Short Methodological Report

Developing and Applying IR-Tree Models: Guidelines, Caveats, and an Extension to Multiple Groups

⑤SAGE

Hansjörg Plieninger[1] ⓘ

Abstract
IR-tree models assume that categorical item responses can best be explained by multiple response processes. In the present article, guidelines are provided for the development and interpretation of IR-tree models. In more detail, the relationship between a tree diagram, the model equations, and the analysis on the basis of pseudo-items is described. Moreover, it is shown that IR-tree models do not allow conclusions about the sequential order of the processes, and that mistakes in the model specification can have serious consequences. Furthermore, multiple-group IR-tree models are presented as a novel extension of IR-tree models to data from heterogeneous units. This allows, for example, to investigate differences across countries or organizations with respect to core parameters of the IR-tree model. Finally, an empirical example on organizational commitment and response styles is presented.

Keywords
item response theory (IRT), IR-tree model, response style, organizational commitment, multiple-group model

Item response tree (IR-tree) models are a class of item response theory (IRT) models that assume that the responses to polytomous items can best be explained by multiple psychological processes (Böckenholt, 2012; Tutz, 1990). Thus, IR-tree models go beyond standard IRT models like the graded response model (GRM) that usually comprise only one process or dimension. Importantly, IR-tree models may contain processes that are not related to the degree of agreement. Therefore, they are well suited to address response styles such as the tendency toward extreme categories (extreme response style, ERS) or toward the midpoint of a scale (midpoint response style, MRS; see also Baumgartner & Steenkamp, 2001).

[1]School of Social Sciences, Department of Psychology, University of Mannheim, Mannheim, Germany

Corresponding Author:
Hansjörg Plieninger, Department of Psychology, University of Mannheim, 68131 Mannheim, Germany.
Email:plieninger@uni-mannheim.de

Even though the idea of modeling multiple processes in categorical data has existed for years (e.g., Maris, 1995; McFadden, 1981; Tutz, 1990), these models didn't gain much interest in psychology until the seminal publications of Böckenholt (2012) and De Boeck and Partchev (2012). Since then, IR-tree models have attained increasing popularity in psychometrics (e.g., Meiser, Plieninger, & Henninger, 2019; Plieninger & Heck, 2018; Tijmstra, Bolsinova, & Jeon, 2018) as well as in applied fields (e.g., DiTrapani, Jeon, De Boeck, & Partchev, 2016; Lang, Lievens, De Fruyt, Zettler, & Tackett, 2019; Zettler, Lang, Hülsheger, & Hilbig, 2016).

The goal of the present article is twofold. The first goal is to make IR-tree models easily accessible to scholars in organizational psychology and beyond. To this end, guidelines for IR-tree models are presented that elaborate on and highlight important points of the relevant psychometric literature. These will be presented in the first section, followed by a section on caveats and pitfalls to illustrate what can go wrong if one does not adhere to these guidelines. The second goal is to present a new extension of IR-tree models to multiple groups that is applicable to data that are collected from potentially heterogeneous units such as different countries or organizations. The last section puts all this into illustrative practice by means of an application to empirical data on organizational commitment. In summary, the present work will help scholars to review published work on IR-tree models, to apply existing IR-tree models to their own single- or multigroup data, and to develop new IR-tree models for research questions specific to their field. The examples used herein will, without loss of generality, focus on response style modeling, but many other applications are possible as well (see below).
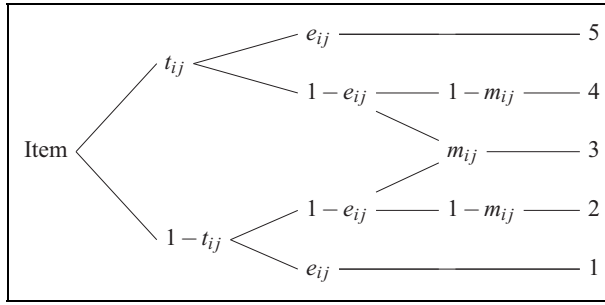
IR-Tree Modeling

As of today, most applications of IR-tree models focus on response styles, and researchers in organizational psychology may make similar use of IR-tree models if they want to control for such response tendencies. Examples of such applications are the study by LaHuis, Blackmore, Bryant-Lees, and Delgado (2019), who investigated factors predicting job performance, or the analysis by Böckenholt (2012), who studied consumer ethics. However, IR-tree models may also be used outside the area of response styles, and examples relevant to organizational psychology include modeling the compromise effect in economic choices (Böckenholt, 2012), and modeling answer changes (Jeon, De Boeck, & van der Linden, 2017) or missing answers (Debeer, Janssen, & De Boeck, 2017; Jeon & De Boeck, 2016). Additional examples for future applications are listed in the section Further IR-Tree Models.

Thus, IR-tree models can be applied in situations where discrete data are suspect to individual differences in more than one latent variable or process. These models may then be used instead of a unidimensional model, and guidelines for IR-tree modeling will be presented in the following. More specifically, the three building blocks of IR-tree models will be presented illustrating how a *tree diagram* leads to a set of *model equations* that can be used to define *pseudo-items* for estimation.

Tree Diagram

IR-tree models have the advantage that tree diagrams can help to both illustrate and develop a model. Such tree diagrams are also used for multinomial processing tree (MPT) models in cognitive psychology (e.g., Riefer & Batchelder, 1988), and IR-tree models are a special case of hierarchical MPT models (Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Plieninger & Heck, 2018). Herein, I will focus on an IR-tree model for items with five response categories k ($k = 1, \ldots, 5$) proposed by LaHuis et al. (2019). It aims to disentangle three different processes ξ_p ($p = 1, \ldots, P$), namely the target trait (e.g., commitment) and two response styles, ERS and MRS. The diagram in

Figure 1. Diagram of the MCN model, an IR-tree model for 5-point items composed of the processes *t* (target trait), *e* (extreme response style), and *m* (midpoint response style).

Figure 1 illustrates that process $\xi_1 = t$ (target trait) leads to agreement, process $\xi_2 = e$ (ERS) leads to extreme categories, and process $\xi_3 = m$ (MRS) leads to the middle category. Because midpoint responses are modeled conditional on nonextreme responses, the model will be called the *midpoint conditionally on nonextreme* (MCN) model.

The model parameters ξ are in fact probabilities such that, for instance, extreme categories are chosen with probability e_{ij} and nonextreme categories with probability $1 - e_{ij}$. These model parameters are specific to person i and item j ($j = 1, \ldots, J$). In more detail, each parameter ξ is reparameterized using a probit-link IRT model with a person ability parameter θ and an item difficulty parameter β (and sometimes also an item discrimination parameter α):

$$\xi_{pij} = \Phi(\theta_{pi} - \beta_{pj}), \tag{1}$$

where Φ is the cumulative normal distribution function. For the MCN model, this leads to

$$t_{ij} = \Phi(\theta_{ti} - \beta_{tj}) \tag{2}$$

$$e_{ij} = \Phi(\theta_{ei} - \beta_{ej}) \tag{3}$$

$$m_{ij} = \Phi(\theta_{mi} - \beta_{mj}). \tag{4}$$

Thus, the probability to agree (i.e., t_{ij}) is higher the higher the person ability θ_{ti} and the lower the item difficulty β_{tj}, and analogous relationships hold for e_{ij} and m_{ij}.

In summary, the model assumes that responses to 5-point items can be explained by the target trait as well as the two response styles MRS and ERS. Often, an IR-tree model fits better than a unidimensional alternative, and this is taken as an indication that response styles are present in the data at hand (e.g., Böckenholt, 2012; Khorramdel & von Davier, 2014; Plieninger & Meiser, 2014). The IR-tree model then allows to measure response styles as well as the response-style free target trait.

Model Equations

The model is formally defined by its model equation that is implied by the tree diagram (Böckenholt, 2012; Plieninger & Heck, 2018; Riefer & Batchelder, 1988). Let X_{ij} be a random variable representing the response of person i to item j, and let x_{ij} be a realization of X_{ij}. Then, the probability Pr for a branch b of the model is given by multiplying all parameters ξ along a branch in the diagram:

$$\Pr(b_{ij}) = \prod_{p=1}^{P} \xi_{pij}{}^{v_{bp}} (1 - \xi_{pij})^{w_{bp}}, \tag{5}$$

where v_{bp} and w_{bp} count how often the parameters ξ_p and $(1 - \xi_p)$ occur in branch b (e.g., if ξ_p occurs 0 times, then $\xi_p{}^0 = 1$). For instance, the probability for the upper branch that leads to Category 3 in Figure 1 is given by multiplying the three terms along this branch, namely, $t_{ij}(1 - e_{ij})m_{ij}$.

If multiple branches b lead to a category k, the respective probabilities of the corresponding branches $b = 1, \ldots, B_k$ are added:

$$\Pr(X_{ij} = x_{ij}) = \sum_{b=1}^{B_k} \Pr(b_{ij}). \tag{6}$$

For the MCN model, this general formulation implies the five equations given below. For example, the probability for Category 3 is given by adding the probabilities of the two respective branches shown in Figure 1.

$$\Pr(X_{ij} = 5) = t_{ij}e_{ij} \tag{7}$$

$$\Pr(X_{ij} = 4) = t_{ij}(1 - e_{ij})(1 - m_{ij}) \tag{8}$$

$$\Pr(X_{ij} = 3) = t_{ij}(1 - e_{ij})m_{ij} + (1 - t_{ij})(1 - e_{ij})m_{ij}$$

$$= (t_{ij} + 1 - t_{ij})(1 - e_{ij})m_{ij} = (1 - e_{ij})m_{ij} \tag{9}$$

$$\Pr(X_{ij} = 2) = (1 - t_{ij})(1 - e_{ij})(1 - m_{ij}) \tag{10}$$

$$\Pr(X_{ij} = 1) = (1 - t_{ij})e_{ij}. \tag{11}$$

These model equations, combined with the reparameterizations specified in Equations 2–4, define the multinomial distribution. They illustrate the model in a condense form and guide the interpretation. Here, for example, Equation 9 shows that the probability to choose Category 3 is governed only by the parameters e and m. This is evident from Figure 1 only on closer inspection: The upper branch leading to Category 3 contains the term t, and the lower branch contains the term $1 - t$. However, they cancel each other out in Equation 9 such that t does not influence $\Pr(X = 3)$. Note further that the individual equations of a multinomial model sum up to 1. This is the case for the MCN model: The probability for an extreme response (i.e., $X = 1$ or $X = 5$) is $te + (1 - t)e = e$; the probability for either $X = 2$ or $X = 4$ simplifies to $(1 - e)(1 - m)$; adding this to the probability for $X = 3$ gives $(1 - e)(1 - m) + (1 - e)m = 1 - e$; finally, the sum of this and the probability for an extreme response equals $\Pr(X \geq 1) = 1 - e + e = 1$.

In summary, both the tree diagram and the model equations are helpful means to develop and interpret a model. While the diagram is more illustrative, the equations show more directly the relationship between each category and the parameters involved. Moreover, the model equations define the pseudo-items that are useful for data analysis.

Pseudo-Items

A convenient feature of IR-tree models is that they can be estimated using software such as Mplus or R (e.g., Böckenholt, 2012; De Boeck & Partchev, 2012). To this end, the original data have to be

Table 1. Pseudo-items for the MCN Model.

Item Response	Equation	Pseudo-Item Response		
x	$\Pr(X = x)$	x_t^*	x_e^*	x_m^*
5	$t\,e$	1	1	—
4	$t\,(1-e)\,(1-m)$	1	0	0
3	$(1-e)\,m$	—	0	1
2	$(1-t)\,(1-e)(1-m)$	0	0	0
1	$(1-t)\,e$	0	1	—

recoded into so-called binary *pseudo-responses* (or *pseudo-items*) according to the following set of rules:

1. The original response x is recoded into P pseudo-responses x_p^*, one for each model parameter ξ_p.
2. If the model equation pertaining to response x contains parameter ξ_p, the pseudo-response x_p^* is coded as 1.
3. If the model equation pertaining to response x contains parameter $1 - \xi_p$, the pseudo-response x_p^* is coded as 0.
4. If the model equation pertaining to response x does neither contain ξ_p nor $1 - \xi_p$, the pseudo-response x_p^* is coded as *Missing*.

This is illustrated for a strongly disagree response (i.e., $x = 1$) in the MCN model in the following: Recall that Equation 11 was defined as $\Pr(X_{ij} = 1) = (1 - t_{ij})e_{ij}$. First, the MCN model is comprised of the three parameters t, e, and m. Thus, each response (or item) is recoded into three pseudo-responses (or pseudo-items). Second, the model equation contains the term e, and thus Rule 2 leads to $x_e^* = 1$. Third, the equation contains the term $1 - t$, and thus Rule 3 leads to $x_t^* = 0$. Fourth, parameter m is not present in the equation, and thus Rule 4 leads to a missing value for x_m^*. This is also illustrated in the last row of Table 1, where the coding schemes for the other response categories are shown as well.

In practice, the recoding procedure is applied to all responses x_{ij}. Thus, the new data set contains $J \times P$ binary pseudo-items instead of J polytomous items. Then, a three-dimensional, binary IRT model can be fit with three correlated latent variables θ_p, one for each group of pseudo-items x_p^*. Apart from that, note that estimation on the basis of pseudo-items works only for classical IR-tree models that do not contain mixtures, that is, for models where the simplified model equations (e.g., Equations 7–11) contain only products of ξ-parameters but not sums (Jeon & De Boeck, 2016; Plieninger & Heck, 2018).

Further IR-Tree Models

A further, very popular IR-tree model for 5-point items was proposed by Böckenholt (2012). It is shown in Figure 2 and will be called the *extremity conditionally on nonmidpoint* (ECN) model. The model is composed of the processes t (target trait), e (extreme response style), and m (midpoint response style), and the tree diagram implies the model equations depicted in Figure 2. These equations lead to the following pseudo-items: $(0, 0, 1, 0, 0)$ for x_m^*, $(0, 0, -, 1, 1)$ for x_t^*, and $(1, 0, -, 0, 1)$ for x_e^*. Thus, the ECN model and the MCN model (see Figure 1) have in common the aim to disentangle the target trait and two response styles. However the
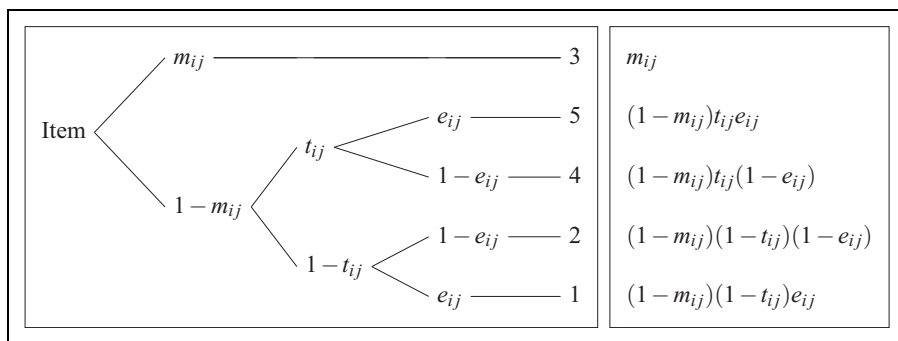
Figure 2. The ECN model: The tree diagram is depicted on the left, the model equations on the right.
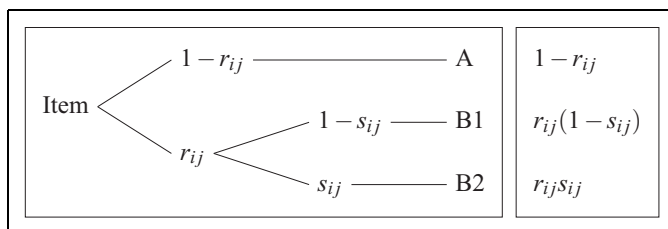
Figure 3. An IR-tree model for items with three outcomes, A, B1, and B2: The tree diagram is depicted on the left, the model equations on the right.

structure of the model and the interpretation of its parameters differ slightly. For example, the process m affects all five categories in the ECN model but only the inner three categories in the MCN model.

Apart from that, IR-tree models can be suitable for categorical data outside the area of response style research. Consider the model in Figure 3 that comprises two processes (see also De Boeck & Partchev, 2012; Tutz, 1990). Such a model may be appropriate in cases where one process is only applicable for "positive" outcomes of the other process. For example, process r may correspond to making a choice or not, and process s may differentiate between two different choices B1 and B2. Or, process r may encode whether someone was absent from work, and process s may differentiate between two different reasons for absenteeism. Or, in a situational judgment test (SJT; see also Lievens et al., 2018), process r may encode whether a chosen option was "correct," and process s may differentiate between different types of correct behavior.

Caveats and Pitfalls of IR-Tree Modeling

The guidelines above describe the steps that are necessary to develop, apply, and interpret an IR-tree model. However, things can go wrong in this process, and two issues that need special attention will be discussed in the following. These caveats together with the guidelines above will help scholars to leverage the full potential of IR-tree models.

Order of Psychological Processes

The first issue concerns the order of the latent processes. Tree diagrams such as the one depicted in Figure 1 are drawn with a specific order of the parameters ξ. This might lead to the belief that the model implies a certain sequential order of the psychological processes.
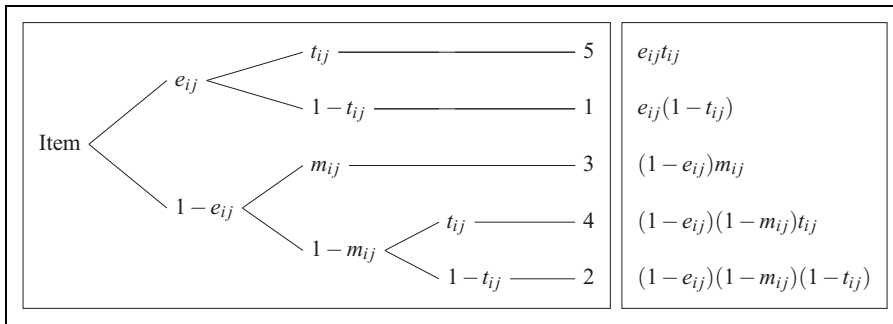
Figure 4. Alternative diagram of the MCN model shown in the box on the left. Even though the diagram is different from that in Figure 1, the implied model equations shown in the box on the right are identical.

However, this is not the case. The model equations contain products (and potentially sums) of parameters, and these operations are commutative; for example, $\Pr(x = 5) = te = et$. In other words, it is irrelevant for the model equations in which order the processes occur on a path in a tree diagram, it is only relevant which processes occur. Thus, IR-tree models make assumptions about the number and type of processes involved but not about their order, and such interpretations can be misleading.

Additionally, a set of model equations such as Equations 7–11 can sometimes be illustrated using different diagrams. For example, the diagram depicted in Figure 4 also represents an IR-tree model. Therein, respondents choose with probability e_{ij} an extreme response and with probability $1 - e_{ij}$ a nonextreme response. Conditionally on an extreme response, respondents can either agree or disagree, and so on. If the order in a tree diagram was meaningful in itself, the models in Figures 1 and 4 would make completely different assumptions. However, when deriving the model equations from the diagram as shown on the right in Figure 4, it becomes clear that they match Equations 7–11. Thus, the two models shown in Figures 1 and 4 are equivalent. This illustrates that both the model equations and the tree diagram are helpful means that guide the interpretation. Whether and how two models differ is most evident from the model equations. Two different diagrams may indeed imply different models (e.g., the MCN depicted in Figure 1 and the ECN model depicted in Figure 2), but they sometimes may also imply equivalent models as shown herein.

Apart from that, it should be noted that a psychological theory about the order of the processes may be expressed in a tree diagram. The only thing to keep in mind is that it does not work the other way round, that is, the IR-tree model does not provide a formal test of the order. Furthermore, one may argue that some probabilities in an IR-tree model can be expressed as conditional probabilities thus implying some sort of order. For example, in the MCN model (see Figures 1 and 4), the process m does only occur in combination with nonextremity $1 - e$. However, such conditional probabilities should not be overinterpreted given the nature of the data in typical applications of IR-tree models: It is probably safe to say that these psychological processes do rather occur in parallel, mutually reinforcing, and/or heterogeneously across persons, items, and situations and not in a fixed and invariant order.

In summary, IR-tree models allow interpretations and make assumptions about the nature of the psychological processes that are involved, but not about their sequential order. Such interpretations are misleading and must be based on other research designs and statistical models if desired.

Definition of Pseudo-Items

The second issue concerns parameter estimation on the basis of pseudo-items. As explained above, an IR-tree model may be estimated in the form of a multidimensional, binary IRT model after the original items have been recoded into binary pseudo-items. This might lead to the misconception that the pseudo-items can be defined independent of each other. For example, a researcher may want to define the third pseudo-item x_m^* as $(0, 0, 1, 0, 0)$ instead of $(-, 0, 1, 0, -)$ in the MCN model (see Table 1) in order to reduce the number of missing values. Or, one might want to define the second pseudo-item x_e^* as $(1, 0, -, 0, 1)$ instead of $(1, 0, 0, 0, 1)$ for some reason. However, this is not legitimate within the IR-tree framework: This leads to a model (as implied by the pseudo-items) that no longer corresponds to the specified model equations and tree diagram. Worse still, such changes can easily lead to an improper model in the sense that the model equations do not sum to 1 and that no tree diagram exists that could illustrate its structure. Figuratively speaking, such an improper model may make predictions such as 50% heads and 60% tails. Strictly speaking, the set of equations implied by the pseudo-items do not even compose a probabilistic model because the definition of a discrete probability distribution requires that the sum of the individual probabilities is 1.

Note that the estimation of an improper model may nevertheless converge without problems, because the software "doesn't know" that an IR-tree model for polytomous data is specified rather than a multidimensional IRT model with three seemingly independent, binary items. Furthermore, it can be shown that improper models lead to biased parameter estimates and severely biased model fit.

Multiple-Group IR-Tree Models

The guidelines and caveats presented above will help to successfully use standard IR-tree models. However, specific research questions sometimes need specific model extensions.

Although many such extensions have already been presented in the literature, all these models are applicable only to data from homogeneous populations—where relevant parameter values hold for all members of that population. However, certain circumstances may lead to violations of this assumptions. For example, levels of the target trait may differ across industries or across measurement occasions. Similarly, countries may differ with respect to variability in extreme response style. Likewise, a researcher may want to investigate whether the item parameters are identical for women and men or across age groups. Ignoring existing group differences may result in misleading conclusions. One way to address such situations are multiple-group models, and these will be presented in the following. Therein, it is assumed that the structure of the tree diagram is the same for all groups, but certain parameters may differ.

Standard IRT models have been extended to multiple-group IRT models, for example, by Bock and Zimowski (1997). Likewise, IR-tree models can be extended to multiple-group IR-tree models by allowing the model parameters to differ between different groups g ($g = 1, \ldots, G$). Thus, adding subscript g and combining, for convenience, Equations 1, 5, and 6 gives:

$$\Pr(X_{gij} = x_{gij}) = \sum_{b=1}^{B_k} \prod_{p=1}^{P} (\Phi(\theta_{pgi} - \beta_{pgj}))^{v_{bp}} \left(1 - \Phi(\theta_{pgi} - \beta_{pgj})\right)^{w_{bp}}. \tag{12}$$

That is, given the tree diagram, which is the same for all groups, the probability of a response of person i in group g to item j depends on his or her latent ability θ and on the item difficulty β, and these item parameter may differ across groups.

Furthermore, in the single-group model ($G = 1$), it is assumed that $\boldsymbol{\theta}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. Thus, the model is identified by centering each latent variable θ_p around 0, and the (co)variances of the latent variables contained in the matrix $\boldsymbol{\Sigma}$ are freely estimated. In the multiple-group case, $\boldsymbol{\Sigma}$ may differ across groups leading to $\boldsymbol{\theta}_{gi} \sim MVN(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. However, the latent means $\boldsymbol{\mu}_g$ need to be constrained for identification purposes, for example, by assuming $\boldsymbol{\mu}_g = \mathbf{0}_g$.

In order to make meaningful comparisons of persons across groups, invariance/equivalence of the item parameters has to be assumed or established. In the IRT literature, this is referred to as the absence of differential item functioning (DIF), while the term strong/scalar invariance is used in factor analysis (for details, see Cheung & Rensvold, 1999; Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993; Tay, Meade, & Cao, 2015; Vandenberg & Lance, 2000). In detail, assuming that the item parameters do not differ across groups reduces β_{pgj} to β_{pj}. This then allows to relax the constraint of $\boldsymbol{\mu}_g = \mathbf{0}_g$ and to freely estimate $\boldsymbol{\mu}_g$ in $G - 1$ groups while constraining only one group, usually the first one to $\boldsymbol{\mu}_{g=1} = \mathbf{0}$. Stated differently, the model is identified by centering all dimensions around 0 in the first group; paired with invariant item parameters, this allows to compare all other groups to the first group and to each other with respect to mean levels of the latent variables. Note further that invariance or the absence of DIF are concepts that are usually associated with the target trait. Extending these concepts to response styles is straightforward and necessary, but it should be kept in mind that invariance of response styles and invariance of target traits may have different interpretations substantively.

Conveniently, multiple-group IR-tree models can be estimated with software for multidimensional, multiple-group IRT models such as Mplus (Muthén & Muthén, 2012) as long as the model does not contain genuine mixtures (see above). An illustrative example will be presented in the next section, and the respective Mplus code is contained in the accompanying research compendium.

In summary, multiple-group IR-tree models are relevant to situations where data were collected in different countries, industries, organizations, or at different points in time, for example. The model class is a straightforward extension of IR-tree models, and thus the guidelines and caveats described above apply to single- and multiple-group models in the same way. Furthermore, multiple-group IR-tree models also build on standard multiple-group models (e.g., Reise et al., 1993; Tay et al., 2015), and thus topics such as DIF or measurement invariance apply to multiple-group IR-tree models as well. Note further that multiple-group models pertain to situations where the groups are known, but models for unknown groups (latent classes) exist as well (e.g., Tijmstra et al., 2018).

Worked Example: Response Styles and Workers' Commitment Across Countries

In the following, an illustrative example on organizational commitment in different countries will be presented to put the first three sections of the article into practice. In detail, the guidelines for single-group models will be applied, and multiple-group models for different countries will be used to illustrate this new model class.

Method

Sample. Data from the International Social Survey Programme were used, namely, from ISSP 2015—Work Orientations IV (ISSP Research Group, 2017). The data set is composed of respondents from 37 countries, and data from Venezuela (see below) were selected, as were data from ten other countries chosen at random (selected countries were Australia, China, Croatia, Hungary, Latvia, Lithuania, Mexico, New Zealand, Slovenia, and United Kingdom). A random subsample

of 500 participants (with complete cases) per country was used herein. Three countries had fewer respondents, and all participants were used in such cases. The final data set comprised 5,311 participants in total.

Measures. Respondents answered three questions about their organizational commitment, for example, "I am proud to be working for my firm or organization," on a 5-point scale from *strongly agree* to *strongly disagree*. The correlations among the three items across all participants where $r_{12} = .56$, $r_{13} = .34$, and $r_{23} = .42$. The response distributions of the three items across countries are shown in the appendix.

Results

First, the application of IR-tree models will be illustrated using data from one specific country, namely, Venezuela. Second, multiple-group models will be used to investigate cross-cultural patterns across all 11 countries. All models were estimated using Mplus 7.4 (Muthén & Muthén, 2012).

IR-Tree Model for Data From Venezuela. The data from Venezuela ($n = 382$) were selected for illustrative purposes because the response distributions of the three items indicated that extreme categories were selected with high prevalence in this country (see the appendix). The MCN model discussed above (see Figure 1) was fit to the data resulting in descriptive fit indices of AIC = 2,590 and BIC = 2,650. This model was compared to a GRM, that is, a standard unidimensional IRT model that does not take response style into account. The fit indices of the GRM were larger signaling inferior fit with AIC = 2,690 and BIC = 2,740. Thus, it may be concluded that a multi-dimensional model that accounts for response styles explained the data better than a standard unidimensional model.

The parameter estimates showed the following pattern: Three β_e parameters were estimated, namely, one for each of the three questionnaire items, and the mean of the three β_e parameters equaled -0.95. This can be plugged into Equation 3 in combination with $\theta_e = 0$ for a person with an average ERS level leading to $\Phi(0 - (-0.95)) = .83$. Thus, the average probability to give an extreme response equaled 83%. This mirrors the large proportion of extreme responses that was observed (see also Figure A1). Likewise, the mean of the three β_m parameters equaled 1.27, and thus the probability to give a midpoint response equaled 10% on average ($\Phi(0 - 1.27) = .10$). Note further that this probability m_{ij} is conditional on a nonextreme response, as can be inferred from Table 1. Furthermore, these probabilities refer to Equations 2–4, that is, to the probabilities of the three model parameters t, e, and m. These probabilities can, of course, also be used to calculate the category probabilities according to Equations 7–11. In detail, averaging across all three items and using an average person with a theta vector equal to 0 leads to $e = .83$, $m = .10$ (see above), and $t = \Phi(0 - (-0.84)) = .80$. Plugging these three values into Equations 7–11 leads to probabilities of .17, .03, .02, .12, and .66 for Categories 1 to 5. These sum to 1, and they closely resemble to observed frequencies (see Figure A1).

With respect to the person parameters, standard deviations of $\sigma_{\theta_t} = 0.66$, $\sigma_{\theta_e} = 1.37$, and $\sigma_{\theta_m} = 0.63$ were observed. Thus, there was considerable variability between persons in, for example, their tendency to use extreme categories. For example, persons with an ERS level one standard deviation above and below the mean are expected to have values of $\theta_e = \pm 1.37$. Again, these values can be plugged into Equation 3 resulting in $\Phi(1.37 + 0.95) = .99$ and $\Phi(-1.37 + 0.95) = .34$, respectively. Thus, the probability to give an extreme response to an average item ranged from 34% for persons relatively low on ERS to 99% for persons relatively high on ERS.
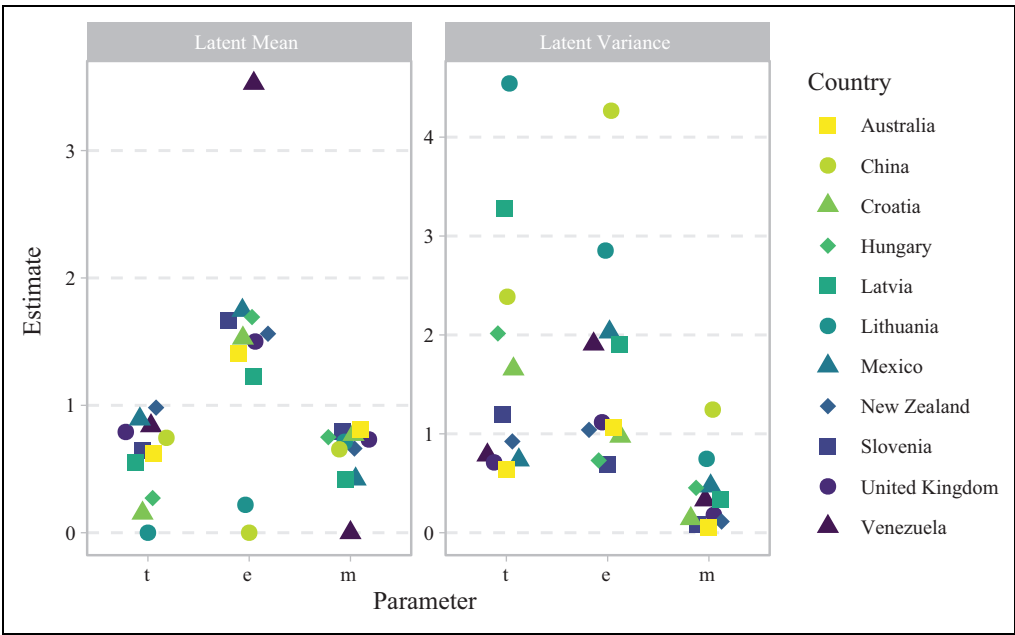
Figure 5. Country-level parameters as estimated from an IR-tree model, namely (relative) latent means and latent variances for the three parameters target trait *t*, extreme response style *e*, and midpoint response style.

Furthermore, the correlations among the dimensions were estimated as $r_{te} = .45, r_{tm} = -.49$, and $r_{em} = -.33$. The negative correlation between ERS and MRS is a typical finding in response style research. The substantial correlations between the target trait and response styles are particularly noteworthy, because such correlations have been found to cause bias (e.g., Plieninger, 2017). In order to investigate potential bias, the θ_t parameters from the MCN model and the GRM were compared. Their correlation equaled .94, and the mean of the absolute differences was 0.21. Thus, differences between the models were mirrored in different estimates of the persons' level of organizational commitment.

In summary, the analyses suggested that response styles were present in the data from Venezuela. Those response styles could be captured by an IR-tree model, which outperformed a standard unidimensional model. Furthermore, the results indicated that the estimates from the unidimensional GRM may be biased by response styles to some extent. This will be discussed further in the following section on data from multiple countries.

Multiple-Group IR-Tree Model for Data From Multiple Countries. A multiple-group model for 11 countries was specified to investigate differences in organizational commitment as well as response styles across countries. The item parameters were set equal across countries—thus assuming strong measurement invariance or no DIF—in order to allow for meaningful comparisons of persons across countries.[1] However, no dispensable restrictions were imposed on the means, variances, and covariances of target trait, ERS, and MRS thus permitting group differences with respect to these parameters.

Again, the IR-tree model (AIC = 66,200, BIC = 66,950) fit better than the GRM (AIC = 67,460, BIC = 67,740). Furthermore, there was considerable heterogeneity of the country-level parameters. In Figure 5, the estimated latent variances and means are shown. More precisely,
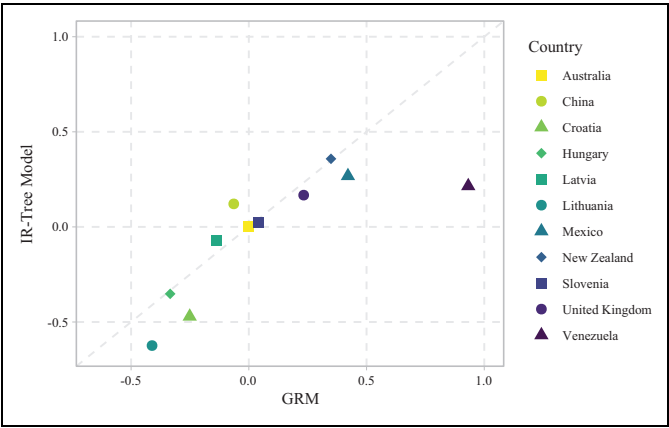
Figure 6. Scatterplot of country level means of organizational commitment estimated from both an IR-tree model and a graded response model (GRM).

relative means are shown (i.e., the lowest country mean level was subtracted from all means) to facilitate interpretation. With respect to the target trait, the country level of organizational commitment was lowest in Lithuania and highest in New Zealand, and the variances showed considerable heterogeneity across countries as well. Even larger differences in country mean levels were observed for ERS, with the highest level of ERS observed in Venezuela. Respondents in China or Lithuania, in contrast, showed rather low levels of ERS and seldom used the Categories 1 and 5 (see also the appendix).

In order to estimate the impact of response styles on substantive conclusions, the country level means of organizational commitment were compared across the IR-tree model and the GRM. The rank correlation of the means across models was $\tau = .78$ as illustrated in Figure 6. As expected, a large correction occurred for Venezuela: In the GRM, this country had by far the largest mean of 0.93. In the IR-tree model, however, the mean equaled 0.22 taking only the third place.[2]

Summary

This illustrative example showed that IR-tree models can add value to data analysis in organizational and also cross-cultural psychology. It was demonstrated how the model parameters can be interpreted and how an IR-tree model differs from a standard unidimensional model. The example also indicated that response styles can have an effect on substantive conclusions, but it should also be noted that such effects are rather small under many conditions (e.g., Plieninger, 2017). Herein, notable differences in organizational commitment after controlling for response styles were found for one out of 11 countries. Moreover, it was shown that IR-tree models can easily be extended to multiple-group models. Finally, it should be noted that the examples were handpicked for illustrative purposes. Thus, even though ISSP 2015 is a very large and reliable data set, substantive conclusions should not rely solely on the presented results.

Discussion

The contribution of the present article is twofold. The first contribution is the development of guidelines for scholars in organizational research and beyond for developing, applying, and

interpreting IR-tree models. Thus, the article serves as a supplement to earlier psychometric work on IR-tree models (e.g., Böckenholt, 2012; Böckenholt & Meiser, 2017). In more detail, the guidelines described above enable researchers to select an existing or even develop a new IR-tree model—by going from a descriptive tree diagram via a set of model equations to the pseudo-items for estimation. Moreover, the worked example illustrated how parameter estimates and model fit can be interpreted. Furthermore, it was shown that this interpretation should focus on the qualitative differences between the model parameters, but not on their relative order in the tree diagram.

The second contribution is the introduction of the novel class of multiple-group IR-tree models. These are applicable when data were collected in possibly heterogeneous groups such as organizations, countries, experimental conditions, or across different time points. Just like multiple-group IRT models, multiple-group IR-tree models allow one to investigate and establish measurement invariance across groups. Furthermore, groups can be compared with respect to all latent dimensions of the IR-tree model (e.g., target trait, ERS, and MRS). Apart from that, note that multiple-group models are best suited for situations with a few, known groups. Many groups such as dozens of companies may make working with a multiple-group model cumbersome, and multilevel models may be an alternative. Furthermore, if the groups are unknown, a latent-class extension may be required (e.g., Tijmstra et al., 2018).

One potential area for IR-tree models in organizational research is the control of response styles such as ERS and MRS. Response styles and other sources of method variance are receiving more and more attention in many areas (e.g., Baumgartner & Steenkamp, 2001; Kam & Meyer, 2015; Moors, 2012; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), and IR-tree models have already been successfully applied in this context. For example, Plieninger and Meiser (2014) predicted measures of academic performance using self-report measures that were controlled for response styles using an IR-tree model. Likewise, LaHuis et al. (2019) predicted job performance using variables such as self-reported teamwork or work ethic.

Further promising areas for using IR-tree models are measures that are multidimensional and result in discrete data (e.g., Likert-type scales, choice data, SJTs). For example, scoring SJTs can be difficult (e.g., Bergman, Drasgow, Donovan, Henning, & Juraska, 2006; Zu & Kyllonen, 2018), and IR-tree models might potentially be used to disentangle different dimensions underlying the responses of a forced-choice SJT. Apart from that, research on choices (e.g., consumer choices) often involves alternative options that differ along multiple dimensions; IR-tree models and related accounts (e.g., Böckenholt, 2012; McFadden, 1981) allow to capture this multidimensionality and to analyze all choices in one model. All this illustrates that IR-tree modeling has become a vibrant topic in certain areas (e.g., Khorramdel, Jeon, & Wang, 2019), and organizational researchers may benefit from adding this model class to their toolbox.

In summary, IR-tree models enable us to shed new light on existing research questions, for instance, about the processes involved in questionnaire responding. Furthermore, they also allow to ask novel research questions, for example, about multidimensionality in self-report data. In either case, the present article offers guidance for working with IR-tree models, in both single-group and multiple-group cases.

Appendix

Response Distribution of Workers' Commitment Across Countries
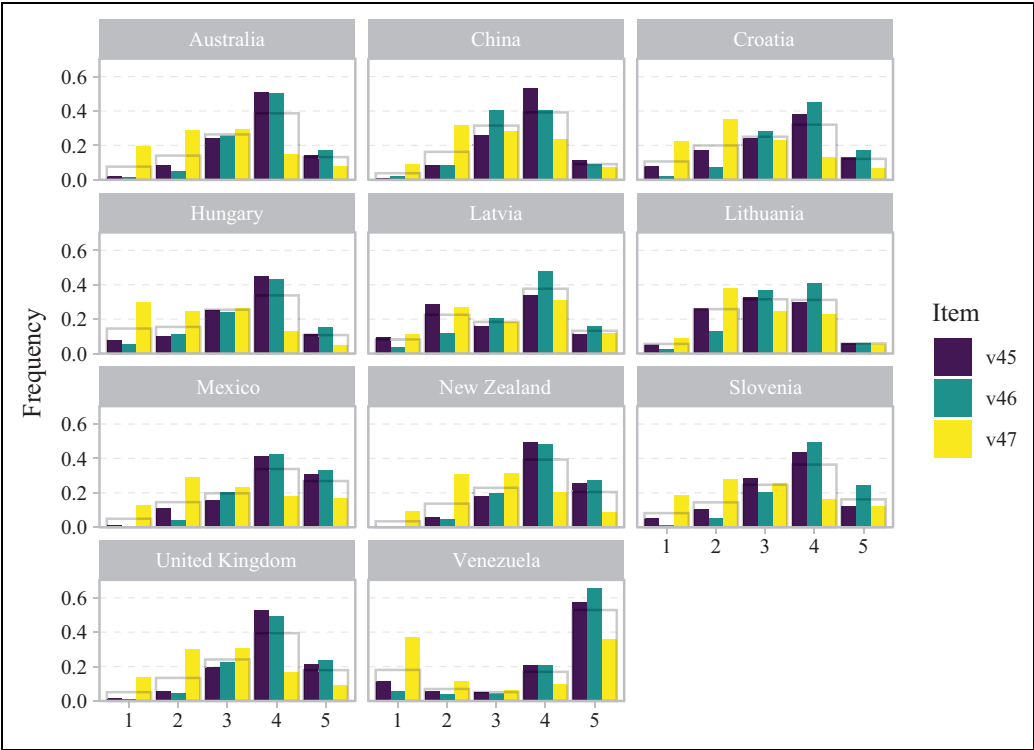
Figure A1. The colored bar plots show, across all respondents within each country, the response distribution of the three items measuring organizational commitment. Depicted in light gray is the response distribution averaged across the three items. The response scale ranged from *strongly agree* (5) to *strongly disagree* (1).

Author Note

Declaration of Conflicting Interests

Funding

ORCID iD

Hansjörg Plieninger https://orcid.org/0000-0002-4416-300X

Notes

1. In substantive research, invariance should be empirically tested (e.g., using model comparisons) as extensively discussed in the general multiple-group literature (e.g., Tay et al., 2015). Herein, invariance was assumed rather than investigated for the sake of brevity of these illustrative analyses.
1. Note that, strictly speaking, this interpretation is to some degree limited by the identification constraint. In detail, the means of one country have to be fixed to identify the model (see above), and Australia was chosen herein (i.e., $\mu_{g=1} = \mathbf{0}$). However, this choice is absolutely arbitrary but not without consequences. One could also fix the means of Venezuela, for example. Figuratively speaking, this would shift all points in Figure 6 to the lower left such that Venezuela took the place of Australia. Then, no correction of Venezuela's absolute mean would occur, because it equaled 0 in both models; for all other countries, however, large upward corrections would occur. Nevertheless, differences between countries would remain the same, and thus the correction of the rank of Venezuela is not affected by this limitation.

References

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*, 143-156. doi:10.1509/jmkr.38.2.143.18840

Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, *14*, 223-235. doi:10.1111/j.1468-2389.2006.00345.x

Bock, R. D., & Zimowski, M. F. (1997). Multiple group IRT. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 433-448). doi:10.1007/978-1-4757-2691-6_25

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665-678. doi:10.1037/a0028111

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. doi:10.1111/bmsp.12086

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*, 1-27. doi:10.1177/014920639902500101

Debeer, D., Janssen, R., & De Boeck, P. (2017). Modeling skipped and not-reached items using IRTrees. *Journal of Educational Measurement*, *54*, 333-363. doi:10.1111/jedm.12147

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, *48*(1), 1-28. doi:10.18637/jss.v048.c01

DiTrapani, J., Jeon, M., De Boeck, P., & Partchev, I. (2016). Attempting to differentiate fast and slow intelligence: Using generalized item response trees to examine the role of speed on intelligence tests. *Intelligence*, *56*, 82-92. doi:10.1016/j.intell.2016.02.012

ISSP Research Group. (2017). *International Social Survey Programme: Work orientations IV—ISSP 2015* [Data file, Version 2.1.0]. doi:10.4232/1.12848

Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods*, *48*, 1070-1085. doi:10.3758/s13428-015-0631-y

Jeon, M., De Boeck, P., & van der Linden, W. (2017). Modeling answer change behavior: An application of a generalized item response tree model. *Journal of Educational and Behavioral Statistics*, *42*, 467-490. doi:10.3102/1076998616688015

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*, 512-541. doi:10.1177/1094428115571894

Khorramdel, L., Jeon, M., & Wang, L. L. (2019). Advances in modelling response styles and related phenomena. *British Journal of Mathematical and Statistical Psychology*, *72*, 393-400. doi:10.1111/bmsp.12190

Khorramdel, 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. doi:10.1080/00273171.2013.866536

LaHuis, D. M., Blackmore, C. E., Bryant-Lees, K. B., & Delgado, K. (2019). Applying item response trees to personality data in the selection context. *Organizational Research Methods*, *22*, 1007-1018. doi:10.1177/1094428118780310

Lang, J. W. B., Lievens, F., De Fruyt, F., Zettler, I., & Tackett, J. L. (2019). Assessing meaningful within-person variability in Likert-scale rated personality descriptions: An IRT tree approach. *Psychological Assessment*, *31*, 474-487. doi:10.1037/pas0000600

Lievens, F., Lang, J. W. B., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. *Journal of Applied Psychology*, *103*, 753-771. doi:10.1037/apl0000280

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*, 523-547. doi:10.1007/BF02294327

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*, 205-235. doi:10.1007/s11336-013-9374-9

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 198-272). Cambridge, MA: MIT Press.

Meiser, T., Plieninger, H., & Henninger, M. (2019). IRTree models with ordinal and multidimensional decision nodes for response styles and trait-based rating responses. *British Journal of Mathematical and Statistical Psychology*, *72*, 501-516. doi:10.1111/bmsp.12158

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transactional, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*, 271-298. doi:10.1080/1359432X.2010.550680

Muthén, L. K., & Muthén, B. O. (2012). *Mplus. Statistical analysis with latent variables* (Version 7.4). Los Angeles, CA: Muthén & Muthén.

Plieninger, H. (2017). 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., & Heck, D. W. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research*, *53*, 633-654. doi:10.1080/00273171.2018.1469966

Plieninger, H., & Meiser, T. (2014). Validity of multiprocess IRT models for separating content and response styles. *Educational and Psychological Measurement*, *74*, 875-899. doi:10.1177/0013164413514998

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903. doi:10.1037/0021-9010.88.5.879

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*, 517-529. doi:10.1037/0021-9010.87.3.517

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552-566. doi:10.1037/0033-2909.114.3.552

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318-339. doi:10.1037/0033-295X.95.3.318

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, *18*, 3-46. doi:10.1177/1094428114553062

Tijmstra, J., Bolsinova, M., & Jeon, M. (2018). General mixture item response models with different item response structures: Exposition with an application to Likert scales. *Behavior Research Methods*, *50*, 2325-2344. doi:10.3758/s13428-017-0997-0

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

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70. doi:10.1177/109442810031002

Zettler, I., Lang, J. W. B., Hülsheger, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self- and observer reports. *Journal of Personality*, *84*, 461-472. doi:10.1111/jopy.12172

Zu, J., & Kyllonen, P. C. (2018). Nominal response model is useful for scoring multiple-choice situational judgment tests. *Organizational Research Methods*, *23*, 342-366. doi:10.1177/1094428118812669