responses that is independent of the content of a response (Johnson, 2003).

In latent trait theory, one can distinguish several approaches to account for response styles by incorporating it into a psychometric model. One approach uses the nominal response model proposed by Bock (1972). Bolt and Johnson (2009), Johnson and Bolt (2010), Bolt and Newton (2011), and Falk and Cai (2016) use the multitrait model to investigate the presence of a response style dimension. Johnson (2003) considered a cumulative type model for ERSs. An alternative strategy for measuring response style is the use of mixture item response theory (IRT). For example, Eid and Rauber (2000) considered a mixture of partial credit models. It is assumed that the whole population can be divided in disjunctive latent classes. After classes have been identified, it is investigated whether item characteristics differ between classes, potentially revealing differing response styles. Finite mixture models for item response data were also considered by Gollwitzer, Eid, and Jürgensen (2005) and Maij-de Meij, Kelderman, and van der Flier (2008). Related latent class approaches were used by Moors (2004), Kankaraš and Moors (2009), Moors (2010), and Van Rosmalen, Van Herk, and Groenen (2010). As Bolt and Johnson (2009) pointed out, in these models, response style is viewed as a discrete qualitative difference in which each respondent is a member of one class, which might be a disadvantage if response style is viewed as a continuous trait.

A quite different more recent strategy uses tree methodology to investigate response styles. Trees, in general, assume a nested structure where first a decision about the direction of the response and then about the strength is obtained. Models of this type have been proposed by Suh and Bolt (2010), De Boeck and Partchev (2012), Thissen-Roe and Thissen, (2013), Jeon and De Boeck (2016), Böckenholt (2012), Khorramdel and von Davier (2014), Plieninger and Meiser (2014), Böckenholt (2017), and Böckenholt and Meiser (2017).

The approach proposed here uses a specific parameterization. In the latent trait model, for each person, an additional parameter is included that indicates whether the person shows a specific tendency to extreme or middle categories. In contrast to mixture models, the response style is a continuous trait that theoretically can take any value. The simultaneous estimation of ability parameters and response style parameters shows whether there is some association between the substantive trait and the response style. The model can be seen as a specific multitrait model with a simple parameterization that continuously models a response style between the poles preference for middle or extreme categories, which is the specific response style considered here. The authors explicitly consider an extension of the partial credit model but the method can also be used to model response styles in alternative ordinal latent trait models.

In “The Extended Partial Credit Model” section, first the partial credit model (PCM) is considered and then the extended model which contains explicit response style parameters is introduced. Also the relation to alternative models is briefly discussed. In “Estimation” section, estimation of parameters is discussed by considering alternative methods. An illustrative example is given in “An Illustrative Example” section. In “Ignoring the Response Style” section, it is investigated how the estimates suffer from the ignorance of response styles. In “Applications” section and the online supplement, further applications that illustrate the method are given.

## **The Extended Partial Credit Model**

Before introducing the model that accounts for the response style, we briefly consider the basic partial credit model.

### The Partial Credit Model

Let  $Y_{pi} \in \{0, 1, \dots, k\}$ ,  $p = 1, \dots, P$ ,  $i = 1, \dots, I$ , denote the ordinal response of person  $p$  on Item  $i$ . The partial credit model (PCM) assumes for the probabilities

$$P(Y_{pi}=r) = \frac{\exp\left(\sum_{l=1}^r \theta_p - \delta_{il}\right)}{\sum_{s=0}^k \exp\left(\sum_{l=1}^s \theta_p - \delta_{il}\right)}, \quad r = 1, \dots, k,$$

where  $\theta_p$  is the person parameter and  $(\delta_{i1}, \dots, \delta_{ik})$  are the item parameters of Item  $i$ . For notational convenience, the definition of the model implicitly uses  $\sum_{k=1}^0 \theta_p - \delta_{ik} = 0$ . With this convention an alternative form is given by

$$P(Y_{pi}=r) = \frac{\exp\left(r\theta_p - \sum_{k=1}^r \delta_{ik}\right)}{\sum_{s=0}^k \exp\left(\sum_{k=1}^s \theta_p - \delta_{ik}\right)}.$$

The PCM was proposed by Masters (1982), see also Masters and Wright (1984). The defining property of the partial credit model is seen if one considers adjacent categories. The resulting presentation

$$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right) = \theta_p - \delta_{ir}, \quad r = 1, \dots, k,$$

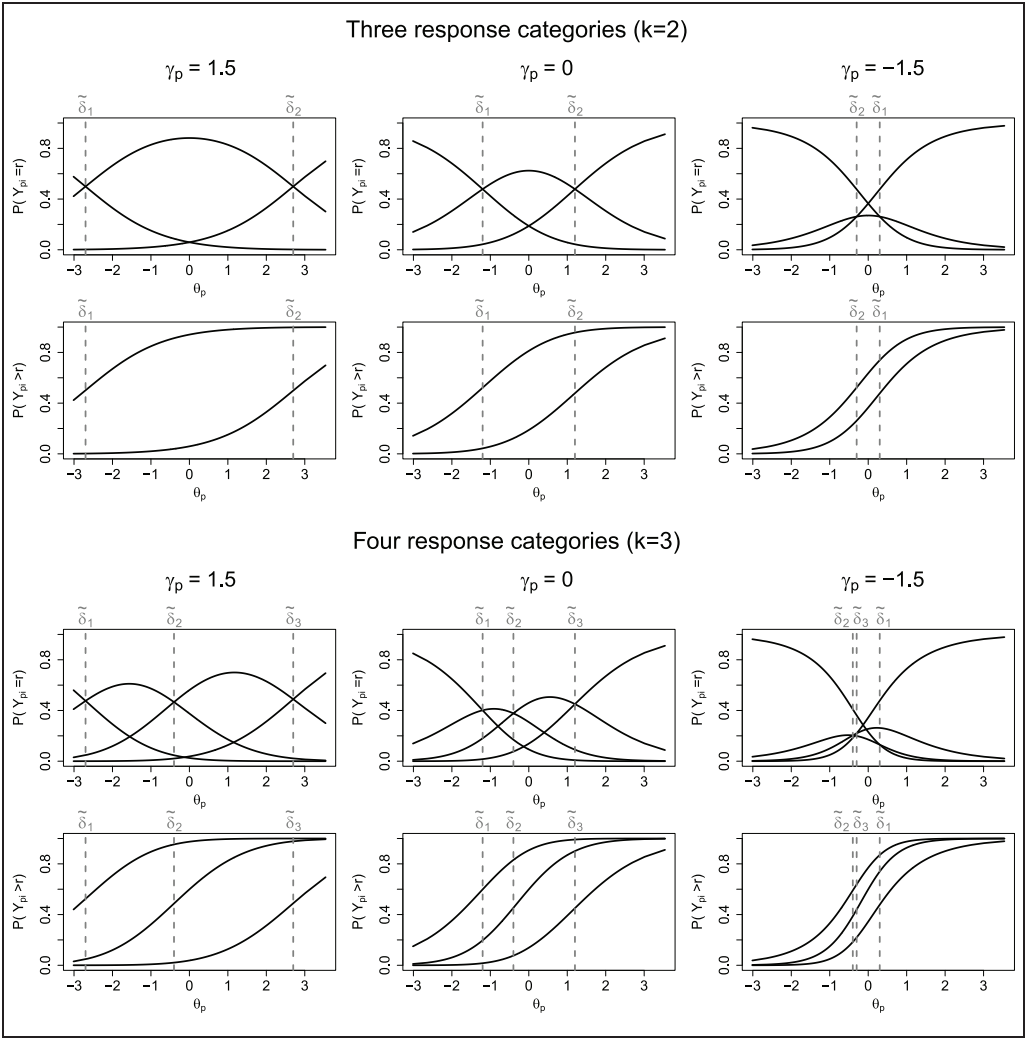
shows that the model is locally (given response categories  $r-1, r$ ) a binary Rasch model with person parameter  $\theta_p$  and item difficulty  $\delta_{ir}$ . It is immediately seen that for  $\theta_p = \delta_{ir}$  the probabilities of adjacent categories are equal, that is,  $P(Y_{pi}=r) = P(Y_{pi}=r-1)$ . That means item response curves never cross, see also middle column of Figure 1.

The partial credit model inherits from the Rasch model a specific property, namely, that the comparison of items does not depend on the person parameters and the comparison of persons does not depend on the item parameters. This special property is often referred to as the specific objectivity of the Rasch model (Irtel, 1995; Rasch, 1966, 1977). More precisely, one obtains for the comparison of items  $i$  and  $j$

$$\log\left(\frac{\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}}{\frac{P(Y_{pj}=r)}{P(Y_{pj}=r-1)}}\right) = -(\delta_{ir} - \delta_{jr}),$$

which does not depend on the person parameter  $\theta_p$ . Thus, items  $i$  and  $j$  can be compared in terms of odds ratios with the odds defined for adjacent categories without the necessity to refer to specific persons. In this sense, it is in accordance with the general principle of specific objectivity that the results of any comparison of two “objects” (items) are independent of the choice of the “agent” (person).

In the same way, the comparison between person  $p$  and person  $\tilde{p}$  is obtained by



**Figure 1.** Probabilities  $P(Y_{pi}=r)$  and item response curves  $P(Y_{pi}>r)$  against  $\theta_p$  for positive, negative  $\gamma_p$ , and  $\gamma_p=0$ .  
Note. In the upper panels, the number of categories is three; in the lower panels, it is four.

$$\log \left( \frac{\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}}{\frac{P(X_{\tilde{pi}}=r)}{P(X_{\tilde{pi}}=r-1)}} \right) = \theta_p - \theta_{\tilde{p}},$$

which does not depend on the the item parameters. Thus, the persons  $p$  and  $\tilde{p}$  can be compared without reference to the item that is used in the measurement.

The authors mention specific objectivity because it is a strong property that separates the measurement from the tool that is used. It will be shown that the extended versions of the model considered in the following are also measurement tools that share the “specific objectivity” property.

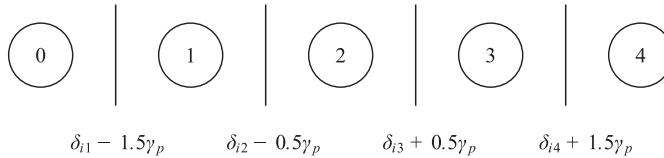
## Explicit Modeling of Response Styles

Let the categories  $0, \dots, k$  represent graded agree–disagree attitudes with a natural symmetry like *strongly disagree*, *moderately disagree*, . . . , *moderately agree*, *strongly agree*. There are two possibilities, the number of categories can be odd with a neutral middle category or the number of categories can be even so that persons have to commit themselves to a positive or negative tendency.

**Odd number of response categories.** Let us start with an odd number of categories, that means  $k$  is even, and let  $m = k/2$  denote the middle category. In the partial credit model, the predictor, when choosing between categories  $r-1$  and  $r$ , has the form  $\eta_{pir} = \theta_p - \delta_{ir}$ . The parameter  $\delta_{ir}$  determines the choice between categories  $r-1$  and  $r$ . Response styles that account for a tendency to extreme categories or a tendency to middle categories are modeled by modifying the thresholds  $\delta_{ir}$ . In the predictor, an additional person parameter  $\gamma_p$  is included that shifts the thresholds of the agreement categories and the thresholds of the disagreement categories into opposite directions. In closed form, the *extended partial credit model with response style* (PCMRS) is given by

$$\log \left( \frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \theta_p + (m-r+0.5)\gamma_p - \delta_{ir} = \theta_p - \tilde{\delta}_{ir}, \quad r = 1, \dots, k, \quad (1)$$

with the modified thresholds defined by  $\tilde{\delta}_{ir} = \delta_{ir} - (m-r+0.5)\gamma_p$ . For illustration, let us consider the case of five response categories ( $k=4$ ,  $m=2$ ). Then the response categories are partitioned into the disagreement categories  $\{0, 1\}$ , the neutral category  $\{2\}$ , and the disagreement categories  $\{3, 4\}$ . The modified thresholds are given by  $\tilde{\delta}_{i1} = \delta_{i1} - 1.5\gamma_p$ ,  $\tilde{\delta}_{i2} = \delta_{i2} - 0.5\gamma_p$ ,  $\tilde{\delta}_{i3} = \delta_{i3} + 0.5\gamma_p$ ,  $\tilde{\delta}_{i4} = \delta_{i4} + 1.5\gamma_p$ , and one obtains the following structure of categories and thresholds.



In general, the parameter  $\gamma_p$  can be seen as a shifting of thresholds. If  $\gamma_p$  is positive, one has for categories  $1, \dots, m$  a shifting of the thresholds  $\delta_{ir}$  to the left, for the agreement categories  $m+1, \dots, k$ , one has a shifting to the right. In both cases the new thresholds are given by  $\tilde{\delta}_{ir} = \delta_{ir} - (m-r+0.5)\gamma_p$ . The scaling factor  $(m-r+0.5)$  on the parameter  $\gamma_p$  determines the direction of the shifting and ensures that the difference between the modified thresholds of adjacent categories changes by a constant, that is,

$$\tilde{\delta}_{ir} - \tilde{\delta}_{i,r-1} = \delta_{ir} - \delta_{i,r-1} + \gamma_p. \quad (2)$$

That means that all the differences become larger for positive  $\gamma_p$  and smaller for negative  $\gamma_p$ . Therefore, for positive  $\gamma_p$ , the probability mass is shifted to the middle categories; for negative  $\gamma_p$ , it is shifted to the extreme categories. It is essential that the modification of the thresholds is done in a symmetric way with a centering in the middle of the response scale.

The meaning of the response style parameter is also seen when considering the extreme values. The extreme case  $\gamma_p \rightarrow \infty$  yields  $P(Y_{pi}=m) \rightarrow 1$ . Then the whole probability mass is in the middle category. For  $\gamma_p \rightarrow -\infty$ , the whole probability mass is in the extreme categories 0 and  $k$ .

Therefore, positive  $\gamma_p$  means that the person has a tendency to middle categories and negative  $\gamma_p$  indicates that the person has a tendency to extreme categories.

A remark to the construction of the extended model concerns the centering of the response style in the middle of the response categories. It can be derived that for any fixed value  $c$  the partial credit model with predictor  $\eta_{pir} = \theta_p + (m-r+c)\gamma_p - \delta_{ir}$  yields an equivalent (reparameterized) model. In particular, the difference property given in equation (2) holds for any chosen value  $c$ . The choice  $c=0.5$  is motivated by the interpretation of the response style parameter  $\gamma_p$ , which can be seen from the local Rasch model:

$$\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} = e^{\theta_p} e^{-\delta_{ir}} = e^{\theta_p} e^{-\delta_{ir}} e^{(m-r+0.5)\gamma_p}.$$

It shows how the odds  $P(Y_{pi}=r)/P(Y_{pi}=r-1)$  are determined in the usual way by the person parameters  $\theta_p$  and  $\delta_{ir}$ . The response style parameter modifies the odds by the factor  $e^{(m-r+0.5)\gamma_p}$  in a symmetric way. The symmetry around the middle category is seen more easily in the example  $k=4$  used before. Then the factor has values  $e^{1.5\gamma_p}$ ,  $e^{0.5\gamma_p}$ ,  $e^{-0.5\gamma_p}$ , and  $e^{-1.5\gamma_p}$ . A consequence of this symmetry is that the factor that compares the lowest categories 0 and 1 and the highest categories  $k$  and  $k-1$  are always the same. In general, one obtains for the factor  $e^{(m-r+0.5)\gamma_p}$ :

$$e^{(0.5-m)\gamma_p} \quad \text{for } r=k, \quad e^{-(0.5-m)\gamma_p} \quad \text{for } r=1.$$

Thus, the tendency to prefer category  $k$  over  $k-1$  is the same as the tendency to prefer 0 over 1, namely,  $e^{(0.5-m)\gamma_p}$ . This symmetry is fulfilled only if  $c$  is fixed at 0.5. Of course, similar relationships can be derived for other categories, for example categories 1 over 2 and  $k-1$  over  $k-2$  ( $k \geq 4$ ).

To illustrate the response style effect, we show in Figure 1 the probabilities of response categories (upper row) and the cumulative probability functions  $P(Y_{pi} > r)$  (lower row) as functions of the person abilities for varying response style parameter  $\gamma_p$ . The upper panels show the case  $k=2$  (three response categories), the lower show  $k=3$  (the model for an even number of response categories is considered later). It is immediately seen that for  $\gamma_p = -1.5$  the extreme categories have much higher probabilities than for  $\gamma_p = 0$ . The inverse is seen for  $\gamma_p = 1.5$ . It is noteworthy that the probabilities of adjacent categories are now equal, that is,  $P(Y_{pi}=r) = P(Y_{pi}=r-1)$ , if  $\theta_p = \tilde{\delta}_{ir}$ .

The extended PCMRS model still shows specific objectivity. For the comparison of items  $i$  and  $j$  one obtains again

$$\log \left( \frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) - \log \left( \frac{P(Y_{pj}=r)}{P(Y_{pj}=r-1)} \right) = -(\delta_{ir} - \delta_{jr}),$$

which does not depend on the person parameter  $\theta_p$  or the response style parameter  $\gamma_p$ . Therefore, the comparison of item difficulties does not depend on the persons. For the comparison between person  $p$  and person  $\tilde{p}$  one obtains

$$\log \left( \frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) - \log \left( \frac{P(Y_{\tilde{p}i}=r)}{P(Y_{\tilde{p}i}=r-1)} \right) = \theta_p - \theta_{\tilde{p}} + (m-r+0.5)(\gamma_p - \gamma_{\tilde{p}}),$$

which does not depend on the the item parameters. Therefore, the comparison of persons is just a function of the person parameters, which now includes the ability parameters  $\theta_p, \theta_{\bar{p}}$  and the response style parameters  $\gamma_p, \gamma_{\bar{p}}$ .

It should be noted that a constraint is needed to obtain identifiable parameters. One can use, for example,  $\theta_p = 0$ , or the symmetric side constraint  $\sum_{p=1}^P \theta_p = 0$ . If not mentioned otherwise, the authors will use the symmetric side constraint.

*Even number of response categories.* If  $k$  is odd, that means the number of categories is even, one has a split into agreement and disagreement categories. Let  $m$  again be defined by  $m = k/2$ . However, now  $m$  is not an integer, and one obtains the disagreement categories  $\{0, \dots, \lfloor m \rfloor\}$  and the agreement categories  $\{\lfloor m \rfloor, \dots, k\}$ , where  $\lfloor \cdot \rfloor$  denotes the floor function. Using the real-valued number  $m$ , the extended PCMRS has the same form for  $k$  odd and  $k$  even, that is,

$$\log \left( \frac{P(Y_{pi} = r)}{P(Y_{pi} = r - 1)} \right) = \theta_p + (m - r + 0.5)\gamma_p - \delta_{ir} = \theta_p - \tilde{\delta}_{ir}, \quad r = 1, \dots, k,$$

where  $\tilde{\delta}_{ir} = \delta_{ir} - (m - r + 0.5)\gamma_p$ . For  $k = 5$ , one obtains  $m = 2.5$ ,  $\lfloor m \rfloor = 2$  and the partitioning into categories  $\{0, 1, 2\}$  and  $\{3, 4, 5\}$  with modified thresholds  $\tilde{\delta}_{i1} = \delta_{i1} - 2\gamma_p$ ,  $\tilde{\delta}_{i2} = \delta_{i2} - \gamma_p$ ,  $\tilde{\delta}_{i3} = \delta_{i3}$ ,  $\tilde{\delta}_{i4} = \delta_{i4} + \gamma_p$ , and  $\tilde{\delta}_{i5} = \delta_{i5} + 2\gamma_p$ . It is noteworthy that for  $k = 1$  one obtains the binary Rasch model as a special case. Therefore, using the real-valued “middle category,”  $m = k/2$  the extended partial credit model can be seen as a generalization of the binary Rasch model to ordered categories.

It is easily seen that for  $k > 1$  again equation (2) holds so that the difference between adjacent thresholds changes by the value  $\gamma_p$ . The response style parameter is again included in a symmetric way yielding the same tendency to extreme categories as in the case of an odd number of response categories.

It should be noted that a similar concept of explicit modeling of a tendency to middle or extreme categories has been used before by Tutz and Berger (2016). However, the authors considered just one item and modeled the effect of covariates. Therefore, no explicit response style parameter, which varies in the population, is present and estimation methods are quite different.

### Embedding Into the Framework of Multidimensional Item Response Models

One approach to model response styles, which has been intensively investigated in recent years, is based on multidimensional IRT models. For nominal models, this approach has been used by Bolt and Johnson (2009), Johnson and Bolt (2010), Bolt and Newton (2011), and Falk and Cai (2016), for the partial credit model by Wetzel and Carstensen (2017) and Plieninger (2016). The partial credit model version uses the extension of the partial credit model to multidimensional data as presented by Kelderman (1996). In the notation used here, the model considered by Wetzel and Carstensen (2017) has the form

$$P(Y_{pi} = r) = \frac{\exp \left( \sum_{l=1}^r \left( \sum_{d=1}^D \omega_{ird} \theta_{pd} - \delta_{il} \right) \right)}{\sum_{s=0}^k \exp \left( \sum_{l=1}^s \left( \sum_{d=1}^D \omega_{ird} \theta_{pd} - \delta_{il} \right) \right)}, \quad r = 1, \dots, k,$$

yielding



$$\log \left( \frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)} \right) = \eta_{pir} = \sum_{d=1}^D \omega_{ird} \theta_{pd} - \delta_{ir}, \quad r = 1, \dots, k,$$

where the person parameter  $\theta_p^T = (\theta_{p1}, \dots, \theta_{pD})$  is now a vector with  $D$  dimensions and  $\omega_{ird}$  are scoring weights. If  $D=1$  and for all items  $\omega_{ird}=1$ , one obtains the simple partial credit model. The scoring  $\omega_{ird}$  can be used to define the dimensions. With weights  $\omega_{ird}=1$  for the first dimension, it is considered the trait to be measured. Alternative weights may be used for the other dimensions defining them as response style parameters. When writing first versions of the present article, the authors were not aware of the strong connection of their approach to the extended partial credit model. However, their approach can be embedded into this general framework. The authors use two dimensions, one for the trait to be measured and the other for the response style. An advantage of their approach is the simple parameterization, which has a simple interpretation in the shifting of thresholds. Moreover, the parameterization uses one person parameter to cover the continuum between the preference of middle categories and extreme categories.

A different class of models, which is worth mentioning as at first sight there are similarities, is the class of random-threshold models as proposed by Johnson (2003), Johnson, (2014), Wang, Wilson, and Shih (2006), and Jin and Wang (2014). The ERS model with partial crediting modeling (ERS-PCM) propagated by Jin and Wang (2014) uses the linear predictor  $\eta_{pir} = \theta_p - (\alpha_i + w_p \delta_{ir})$ ,  $r = 1, \dots, k$ , where  $w_p$  is a weight parameter of respondent  $p$  on thresholds and is assumed to follow a log-normal distribution with a mean of one and variance of  $\sigma_w$ . According to Jin and Wang (2014), the model is designed to account for the ERS only, that is, a systematic tendency for a person to endorse extreme options. Although the approach considered here uses an additive effect, which is centered in the middle, the random-threshold model assumes an multiplicative effect. Multiplicative effects of two unobserved values are much harder to estimate; therefore, Jin and Wang (2014) use Bayesian methods. An additional assumption of the ERS-PCM, which is a restrictive assumption, is that the trait parameter  $\theta_p$  and the weight parameter  $w_p$  are independent.

### Alternative Approaches

Several alternative approaches to the modeling of response styles within a latent trait framework have been considered in the literature. An approach with some tradition is the mixture approach as considered, for example, by Rost (1991) and von Davier and Rost (1995), for an overview on mixture and latent class models, see also von Davier and Yamamoto (2007). Mixture models assume that respondents come from different latent classes. Different item response models are fitted within these classes, some may represent the substantive trait, some may represent response style behavior. A problem that arises when using mixture models is the number of classes, which is unknown. One might get quite different models if one fits, for example, two or three classes, as all the parameters can change when considering one more class. If one has chosen a number of classes, it is sometimes still difficult to interpret the difference between classes and explain what feature is represented by a class, it might be a response style or some other dimension that is involved when responding to items. As the classes are not prespecified, for example, by explicitly modeling of response style behavior, there is much uncertainty involved and the interpretation of the model within classes sometimes tends to be vague. An advantage of the PCMRS model is the explicit modeling of the response style, which allows to decide whether it is present, and if, how strong it is. In mixture models, response styles are viewed as a discrete trait and respondents are classified into multiple classes



representing different response behavior (Bolt & Johnson, 2009). If one considers response style as a continuous trait, as is quite common in parts of the literature, alternative models are more appropriate.

More recently, response styles have been modeled by tree methodology also considered as multiprocess models, see, for example, Suh and Bolt (2010), De Boeck and Partchev (2012), Böckenholt, (2012), Plieninger and Meiser (2014), and Böckenholt (2017). The methods provide flexible tools to model item response data. In item response trees, one can also model various distinct response style traits separately. However, flexibility seems to come at a price. One has typically many options how to construct a tree. As Böckenholt (2017) mentions “the application of IR-tree models requires conceptualizing how respondents mentally translate rating questions and map their internal assessments onto the rating scale” (p. 80). The construction of a corresponding tree uses binary decisions taken in latent binary subitems. As the subitems are not observed, it is sometimes hard to decide which tree is appropriate. Nevertheless, the method is a strong and flexible tool with a potential far beyond response styles.

## Estimation

As the PCM, also the PCMRS can be embedded into the framework of (multivariate) generalized linear models (GLMs). Then estimates can be obtained using program packages that fit multivariate GLMs. This joint likelihood approach yields estimates for all the parameters.

The embedding into the framework of GLMs is obtained in the following way. Let the parameters be collected in the vectors  $\boldsymbol{\theta}^T = (\theta_1, \dots, \theta_{P-1})$ ,  $\boldsymbol{\delta}_i^T = (\delta_{i1}, \dots, \delta_{ik})$ ,  $\boldsymbol{\gamma}^T = (\gamma_1, \dots, \gamma_{P-1})$ . With  $\mathbf{1}_m^{(q)}$  denoting a unit vector of length  $q$  with a 1 in component  $m$  one has for even  $k$ :

$$\log \frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)} = \left(\mathbf{1}_p^{(P-1)}\right)^T \boldsymbol{\theta} - \left(\mathbf{1}_r^{(k)}\right)^T \boldsymbol{\delta}_i + (m-r+0.5) \left(\mathbf{1}_p^{(P)}\right)^T \boldsymbol{\gamma}.$$

Therefore, by appropriate specifications of the model components joint ML (maximum likelihood) estimates can in principle be obtained using the R package VGAM (Yee, 2014).<sup>a</sup> For odd  $k$ , the predictor has to be modified accordingly.

A disadvantage of the joint likelihood estimation is that many parameters have to be estimated, which makes estimates unstable. Moreover, estimates typically are asymptotically biased as the increase of the number of persons also increases the number of parameters to be estimated. An alternative method, which works much better, is marginal likelihood estimation. To reduce the number of parameters, one assumes that the response style parameters are drawn from a normal distribution  $N(0, \sigma^2)$ . The corresponding marginal likelihood with  $\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1^T, \dots, \boldsymbol{\delta}_I^T)$  is

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2) = \prod_{p=1}^P \int P(\{Y_{p1}, \dots, Y_{pI}\}) f(\gamma_p) d\gamma_p,$$

where  $f(\gamma_p)$  is the density  $N(0, \sigma^2)$  of the random effects. The corresponding log likelihood (for even  $k$ ) simplifies to

$$l(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2) = \sum_{p=1}^P \log \left( \int \prod_{i=1}^I \prod_{r=1}^k \left\{ \frac{\exp\left(\sum_{l=1}^r \theta_p + (m-r+0.5)\gamma_p - \delta_{il}\right)}{\sum_{s=0}^k \exp\left(\sum_{l=1}^s \theta_p + (m-r+0.5)\gamma_p - \delta_{il}\right)} \right\}^{y_{pir}} f(\gamma_p) d\gamma_p \right),$$

where  $y_{pir} = 1$  if  $Y_{pi} = r$  and  $y_{pir} = 0$  otherwise. Maximization of the marginal log likelihood can be obtained by integration techniques.

Typically, one first wants to obtain good estimates of the item parameters and estimate person parameters later for the validated test tool. Therefore, one also assumes a distribution for the person effects, which yields the marginal likelihood:

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \prod_{p=1}^P \int P(\{Y_{p1}, \dots, Y_{pI}\}) f(\gamma_p, \theta_p) d\gamma_p d\theta_p,$$

where  $f(\gamma_p, \theta_p)$  now denotes the two-dimensional density of the person parameters,  $N(\mathbf{0}, \boldsymbol{\Sigma})$ . The diagonals of the matrix  $\boldsymbol{\Sigma}$  contain the variance of the response style parameters  $\sigma_\gamma^2$  and the variance of the person effects,  $\sigma_\theta^2$ , the off diagonals are the covariances between response style and location effects,  $\text{cov}_{\gamma\theta}$ . The corresponding log likelihood is

$$l(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \sum_{p=1}^P \log \left( \int \prod_{i=1}^I \prod_{r=1}^k \left\{ \frac{\exp \left( \sum_{l=1}^r \theta_p + (m-r+0.5)\gamma_p - \delta_{il} \right)}{\sum_{s=0}^k \exp \left( \sum_{l=1}^s \theta_p + (m-r+0.5)\gamma_p - \delta_{il} \right)} \right\}^{y_{pir}} f(\gamma_p, \theta_p) d\gamma_p d\theta_p \right).$$

The embedding into the framework of generalized mixed models allows to use methods that have been developed for this class of models. One strategy is to use joint maximization of a penalized log likelihood with respect to parameters and random effects appended by estimation of the variance of random effects, see Breslow and Clayton (1993) and McCulloch and Searle (2001). However, joint maximization algorithms tend to underestimate the variances and, therefore, the true values of the random effects. An alternative strategy, which is used here, is numerical integration by Gauss-Hermite integration methods. Early versions for univariate random effects date back to Hinde (1982) and Anderson and Aitkin (1985). For an overview on estimation methods for generalized mixed model see McCulloch and Searle (2001) and Tutz (2012).

## An Illustrative Example

As an example, the authors consider data from the ALLBUS, the general survey of social science carried out by the German institute GESIS. They are available from <http://www.gesis.org/allbus>. The authors use data containing the answers of 2,535 respondents from the questionnaire in 2012. One part of the survey, comprising eight questions, asks for the degree of confidence in public institutions and organizations. Examples are the federal constitutional court, the justice, or the political parties. The answers are all measured on a scale from 1 (*no confidence at all*) to 7 (*excessive confidence*). The authors fitted a simple PCM and the extended PCMRS using scaled shifting of thresholds. In both cases, marginal estimation is applied assuming normally distributed person parameters for the PCM and two-dimensional normally distributed person and response style parameters for the PCMRS. The estimated variance of the person parameters when fitting the PCM is  $\hat{\sigma}^2 = 0.682$ . When fitting the PCMRS one obtains the estimated covariance matrix between person and response style parameters:

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\sigma}_\theta^2 & \hat{\text{cov}}_{\gamma\theta} \\ \hat{\text{cov}}_{\gamma\theta} & \hat{\sigma}_\gamma^2 \end{pmatrix} = \begin{pmatrix} 0.773 & 0.025 \\ 0.025 & 0.246 \end{pmatrix}.$$

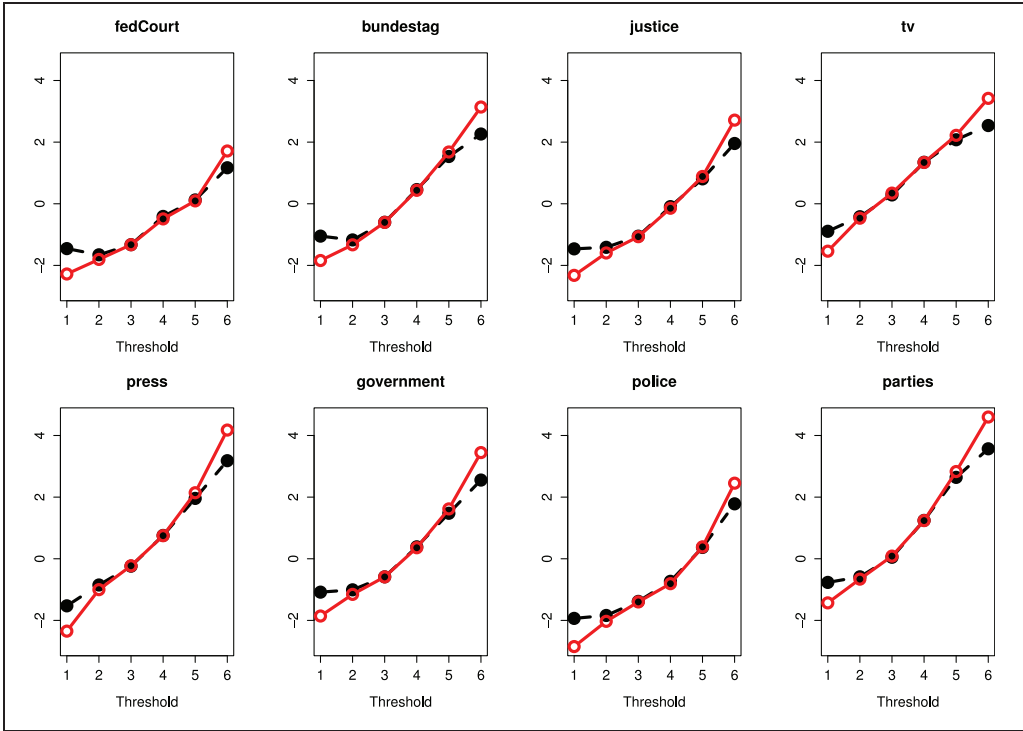
In this application, the correlation between the ability parameters and the response style parameters is rather small ( $\hat{\rho}_{\gamma\theta} = .058$ ). However, the standard deviation of the response style parameters ( $\hat{\sigma}_{\gamma} = 0.496$ ) indicates that response styles should not be neglected when analyzing the data. The estimates of the item parameters are shown in Figure 2, separately for each item. With seven response categories, there are six thresholds each. The red, solid lines correspond to the estimates for the extended PCMRS and the black, dashed lines correspond to the estimates for the simple PCM. It is striking that the estimates for the first and the last threshold strongly differ for all items, while the middle thresholds are fairly equal. If the presence of response style parameters is ignored in particular, the parameters of extreme categories seem to be attenuated. The effect might be connected to the variance of the latent trait  $\theta_p$ . The variance of  $\theta$  raises from .682 to .773 if the model accounts for the response style. Therefore, the variability in the population seems to be underestimated if the response style is ignored. As will be shown later, ignoring the response style has also consequences for the estimation of person parameters.

Although many estimates of the item parameters coincide, there are still big differences between both models due to the presence of the response styles. For illustration, the item response curves for the Item “justice” are given in Figure 3. The curves are plotted against person parameters, which are chosen between the 10% and the 90% quantile of the person parameters according to its estimated one-dimensional normal distribution  $\hat{\theta} \sim N(0, 0.773)$ . Again, the solid lines show the estimates for the extended PCMRS and the black lines show the estimates for the simple PCM. The first row corresponds to the curves for category 1, the second row to the curves for category 4, and the third row shows the curves for category 7. The middle panel represents persons without response style ( $\gamma_p = 0$ ), the left panel represents persons with a tendency to the middle ( $\gamma_p = \sigma_{\gamma}$ ), and the right panel represents persons with a tendency to the extremes ( $\gamma_p = -\sigma_{\gamma}$ ). With values  $\gamma_p = -\hat{\sigma}_{\gamma}$  and  $\gamma_p = \hat{\sigma}_{\gamma}$ , the left and the right figures represent the extremes of a continuum containing 68% of the population. Obviously, the curves obtained by the PCM are the same in each case. There are only minor differences between the two models for persons without response styles. However, for persons with  $\gamma_p = \sigma_{\gamma}$ , the probability for the extreme categories decreases while the probability for category 4 strongly increases when fitting the extended PCMRS. For persons with  $\gamma_p = -\sigma_{\gamma}$  the effect is the opposite.

## Ignoring the Response Style

The illustrative example showed that there are notable differences between the fits of the simple PCM and the extended PCMRS if response styles are present. In particular, the item parameters of the extreme categories were attenuated in the PCM. In the following simulations, it is demonstrated that this is the result of strongly biased estimates of the  $\delta$ -parameters, which are the parameters of interest in most studies.

In the simulation study, the authors consider exemplarily several settings with  $P = 500$  persons and  $I = 10$  items composed of seven categories ( $k = 6$ ). The data generating model is the PCMRS with  $k$ -dimensional normally distributed item parameters,  $\delta_i \sim N_k(\mathbf{0}, \mathbf{I})$ . The person parameters are drawn from a two-dimensional normal distribution  $(\theta_p, \gamma_p) \sim N_2(\mathbf{0}, \Sigma)$ , where  $\sigma_{\theta}^2 = 1$ . In each setting, the variance of the response style parameters varies:  $\sigma_{\gamma}^2 \in \{0, .0.4\}$ . For  $\sigma_{\gamma}^2 = 0$ , the data generating model corresponds to the simple PCM. The authors consider simulations with and without correlation between the ability parameters and the response style parameters ( $\rho_{\gamma\theta} = 0$  or  $\rho_{\gamma\theta} = .3$ ). In many applications, the item parameters are ordered, that is,  $\delta_{i1} \leq \dots \leq \delta_{ik}$ , but that is not necessarily the case. Therefore, the authors distinguish between ordered (ascending) and nonordered item parameters. To obtain ordered item parameters, they



**Figure 2.** Estimates of item parameters (ALLBUS).

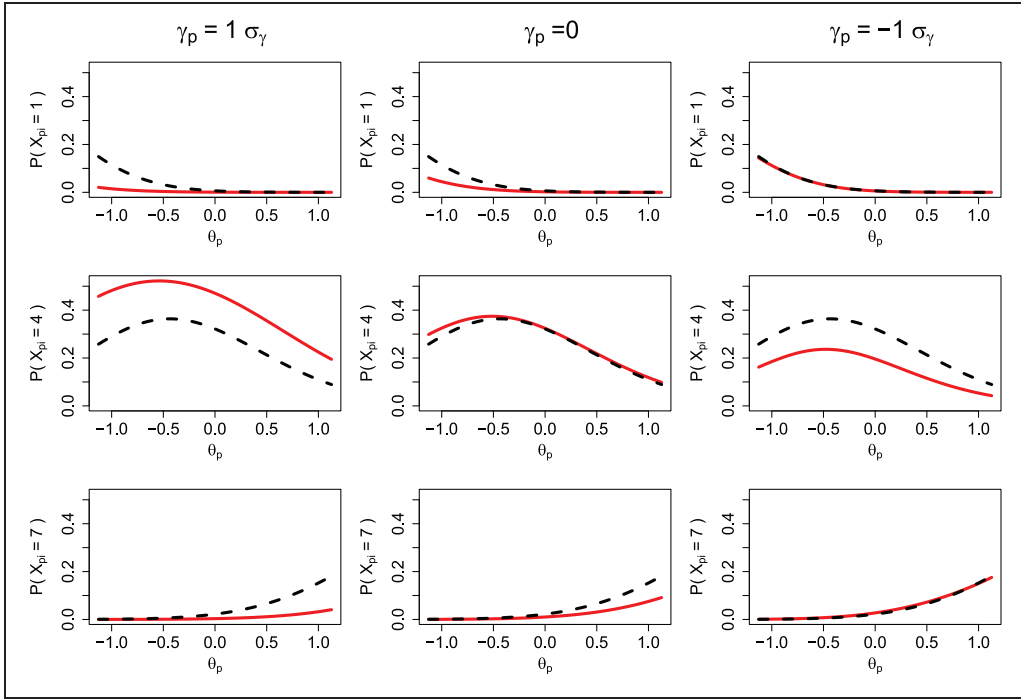
Note. Solid (red) line represents estimates for PCMRS, dashed (black) line represents estimates for PCM. PCMRS = partial credit model with response style; PCM = partial credit model.

are first drawn from the multivariate normal distribution and subsequently ordered by size. In each setting, 100 data sets were generated.

Figure 4 shows the boxplots of the mean squared errors (MSEs) summarized over all item parameters, computed by  $\frac{1}{60}(\sum_{i=1}^{10} \sum_{r=1}^6 (\hat{\delta}_{ir} - \delta_{ir})^2)$ , for the two settings with ordered item parameters. For each value of  $\sigma_\gamma^2$ , the results of the PCMRS are given on the left (not colored) and the results of the PCM are given on the right (gray colored). In both settings (with and without correlation), it is seen that the MSEs are very similar for very small values of  $\sigma_\gamma^2$ , but for large values of  $\sigma_\gamma^2$  the MSEs are much larger if the response style is ignored. The picture is very similar in the case of nonordered item parameters (Figure 5) but the increase of the MSEs with growing  $\sigma_\gamma^2$  is slightly weaker.

The poor estimation accuracy of the PCM is mainly caused by the bias, which is illustrated for one item in Figure 6. The figure shows the bias of the item parameters  $\delta_{1r}$ ,  $r = 1, \dots, 6$  for the setting with ordered item parameters (left panel) and for the setting with nonordered item parameters (right panel) without correlation ( $\rho_{\gamma\theta} = 0$ ). One obtains strongly biased estimates even for moderate values of  $\sigma_\gamma^2$  when fitting the PCM (solid lines). Conspicuously, in the ordered case, mainly the parameters of the extreme Categories 1 and 6 are affected. The positive bias for parameter  $\delta_{11} = -1.25$  and the negative bias for parameter  $\delta_{16} = 1.71$  both indicate that the parameter sizes are strongly underestimated. These effects were already seen in the illustrative example.

In the nonordered case, one observes biased results for all categories and the bias is less systematic. Again, a positive bias for the parameter  $\delta_{11} = -0.84$  corresponds to an attenuation of



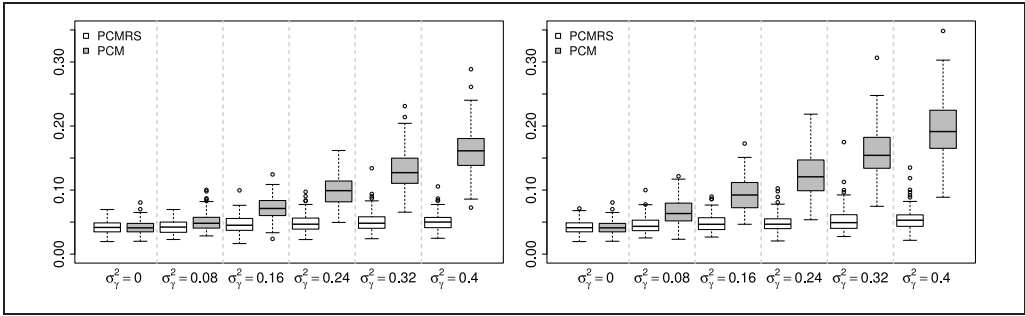
**Figure 3.** Response curves for Item “Justice” (ALLBUS) along person parameter  $\theta$  (between 10% and 90% quantile) for Category 1 (upper panel), Category 4 (middle panel), and Category 7 (lower panel). Note. Columns represent different response styles: tendency to the middle (left), no response style (middle), and tendency to the extremes (right). Solid (red) lines represent estimates for PCMRs, dashed (black) lines represent estimates for PCM. PCMRs = partial credit model with response style; PCM = partial credit model.

the effect, but, for example, the positive bias for parameter  $\delta_{15} = 1.71$  indicates an overestimation of the effect. These findings are very similar for the settings with correlation  $\rho_{\gamma\theta} = .3$  (not shown).

The estimates of the components of the covariance matrix  $\Sigma$  of the person parameters  $(\theta_p, \gamma_p)$  are given in Figure 7. It shows the results for the settings with ordered item parameters, without correlation (left panel) and with correlation (right panel). The true values are marked by red crosses. Both models yield estimates of the variance of the ability parameters  $\sigma_\theta^2$  (given in the first row), whereas only the extended PCMRs yields estimates of the variance of the response style parameters  $\sigma_\gamma^2$  and the covariance  $cov_{\gamma\theta}$ . It is seen that both components are estimated with sufficient accuracy in both settings also for high values of  $\sigma_\gamma^2$ . In the setting with correlation, the covariance (given in the third row) increases with increasing  $\sigma_\gamma^2$ . The variance of the ability parameters is slightly underestimated by the PCMRs. In summary, the structure parameters contained in the covariance matrix are estimated rather well.

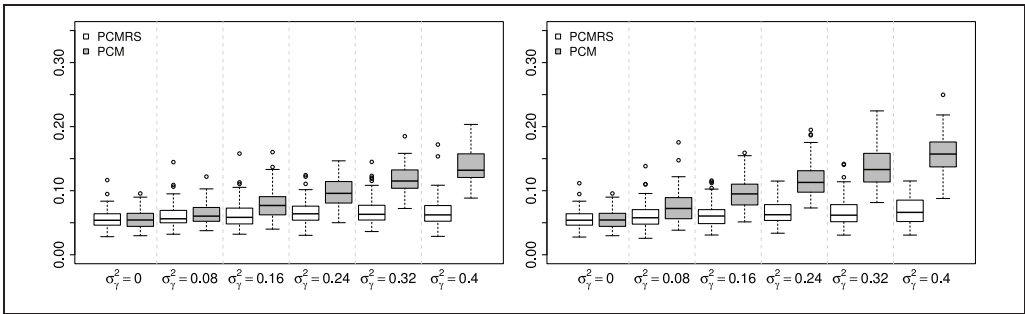
### Effect on Estimated Person Parameters

If the response style is ignored, estimates of the item parameters can be strongly biased. Consequently, also the estimated person parameters will be affected. Posterior mean estimates of person parameters  $\alpha_p = (\theta_p^m, \gamma_p^m)^T$  can be obtained as,



**Figure 4.** MSEs of all item parameters for the simulations with ordered item parameters.

Note. The correlation is  $\rho_{\gamma 0} = 0$  (left) and  $\rho_{\gamma 0} = .3$  (right). MSE = mean squared error; PCMRs = partial credit model with response style; PCM = partial credit model.



**Figure 5.** MSEs of all item parameters for the simulations with nonordered item parameters.

Note. The correlation is  $\rho_{\gamma 0} = 0$  (left) and  $\rho_{\gamma 0} = .3$  (right). MSE = mean squared error; PCMRs = partial credit model with response style; PCM = partial credit model.

$$\alpha_p^m = E(\alpha_p | Y_p) = \int \alpha_p f_p(\alpha_p | Y_p, \hat{\delta}, \hat{\Sigma}) d\alpha_p$$

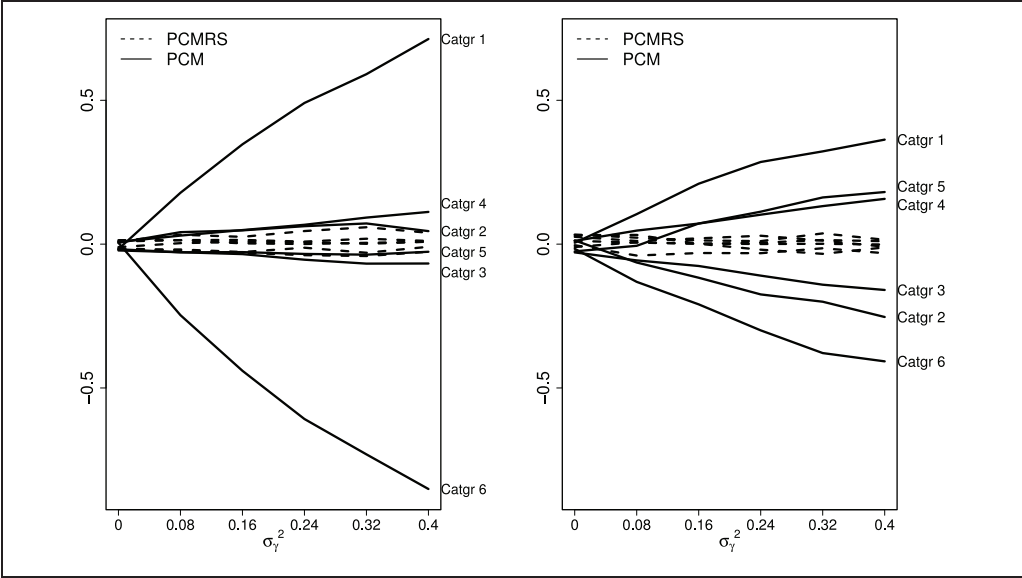
where  $Y_p^T = (Y_{p1}, \dots, Y_{pI})$ ,  $\hat{\delta}^T = (\hat{\delta}_{11}, \dots, \hat{\delta}_{lk})$ , and

$$f_p(\alpha_p | Y_p, \hat{\delta}, \hat{\Sigma}) = \frac{P(Y_p | \alpha_p, \hat{\delta}) f(\alpha_p | \hat{\Sigma})}{\int P(Y_p | \alpha_p) f(\alpha_p | \hat{\delta}, \hat{\Sigma}) d\alpha_p},$$

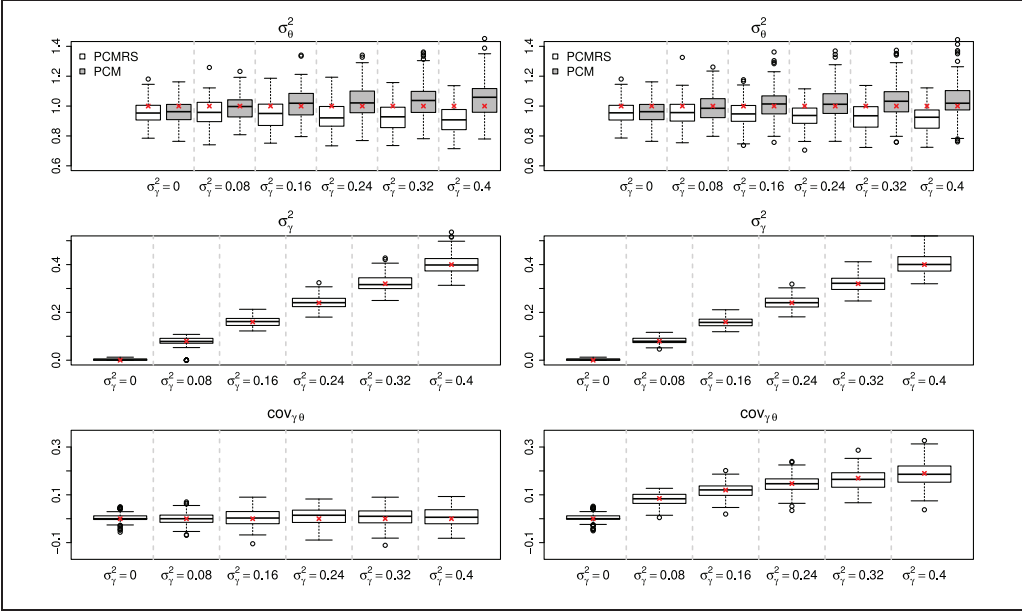
$$P(Y_p | \alpha_p, \hat{\delta}) = \prod_{i=1}^I P(Y_{pi} | \alpha_p, \hat{\delta}).$$

For the computation, numerical integration techniques are needed, for details, see, for example, Fahrmeir and Tutz (1997).

For illustration, Figure 8 shows the mean squared errors (MSEs) summarized over all person parameters for the two simulation settings with ordered item parameters. Again, for each value of  $\sigma_\gamma^2$ , the results for the PCMRs are given on the left (not colored) and the results for the PCM are given on the right (gray colored). The effect on the person parameters is similar to the effect on the item parameters, the MSEs are very similar for very small values of  $\sigma_\gamma^2$ , but for large

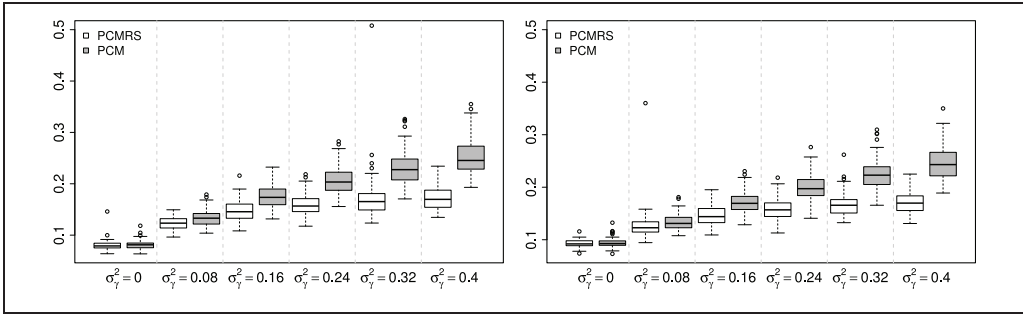


**Figure 6.** Bias of the item parameters of the first item for the simulation with ordered item parameters (left) and nonordered item parameters (right) and  $\rho_{\gamma\theta} = 0$ .  
Note. PCMRs = partial credit model with response style; PCM = Partial Credit Model.



**Figure 7.** Estimates of the variance components for the simulations with ordered item parameters and  $\rho_{\gamma\theta} = 0$  (left) and  $\rho_{\gamma\theta} = 0.3$  (right).  
Note. The true values are marked by red crosses. PCMRs = partial credit model with response style; PCM = partial credit model.





**Figure 8.** MSEs of person parameters for the simulations with ordered item parameters.

Note. The correlation is  $\rho_{\gamma\theta} = 0$  (left) and  $\rho_{\gamma\theta} = .3$  (right). MSE = mean squared errors; PCMRs = partial credit model with response style; PCM = partial credit model.

values of  $\sigma_\gamma^2$ , the MSEs are much larger if the response style is ignored. The picture is very similar in the case of nonordered item parameters (not given).

## Applications

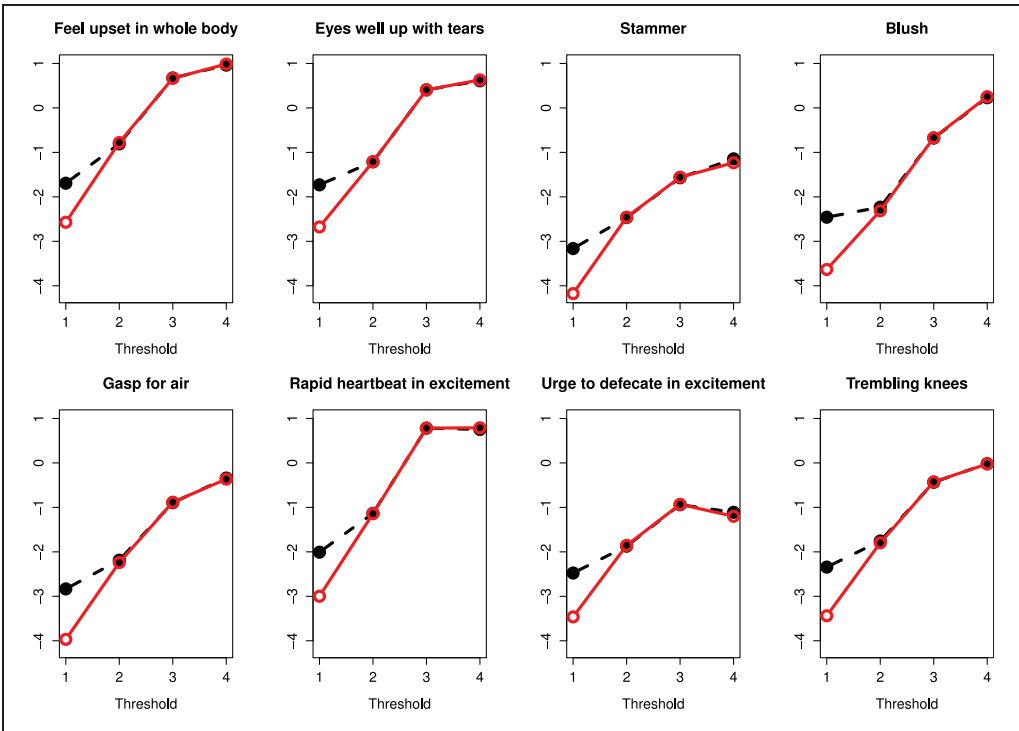
Although the model was introduced for symmetric response categories, a response style as a tendency to middle or extreme categories is often also found for nonsymmetric responses. In the following, we consider response categories that represent the frequency of complaints ranging from *never* to *almost every day*. We use data from the standardization sample of the Freiburg Complaint Checklist (FCC; Fahrenberg, 2010; ZPID, 2013). The FCC is a questionnaire that is used to assess physical complaints of adults. The revised version of the FCC contains 71 items that can measure complaints on nine different scales, as for example the scales *general condition*, *tenseness*, or *emotional reactivity*. The primary data of the standardization sample contains data on 2,070 participants (2,032 complete cases). Each of the 71 items is measured on a 5-point response scale that refers to the frequency of the complaint: *never*, *about 2 times a year*, *about 2 times a month*, *approximately 3 times a week*, or *almost every day*.

The PCMRs and, for comparison, also the simple PCM is applied to the items referring to the scale *emotional reactivity*. When fitting a simple PCM one obtains  $\hat{\sigma}_\theta^2 = 0.785$ , for the extended PCMRs one obtains the covariance matrix:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_\theta^2 & \hat{cov}_{\gamma\theta} \\ \hat{cov}_{\gamma\theta} & \hat{\sigma}_\gamma^2 \end{pmatrix} = \begin{pmatrix} 0.901 & 0.266 \\ 0.266 & 0.562 \end{pmatrix}.$$

The estimated standard deviation of the response style parameters ( $\hat{\sigma}_\gamma = 0.749$ ) is quite high and, therefore, response styles are definitely present. Figure 9 shows the estimates for the item parameters, separately for each item. The solid (red) lines represent estimates for the PCMRs and the dashed (black) lines represent estimates for the PCM. It can be seen that the estimates for the first category differ strongly while the estimates for the other categories mostly coincide. Except for the Item “Urge to defecate in excitement,” all items show ascending category-specific parameters.

To further illustrate the differences between both models due to the presence of the response style parameters in the PCMRs, Figure 10 shows the item response curves for the second item “Eyes well up with tears” for Categories 1, 3, and 5 (rows). In analogy to Figure 3, the curves



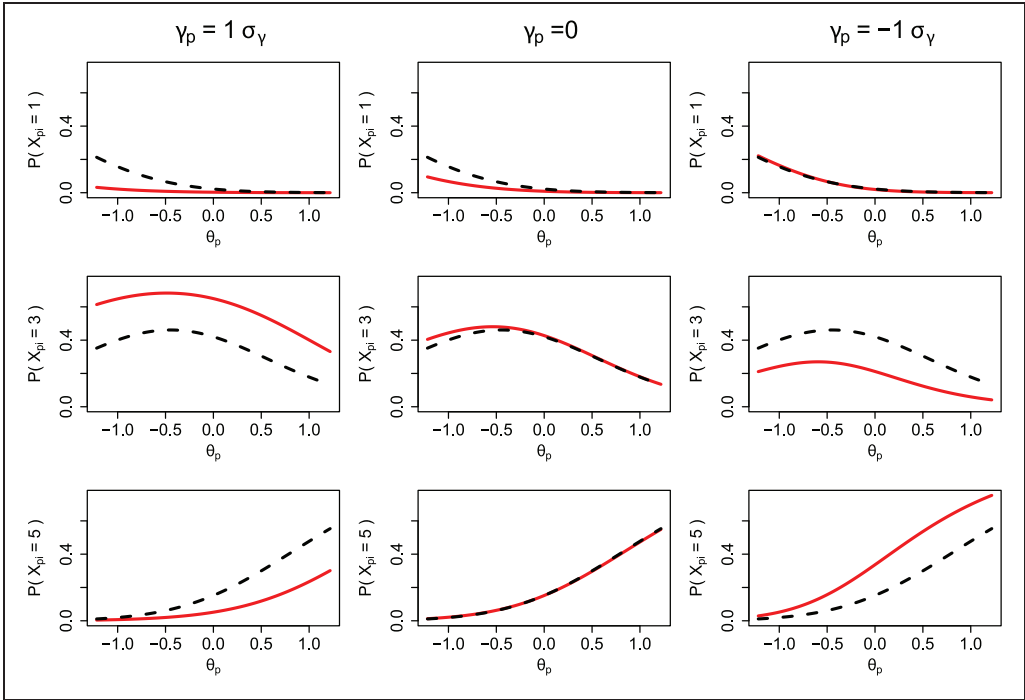
**Figure 9.** Estimates of item parameters for items on emotional reactivity (FCC), separately for each item.

Note. Solid (red) line represents estimates for PCMRs, dashed (black) line represents estimates for PCM. FCC = Freiburg Complaint Checklist; PCMRs = partial credit model with response style; PCM = partial credit model.

are plotted along different values of the person parameters, which are chosen between the 10% and the 90% quantile of the person parameters according to its estimated unidimensional normal distribution  $\hat{\theta} \sim N(0, 0.901)$ . The middle panel represents persons without response style ( $\gamma_p = 0$ ), the left panel represents persons with a tendency to the middle ( $\gamma_p = \sigma_\gamma$ ) and the right panel represents persons with a tendency to the extremes ( $\gamma_p = -\sigma_\gamma$ ). For  $\gamma_p = 0$ , small differences only show up for lower values of the person parameter  $\theta_p$ . However, huge differences appear for  $\gamma_p = -\hat{\sigma}_\gamma$  and  $\gamma_p = \hat{\sigma}_\gamma$ . Although for  $\gamma_p = -\hat{\sigma}_\gamma$  the highest category becomes quite dominant, for  $\gamma_p = \hat{\sigma}_\gamma$  the middle category is the most probable category along the whole range of  $\theta$  values. For more details and an analysis of the items that correspond to the scale *tenseness* see the online supplement.

## Concluding Remarks

A strength of the proposed model is that it disentangles the substantive traits from the response style  $\gamma_p$  by specifying separate parameters  $\theta_p$  and  $\gamma_p$ . The parameterization is simple with a straightforward interpretation as shifting of thresholds. Moreover, the parameterization uses one person parameter to cover the continuum between the preference of middle categories and extreme categories. As it is not assumed that the response style and the substantive trait are independent, one can investigate the association between these latent traits.



**Figure 10.** Response curves for Item “Eyes well up with tears” (FCC) along person parameter  $\theta$  (between 10% and 90% quantile) for Category 1 (upper panel), Category 3 (middle panel), and Category 5 (lower panel).

Note. Columns represent different response styles: tendency to the middle (left), no response style (middle), and tendency to the extremes (right). Solid (red) lines represent estimates for PCMRs, dashed (black) lines represent estimates for PCM. FCC = Freiburg Complaint Checklist; PCMRs = partial credit model with response style; PCM = partial credit model.

The specific response style captured by the model, namely, the tendency to middle or extreme categories, is itself seen as a latent trait, which becomes obvious from the embedding into the framework of multidimensional IRT (“Embedding into the Framework of Multidimensional Item Response” section). Response style is modeled on the respondent level in contrast to response style on the stimulus level. It is seen as a personal characteristic. As such it can be linked to other latent traits, as, for example, intolerance of ambiguity, decisiveness or extraversion, and conscientiousness (see Austin, Deary, & Egan, 2006; Naemi, Beal, & Payne, 2009). Because of the relationship between personality and response style, it is important to diagnose and correct for response style. For more literature, including the dependence of response styles on external variables (see Van Vaerenbergh & Thomas, 2013).

The authors explicitly considered a partial credit model that accounts for a specific response style. However, the basic concept can also be used to model response styles in alternative ordinal latent trait models like the graded response model (Samejima, 1997). The only complication in graded response models is that it is based on a threshold concept with thresholds that have to be ordered. The ordering has to hold also after incorporating response style parameters. The model could also be extended by considering alternative functions to spread the thresholds. For

example, for even  $k$  one could use  $(m-r)^p$ , which includes quadratic spreading if  $p=2$ . Considering  $p$  as a tuning parameter one could optimize the fit.

The proposed method has been implemented in R (R Core Team, 2016), it is available on request from the authors and will be available from CRAN (Comprehensive R Archive Network) soon. For the joint normal distribution of the person parameters and the response style parameters, two-dimensional Gauss-Hermite integration is used. For faster performance, it is (in a parallel manner) implemented in C++ and integrated into R using the package Rcpp (Eddelbuettel, 2013). Optimization of the marginal likelihood is done numerically using the algorithm L-BFGS-B (see Byrd, Lu, Nocedal, & Zhu, 1995).

## 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: The authors gratefully acknowledge funding from Deutsche Forschungsgemeinschaft (DFG project TU 62/8-1).

## Supplemental Material

Supplementary material is available for this article online

## References

- Anderson, D. A., & Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 47, 203-210.
- Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, 40, 1235-1245.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, 17, 665-678.
- Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, 22, 69-83.
- Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology*, 70, 159-181.
- Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, 71, 814-833.
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association*, 88(421), 9-25.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16, 1190-1208.
- De Boeck, P., & Partchev, I. (2012). Irtrees: Tree-based item response models of the glmm family. *Journal of Statistical Software*, 48, 1-28.
- Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. New York, NY: Springer-Verlag.

- Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment, 16*, 20-30.
- Fahrenberg, J. (2010). *Freiburg Complaint Checklist* [Freiburger Beschwerdenliste (FBL)]. Göttingen, Germany: Hogrefe.
- Fahrmeir, L., & Tutz, G. (1997). *Multivariate statistical modelling based on generalized linear models*. New York, NY: Springer-Verlag.
- Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods, 21*, 328-347.
- Gollwitzer, M., Eid, M., & Jürgensen, R. (2005). Response styles in the assessment of anger expression. *Psychological Assessment, 17*, 56-69.
- Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist (Ed.), *International Conference on Generalized Linear Models* (pp. 109-121). New York, NY: Springer-Verlag.
- Irtel, H. (1995). An extension of the concept of specific objectivity. *Psychometrika, 60*, 115-118.
- Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods, 48*, 1070-1085.
- Jin, K.-Y., & Wang, W.-C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement, 74*, 116-138.
- Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika, 68*, 563-583.
- Johnson, T. R. (2014). A note on conditional and quasi-conditional maximum likelihood estimation for polytomous item response models: Robust inference in the presence of individual differences in response scale use. *Applied Psychological Measurement, 38*, 599-613.
- Johnson, T. R., & Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics, 35*, 92-114.
- Kankaraš, M., & Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe insights from a multiple-group latent-class factor approach. *International Sociology, 24*, 557-579.
- Kelderman, H. (1996). Multidimensional rasch models for partial-credit scoring. *Applied Psychological Measurement, 20*, 155-168.
- Khorrandel, L., & von Davier, M. (2014). Measuring response styles across the big five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research, 49*, 161-177.
- Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement, 32*, 611-631.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.
- Masters, G. N., & Wright, B. (1984). The essential process in a family of measurement models. *Psychometrika, 49*, 529-544.
- McCulloch, C., & Searle, S. (2001). *Generalized, linear, and mixed models*. New York, NY: John Wiley.
- Messick, S. (1991). Psychology and methodology of response styles. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 161-200). Hillsdale, NJ: Lawrence Erlbaum.
- Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review, 20*, 303-320.
- Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research, 22*, 93-119.
- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality, 77*, 261-286.
- Plieninger, H. (2016). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement, 77*, 32-53. doi:10.1177/0013164416636655
- Plieninger, H., & Meiser, T. (2014). Validity of multiprocess IRT models for separating content and response styles. *Educational and Psychological Measurement, 74*, 875-899.

- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57.
- Rasch, G. (1977). On specific objectivity. an attempt at formalizing the request for generality and validity of scientific statements in symposium on scientific objectivity. *Danish Year-Book of Philosophy Kobenhavn*, 14, 58-94.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology*, 44, 75-92.
- Samejima, F. (1997). Graded response model. In W. Van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York, NY: Springer.
- Suh, Y., & Bolt, D. M. (2010). Nested logit models for multiple-choice item response data. *Psychometrika*, 75, 454-473.
- Thissen-Roe, A., & Thissen, D. (2013). A two-decision model for responses to likert-type items. *Journal of Educational and Behavioral Statistics*, 38, 522-547.
- Tutz, G. (2012). *Regression for categorical data*. Cambridge, UK: Cambridge University Press.
- Tutz, G., & Berger, M. (2016). Response styles in rating scales: Simultaneous modeling of content-related effects and the tendency to middle or extreme categories. *Journal of Educational and Behavioral Statistics*, 41, 239-268.
- Van Rosmalen, J., Van Herk, H., & Groenen, P. (2010). Identifying response styles: A latentclass bilinear multinomial logit model. *Journal of Marketing Research*, 47, 157-172.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25, 195-217.
- von Davier, M., & Rost, J. (1995). Polytomous mixed rasch models. In G. Fischer & I. Molenaar (Eds.), *Rasch models* (pp. 371-379). New York, NY: Springer-Verlag.
- von Davier, M., & Yamamoto, K. (2007). Mixture-distribution and hybrid rasch models. In M. von Davier & C. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 99-115). New York, NY: Springer-Verlag.
- Wang, W.-C, Wilson, M., & Shih, C.-L. (2006). Modeling randomness in judging rating scales with a random-effects rating scale model. *Journal of Educational Measurement*, 43, 335-353.
- Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, 33, 352-364. doi:10.1027/1015-5759/a000291
- Yee, T. W. (2014). *VGAM: Vector generalized linear and additive models* (R package version 0.9-4). Retrieved from <https://www.stat.auckland.ac.nz/~yee/VGAM>
- ZPID. (2013). *PsychData of the Leibniz Institute for Psychology Information ZPID*. Trier, Germany: Center for Research Data in Psychology.