Deepening Response-Style Modeling: Framework Evolution and Applications of the Item Response Tree*

Dong Jieyuan, Liu Hongyun

2026-08-19
ChinaRxiv

Abstract

Item response trees provide a refined framework for separating and understanding response styles by decomposing responses to Likert scales into multiple decision nodes. In recent years, the item response tree framework has continued to evolve, promoting deeper research on response styles. At the node level, item response tree models have been continuously expanded from within-node unidimensional structures to within-node multidimensional structures, and from models in which all nodes follow a dominance process to attempts to introduce ideal point processes, thereby enriching the representation of the theoretical mechanisms underlying responses. At the response process level, research on heterogeneity has gradually accumulated through between-person mixture modeling, modeling of within-response path uncertainty, and within-person dynamic modeling, leading to increasingly in-depth characterization of the process. Based on a systematic review of key advances, this paper summarizes a progressive modeling and analysis workflow and provides corresponding R implementation functions and empirical examples. Finally, it discusses two core research perspectives in response style modeling—controlling measurement bias and understanding the response process—as well as future directions for the development of item response trees.

Full Text

Deepening Response-Style Modeling: Framework Evolution and Applications of the Item Response Tree*

Dong Jieyuan Liu Hongyun

(Department of Psychology, Beijing Normal University; Beijing Key Laboratory of Applied Experimental Psychology; National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Beijing 100875)

Abstract The item response tree provides a refined framework for separating and understanding response styles by decomposing Likert-scale responses into multiple decision nodes. In recent years, the item response tree framework has continued to evolve, promoting the deepening of response-style research. At the node level, item response tree models have been continuously expanded: from unidimensional within-node structures to multidimensional within-node structures; from treating nodes uniformly as dominance processes to attempts to introduce ideal-point processes; and the representation of response-theory mechanisms has become richer. At the response-process level, investigations of heterogeneity have gradually accumulated, including mixed modeling across individuals, modeling of uncertainty in within-response paths, and dynamic modeling within individuals, leading to increasingly deep depictions of the process. On the basis of a systematic review of key developments, this paper summarizes a progressive modeling-analysis workflow and provides corresponding R-language implementation functions and validation examples. Finally, it discusses two core research perspectives in response-style modeling—control of measurement bias and understanding of the response process—as well as future directions for the development of item response trees.

Keywords response style; item response tree; Likert scale; item response theory

Classification number B841

In fields such as psychology, sociology, and education, self-report questionnaires have become a mainstream tool for measuring individuals' attitudes, beliefs, and traits because they are easy to administer and efficient to test. However, respondents' answers are often not based entirely on item content; they may also be affected by response style (RS). RS refers to a systematic tendency for respondents to answer questionnaire items independently of item content and on the basis of irrelevant factors (Baumgartner & Steenkamp, 2001; Paulhus, 1991). Within the framework of item response theory (IRT), it is typically manifested as excessive selection of certain specific options on Likert scales (Bolt & Meng, 2025; van Vaerenbergh & Thomas, 2013). Common types include the extreme response style (ERS), which tends to select endpoint options; the midpoint response style (MRS), which tends to select middle options; and the acquiescence response style (ARS), which tends to select "agree" -side options.

Date received: 2026-03-25

*Supported by the National Natural Science Foundation of China (32471145) and by the achievements of the Beijing Universities Philosophy and Social Sciences Innovation Center.

Corresponding author: Liu Hongyun, E-mail: hyliu@bnu.edu.cn

etc. As an important source of common method bias (Podsakoff et al., 2012), RS can contaminate latent trait estimation, weaken questionnaire reliability, and distort analytical conclusions about measurement models and relationships among variables (Adams et al., 2019; Zhang et al., 2022; Zóltak et al., 2025). However, from cognitive and cross-cultural perspectives, RS is also a key element of response strategies: at the cognitive level, RS may manifest as a satisficing strategy adopted by respondents when coping with the cognitive load of questionnaires, possibly driven by a tendency to conserve cognitive resources (Merhof & Meiser, 2023); at the cross-cultural level, RS is a cultural-psychological phenomenon worth exploring. For example, ERS is associated with cultural adaptation strategies among immigrant groups and is reflected in argumentation styles and their intergenerational differences (Morren et al., 2012), and on the same scales, Chinese samples often show less ERS and more MRS than Western samples (Guo & Spina, 2019; Harzing et al., 2012). It can therefore be seen that two analytical perspectives coexist in RS research: one regards RS as measurement bias that needs to be controlled, while the other regards it as a cognitive process worthy of investigation. The two perspectives each have their own emphasis and can also complement one another.

To measure and control RS, researchers have carried out many explorations. Early counting methods quantified RS by calculating the frequency or proportion with which respondents used particular categories (Cheung & Rensvold, 2000), but they could not separate RS from the trait of interest (TOI). Subsequently, IRT-based modeling methods gradually became mainstream, including mixed IRT models (Austin et al., 2006), multidimensional IRT models (Bolt & Johnson, 2009), and others. Likewise within the IRT framework, Böckenholt (2012) proposed the item response tree (IRTree) model. This model quickly attracted attention and became a frontier approach for RS modeling. Unlike previous IRT models, the IRTree model decomposes ordered categorical responses into a series of binary decision nodes, each corresponding to a different decision stage (e.g., whether to agree, whether to choose an extreme option). It assumes that RS and TOI operate at different decision stages, thereby both enabling their separation and helping to reveal respondents' response processes. Thus, the IRTree model can take into account the dual analytical perspectives of bias control and process understanding. In recent years, the IRTree model has continued to evolve and has generated multiple variants: at the level of node structure, it has expanded from unidimensional within-node structures to multidimensional within-node structures, and from dominant process assumptions to ideal-point process assumptions; at the level of response processes, hybrid modeling, single-outcome multi-path modeling, and dynamic modeling have successively appeared, in order to capture interindividual differences, uncertainty in within-response paths, and intraindividual dynamic change. Today, the IRTree model has become an important tool for RS modeling; Henninger et al. (2025) evaluated it as the current state-of-the-art method for RS modeling.

However, although variants of IRTree models have become increasingly abundant, questions about what theoretical assumptions different models each imply and what analytical scenarios they are suitable for—these basic issues, while discussed sporadically in the original literature on the various models—still lack systematic sorting, comparison, and integration. Existing tutorials or reviews of IRTree models either focus only on a particular class of extensions (Plieninger, 2021; Xue & Chen, 2023), or, when reviewing various RS modeling methods, discuss IRTree models only in general terms (Bolt & Meng, 2025; Zhang Yanyi, Wang Huahui, 2019); especially in China, because the relevant reviews are relatively early, they only introduced IRTree

developmental period (Zhang Jiyuan, Wang Yehui, 2019). This situation poses certain obstacles for researchers: methodological researchers, lacking a systematic sorting of the evolutionary lineage of IRTree, find it difficult to further improve and extend the model; applied researchers, lacking clear theoretical assumptions and practical guidance, can easily be at a loss when faced with numerous variants.

To address the above issues, this article systematically reviews the framework evolution of the IRTree model in the field of RS modeling. The focus here is mainly based on the following considerations. First, methodological research around IRTree has continued to accumulate, promoting a shift in RS research from simple bias control toward attention to process understanding; the value of this shift in research orientation merits a staged review. Second, a series of extensions of the IRTree model are not merely sporadic technical updates, but rather refinements of modeling around the core proposition of representing the response process, which constitutes another important level in the deepening of RS research. In addition, current domestic applications of IRTree remain very limited; the crux lies in the lack of Chinese-language guidance that combines theoretical depth with practical operation. Therefore, this article first systematically reviews the evolutionary course of IRTree from basic models to frontier extensions, summarizes the statistical structures, theoretical assumptions, and applicable boundaries of major IRTree variant models in light of existing research findings, and provides a stepwise analytic procedure and empirical demonstrations, so as to serve researchers' theoretical understanding and practical application. The following first elaborates the basic framework of IRTree, then examines its evolution in modeling depth: first introducing innovations at the node level; reviewing extensions from unidimensional structures within nodes to the incorporation of multidimensional structures within nodes, and from dominance processes to the consideration of ideal-point processes; then introducing the incorporation of heterogeneity at the response-process level, covering mixture modeling, single-outcome multipath modeling, and dynamic modeling. Finally, based on a review of model variants, general procedural recommendations for empirical analysis are given, and the practical value and research prospects of the IRTree framework are presented through an empirical application example and a comprehensive discussion.

1 Measurement Response Styles: The Basic Framework of Item Response Trees

1.1 Construction of the Item Response Tree Model

Since it was first proposed and applied to RS modeling, the core of the IRTree model has always revolved around three elements: tree diagrams, model equations, and pseudo-items. The construction of an IRTree model usually begins with drawing a tree diagram, which can intuitively show researchers' theoretical assumptions about how responses are decomposed into multiple decision processes. Taking the commonly used 5-point Likert scale as an example ("1" indicates "strongly disagree," and "5" indicates "strongly agree"), one widely adopted decomposition method is shown in Figure 1. It assumes that behind respondent i's response Y_{ij} to item j, there are three decision processes: whether to choose the middle option (recorded as node 1), whether to choose the "agree" side (recorded as node 2), and whether to choose an extreme option (recorded

Figure 1. Hypothesized response process of the 5-point scoring IRTree model (number of nodes = 3)

Figure 1: Figure 1. Hypothesized response process of the 5-point scoring IRTree model (number of nodes = 3)

as node 3).

Figure 1 Hypothesized response process of the 5-point scoring IRTree model (number of nodes = 3)

Note: Y_{ij} denotes examinee i' s response to item j; Pr_{1ij}, Pr_{2ij}, and Pr_{3ij} respectively denote the probability of choosing the middle option at node 1, the probability of choosing the "agree" side at node 2, and the probability of choosing an extreme option at node 3.

Let Pr_{1ij}, Pr_{2ij}, and Pr_{3ij} respectively denote the probabilities of making a "yes" decision at the corresponding nodes. Based on the multiplication rule for conditional probabilities, the selection probabilities for each response category can be expressed as follows:

$$P(Y_{ij} = 5) = (1 - Pr_{1ij})Pr_{2ij}Pr_{3ij}, \qquad (1)$$

$$P(Y_{ij} = 4) = (1 - Pr_{1ij})Pr_{2ij}(1 - Pr_{3ij}), \qquad (2)$$

$$P(Y_{ij} = 3) = Pr_{1ij}, \qquad (3)$$

$$P(Y_{ij} = 2) = (1 - Pr_{1ij})(1 - Pr_{2ij})(1 - Pr_{3ij}), \qquad (4)$$

$$P(Y_{ij} = 1) = (1 - Pr_{1ij})(1 - Pr_{2ij})Pr_{3ij}. \qquad (5)$$

The basic IRTree model assumes that the three decision processes are respectively influenced by the respondent parameters MRS, TOI, and ERS. The probabilities at the respective nodes, Pr_{1ij}, Pr_{2ij}, and Pr_{3ij}, can further be parameterized by an IRT model; for example, they may be specified as a one-parameter logistic model:

$$Pr_{1ij} = logit^{-1}(\eta_{1i} - \beta_{1j}), \qquad (6)$$

$$Pr_{2ij} = logit^{-1}(\theta_i - \beta_{2j}), \qquad (7)$$

$$Pr_{3ij} = logit^{-1}(\eta_{2i} - \beta_{3j}), \qquad (8)$$

where $logit^{-1}(x) = 1/(1+\exp(-x))$; β_{1j}, β_{2j}, and β_{3j} denote the difficulty parameters of item j at the respective nodes; η_{1i}, θ_i, and η_{2i} denote examinee i' s levels of MRS, TOI, and ERS, respectively (in this article, θ is used to denote TOI, and η is used to denote RS). Equations (1) through (8) together constitute the equations of a complete IRTree model.

The above IRTree model can be estimated using the software *Mplus* (Muthén & Muthén, 1998-2022) or the R package *mirt* (Chalmers, 2012). Before estimation, the original responses need to be recoded as pseudo-item responses corresponding to each decision node. The pseudo-

the item records the binary decision result for each node (usually 1 denotes "yes," and 0 denotes "no"). According to the logic of the tree diagram, the original response Y_{ij} can be encoded as $Y_{1ij}, Y_{2ij}, Y_{3ij}$ following Table 1. Through this discretized encoding, the IRTree model can be transformed into a three-dimensional 0–1 scoring IRT model for parameter estimation; here, "three-dimensional"means that the three pseudo-items decomposed from each original item correspond respectively to the three dimensions MRS, TOI, and ERS, thereby forming a multidimensional structure among nodes.

Table 1 Pseudo-items of the 5-point scoring IRTree model (number of nodes = 3)

Pseudo-item	Response category	Response category	Response category	Response category	Response category
Response	1	2	3	4	5
Y_{1ij}	0	0	1	0	0
Y_{2ij}	0	0	–	1	1
Y_{3ij}	1	0	–	0	1

Note: $Y_{1ij}, Y_{2ij}, Y_{3ij}$ correspond respectively to node 1 (whether a midpoint response is selected), node 2 (whether the response is agreement), and node 3 (whether an extreme response is selected). "–" indicates that the node is not applicable (planned missing), that is, under the corresponding response category, no judgment is required for that pseudo-item.

1.2 Understanding the Structure of the Item Response Tree Model

Unlike traditional polytomous-scoring IRT models such as the graded response model (GRM) or the generalized partial credit model (GPCM), the IRTree model does not assume that all category differences are determined by a single latent trait TOI and its thresholds. Instead, it decomposes polytomous responses into multiple binary decision nodes, and each node can be assigned different person parameters. Although information is decomposed into different pseudo-items, the entire set of pseudo-items as a whole still completely preserves all information in the original ordered response. This completeness of information is directly guaranteed by the likelihood decomposition structure of the IRTree; under the premise that the tree structure is reasonably specified, the product of the conditional independent likelihoods of the pseudo-items is equal to the likelihood function of the original response (Li, 2022). In addition, the tree structure of the IRTree also provides a cornerstone for separating different cognitive processes, and each node parameter (such as difficulty) has an intuitive interpretation that directly corresponds to a specific decision process (Khorramdel & Von Davier, 2014).

However, when implementing the tree structure as model equations, it is statistically necessary to clarify that, except for the starting node, the probabilities at the nodes are all conditional probabilities. For example, in the 5-point scoring example in Section 1.1, the probability that node 1 selects the midpoint category "3" is $Pr_{1ij} = P(Y_{1ij} = 1)$; the probability that node 2 selects "4" or "5" under the condition of not selecting "3" is $Pr_{2ij} = P(Y_{2ij} = 1|Y_{1ij} = 0)$; and the probability that node 3 selects "1" or "5" under the condition of not selecting "3" and having selected "4" or "5" is $Pr_{3ij} = P(Y_{3ij} = 1|Y_{1ij} = 0, Y_{2ij} = 1)$. Thus, formula (1) can be regarded as a decomposition of the marginal probability of selecting category "5" :

$$P(Y_{1ij} = 0, Y_{2ij} = 1, Y_{3ij} = 1) = P(Y_{1ij} = 0)P(Y_{2ij} = 1|Y_{1ij} = 0)P(Y_{3ij} = 1|Y_{1ij} = 0, Y_{2ij} = 1). \tag{9}$$

It should be noted that conditional probabilities only represent logical dependence and are not equivalent to the temporal order of psychological processing. A tree diagram merely depicts the number and types of psychological processes involved in each response category, and does not presuppose that these processes occur in a particular order (Böckenholt & Meiser, 2017; Plieninger & Heck, 2018). For example, behind the behavior of choosing "strongly agree" may lie three decisions—whether to make an extreme response, whether to agree, and whether to make an extreme response—but these are not necessarily experienced sequentially. This can be demonstrated by the chain multiplication rule: if the order of any nodes is exchanged, the joint probability can be expressed as a product of different conditional probabilities. For example, after exchanging the order of node 1 and node 3, the right-hand side of Equation (9) becomes as shown in Equation (10):

$$P(Y_{1ij} = 0, Y_{2ij} = 1, Y_{3ij} = 1) = P(Y_{3ij} = 1)P(Y_{2ij} = 1|Y_{3ij} = 1)P(Y_{1ij} = 0|Y_{2ij} = 1, Y_{3ij} = 1), \tag{10}$$

Here, the first term on the right-hand side, $P(Y_{3ij} = 1)$, is the marginal probability of choosing the extreme category "1" or "5"; the second term is the probability of choosing "4" or "5" conditional on having already made an extreme response; and the third term is the probability of not choosing "3" conditional on having made the first two decisions (which obviously takes the value 1). This indicates that the IRTree model is insensitive to node order. Plieninger (2021) also emphasized that a tree structure characterized by conditional probabilities appears to imply some kind of order, but it should not be overinterpreted: what the tree diagram presents is only one possible order, and one cannot infer in reverse that this order is a necessary assumption of the model. In practice, when describing modeling assumptions, one may use an "ordered interpretation" of the response process, but it must be recognized that the statistical structure described by the IRTree model concerns which processes are included and which parameters affect each process, and does not restrict their order. When interpreting results, a "component interpretation" (i.e., one that emphasizes only the psychological components represented by each node and the meanings of their parameters), which is weaker than an "ordered interpretation," is more robust.

The above is an explanation of the general structure of IRTree. For a tree structure that includes symmetric branches for measuring ERS, the issue of directional invariance must also be considered. Directional invariance in IRTree usually means that the effect of ERS is completely consistent on the "agree" and "disagree" sides. This implies that the model uses only one ERS latent variable and one set of pseudo-item parameters to simultaneously depict extreme responses in the two attitude directions. If directional invariance holds, then it supports treating ERS as a cross-directional general tendency toward extreme responding, which facilitates model simplification and interpretation. In reality, however, respondents may show different preferences for positive and negative extremes—for example, they may tend to choose "strongly agree" while avoiding "strongly disagree" ; some items may also be more likely to elicit positive extreme responses rather than negative extreme responses. At this point, relaxing the assumption of directional invariance and modeling with two latent variables (positive ERS and negative ERS) and two sets of pseudo-item parameters can capture such differences more finely. Jeon and De Boeck (2019) recommended beginning with a benchmark model and then gradually relaxing the assumption of directional invariance, selecting the model structure with greater explanatory power by comparing model fit. Their research results based on multiple sets of empirical data showed that directional invariance at the latent-variable level usually holds—that is, there exists a respondent parameter for ERS that is consistent across directions; however, directional invariance at the pseudo-item parameter level often does not hold, and relaxing the parameter constraints can significantly improve model fit. This suggests that when applying the IRTree model to measure ERS, it is best to regard the extreme-response processes in the two attitude directions as different nodes. For example, for a 5-point scale, the IRTree model can be extended to the four-way pseudo-item coding shown in Table 2, with the corresponding pseudo-item parameter at node 4 added to the model equation.

Table 2 Pseudo-items of the 5-point scored IRTree model (number of nodes =

4)

Pseudo-item	Response category	Response category	Response category	Response category	Response category
Response	1	2	3	4	5
Y_{1ij}	0	0	1	0	0
Y_{2ij}	0	0	–	1	1
Y_{3ij}	1	0	–	–	–
Y_{4ij}	–	–	–	0	1

Note: Y_{1ij}, Y_{2ij}, Y_{3ij}, and Y_{4ij} correspond respectively to node 1 (whether the response is a midpoint response), node 2 (whether the response is agreement), node 3 (whether, when disagreeing, the response is an extreme response), and node 4 (whether, when agreeing, the response is an extreme response). "–" indicates that the node is not applicable (planned missing), that is, the pseudo-item does not need to be judged under the corresponding response category.

Existing literature also often discusses the IRTree model within two broader modeling frameworks: the multinomial processing tree (MPT) model and the generalized linear mixed model (GLMM). The MPT model originated in cognitive psychology. It likewise assumes that, when completing a given task, respondents undergo multiple cognitive processing stages; by fitting behavioral data, the MPT model can infer latent psychological processes and their contribution rates (Liu Yuanyuan et al., 2019). Traditional MPT models typically assume that respondents have homogeneous parameters under the same experimental condition, but subsequent studies have gradually introduced multilevel modeling to examine individual differences (Klauer, 2010; Matzke et al., 2015). The proposal and development of MPT has had an important influence on IRTree. The tree representation of IRTree is in the same structural line of thought as MPT, and early research often regarded IRTree as an application of multilevel MPT in psychological measurement (Böckenholt, 2012; Khorramdel & Von Davier, 2014). On the other hand, De Boeck and Partchev (2012) systematically interpreted the IRTree model from the GLMM perspective: the pseudo-items at each node correspond, through the link function, to a linear predictor, which includes both fixed effects at the item level (difficulty parameters) and random effects at the respondent level (TOI or RS). This perspective allows IRTree to be implemented with the aid of GLMM estimation systems (e.g., the R package lme4 (Bates et al., 2015)) and facilitates model extensions, such as Jeon and De Boeck' s (2016) discussion of how to incorporate node-level effects. However, a limitation of GLMM is that its linear predictor can correspond only to single-parameter IRT models. With the development of IRTree models, two-parameter IRT models have also gradually been used to fit pseudo-items (Böckenholt, 2017; Jeon & De Boeck, 2019). For example, formula (6) can be extended as

$$Pr_{1ij} = logit^{-1}(\alpha_{1j}\eta_{1i} - \beta_{1j}), \qquad (11)$$

where α_{1j} is the discrimination parameter of the pseudo-item, reflecting the ability of item j at node 1 to distinguish different levels of MRS.

1.3 Theoretical Features of the Item Response Tree Model for Modeling Response Styles

Before reviewing recent advances in IRTtree models, it is necessary to place them among the various methods of RS modeling and to distinguish the distinctive features of their theoretical assumptions. The theoretical differences among RS modeling methods are first reflected in whether RS is conceptualized as latent categories or as latent traits. The former, such as mixture IRT models (Austin et al., 2006; Huang, 2016) or mixture factor-analytic models (Arias et al., 2023), assumes that there are latent categories such as "ordinary responding" and "extreme responding," allowing the parameters of IRT models or factor-analytic models to vary across different categories, thereby identifying groups with different response strategies. In contrast, IRTtree models regard ERS, MRS, and the like as continuous latent traits, which can not only model differences in RS intensity across individuals but also estimate the correlations between RS and TOI. Another representative model of continuous-trait theory is the multidimensional nominal response model (MNRM), which is based on a "divide-by-total" structure and, together with a predefined scoring function, models all response categories in a single step (Henninger & Meiser, 2020). Compared with the latent-category method, which assumes homogeneous ERS levels within categories, the continuous-trait method can model differences in ERS intensity across individuals and is therefore more flexible (Schoenmakers et al., 2024).

Schoenmakers et al. (2024) further revealed key differences in the theoretical assumptions of MNRM and IRTtree models: the former assumes that RS and TOI jointly exert a global influence on the choice probabilities of all categories, whereas the latter adopts a multiprocess decision-making perspective, allowing RS to affect only specific nodes. For example, in a typical IRTtree model, ERS usually acts only on the node of "whether to choose an extreme option" ; in MNRM, however, changing only the ERS level also changes the probability expectation that an individual will choose the "agree" side—an unintended side effect that researchers may not desire and yet often overlook (Bolt & Meng, 2025; Schoenmakers et al., 2024). At present, there is still no consensus on which of the two types of models is superior; both are mainstream RS modeling methods. Considering that both belong to a confirmatory framework, researchers in practice need to emphasize theoretical construction and choose an appropriate method based on theoretical assumptions: when there is no explicit node-decomposition theory or when researchers do not care about the components of the response process, MNRM' s one-step global modeling may be more convenient; when theory clearly indicates that RS affects specific decision stages, the structured decomposition of IRTtree is more suitable.

2 Framework Evolution at the Node Level: Extensions of Latent-Variable Structures and Response Functions

On the basis of the early IRTtree framework, a series of methodological studies have promoted sustained evolution of the IRTtree framework in different directions. Among them, structural extensions at the node level are especially central; when this paper reviews the evolution of the framework, it treats them as the first main line. The basic IRTtree implicitly contains two key assumptions at the node level: first, a unidimensional structure within each node, that is, the response at each node is driven by only a single latent variable; second, all response functions are dominance processes, that is, the response probability at each node is assumed to vary monotonically with the corresponding latent variable. As research has deepened, researchers have hoped to develop theories that more closely fit actual response mechanisms, and therefore have attempted to relax the above assumptions to improve the representational flexibility of IRTtree models at the node level. This section therefore reviews node-level ⋯

expansions: first, the incorporation of multidimensional structures within nodes is discussed, allowing a single node to be simultaneously influenced by multiple person parameters; next, the incorporation of ideal-point processes is discussed, allowing nodes to use nonmonotonic response functions; finally, the combined use of the two types of extensions is discussed.

2.1 Extension of the Latent-Variable Structure: Incorporating Multidimensional Structures Within Nodes

Although the basic IRTree model as a whole presents a "between-node multidimensional" structure (different nodes measure different person parameters), within each node it still maintains the assumption of "within-node unidimensionality." For example, it assumes that TOI affects only the direction decision (agree or disagree), ERS affects only the extreme-response decision, and TOI and RS do not interfere with each other in their respective decision processes. However, inspired by the dual-process model of Thissen-Roe and Thissen (2013) and by satisficing response strategies (Krosnick, 1991), the Meiser group (Meiser et al., 2019; Merhof & Meiser, 2023) further extended the assumptions about the roles of TOI and RS: RS operates on local decisions, whereas the influence of TOI runs through the entire response process. To characterize this "TOI global, RS local" assumption, which resembles a bifactor model, TOI is loaded on all pseudo-items, forming a compensatory multidimensional structure of "TOI+RS" at nodes where originally only RS operated. Meiser et al. (2019) therefore proposed an IRTree model containing multidimensional structures within nodes.

For ease of understanding, the following introduces the within-node multidimensional structure using a 4-point scoring scale as an example. The response is decomposed into two processes, "whether to agree" and "whether to respond extremely." After considering the issue of directional invariance mentioned in Section 1.2, it is represented in Figure 2 as three nodes.^1 According to the

Figure 2. Response-process assumptions of a 4-point-scoring IRTree model (number of nodes = 3)

Figure 2: Figure 2. Response-process assumptions of a 4-point-scoring IRTree model (number of nodes = 3)

improved theoretical assumptions, TOI plays a global role at all nodes, whereas ERS plays a local role at node 2 and node 3. The corresponding model equations are as follows:

$$Pr_{1ij} = logit^{-1}(\alpha_{1j}\theta_i + d_{1j}), \qquad (12)$$

$$Pr_{2ij} = logit^{-1}\left(\alpha_{2j}^{(\eta)}\eta_i + \alpha_{2j}^{(\theta)}\theta_i + d_{2j}\right), \qquad (13)$$

$$Pr_{3ij} = logit^{-1}\left(\alpha_{3j}^{(\eta)}\eta_i + \alpha_{3j}^{(\theta)}\theta_i + d_{3j}\right), \qquad (14)$$

where θ_i, η_i respectively denote the TOI and ERS levels of person i; d_{1j}, d_{2j}, d_{3j} are the intercept parameters of item j at each node; $\alpha_{1j}, \alpha_{2j}^{(\eta)}, \alpha_{2j}^{(\theta)}, \alpha_{3j}^{(\eta)}, \alpha_{3j}^{(\theta)}$ are the slope parameters or discrimination parameters of item j at each node, with the superscripts used to distinguish the slopes of RS and TOI within the same node. At this point, TOI acts in opposite directions at the two extreme-response nodes: at node 2, the higher the TOI level, the more likely the person is to choose "2" rather than "1," and the expected TOI slope is $\alpha_{2j}^{(\theta)} < 0$; whereas at node 3, the higher the TOI level, the more likely the person is to choose "4" rather than "3," and the expected TOI slope is $\alpha_{3j}^{(\theta)} > 0$. When $\alpha_{2j}^{(\theta)} = \alpha_{3j}^{(\theta)} = 0$, the model reduces to a within-node unidimensional structure.

---

^1 Referring to conventions in the existing literature (Bolt & Meng, 2025; Meiser et al., 2019; Merhof & Meiser, 2024; Schoenmakers et al., 2024), this article presents only the decomposition of the response process in the tree diagram, and does not indicate whether the latent-variable structure within nodes is unidimensional or multidimensional; the latent-variable structure within nodes is embodied through the specific model equations. Thus, the same tree diagram can correspond to different model equations.

Figure 2. Response-process assumptions of a 4-point-scoring IRTree model (number of nodes = 3)

Note: Y_{ij} denotes examinee i' s response to item j; Pr_{1ij}, Pr_{2ij}, and Pr_{3ij} respectively denote the probability of choosing the "agree" side at node 1, the probability of choosing an extreme option on the "disagree" side at node 2, and the probability of choosing an extreme option on the "agree" side at node 3.

Compared with the within-node unidimensional IRTree model, the IRTree model with within-node multidimensional structure has more refined theoretical assumptions, and its advantages have also received preliminary empirical support. Based on analyses of empirical data, Meiser et al. (2019) found that the fit of the within-node multidimensional structure was superior to that of the within-node unidimensional structure. Merhof et al. (2024) further raised the issue of mimicry effects when IRTree models handle ERS, thereby further highlighting the value of within-node multidimensional structure. Mimicry effects refer to the situation in which, when the response distribution of the actual data is asymmetric (e.g., the trait distribution is skewed), in a within-node unidimensional model, ERS becomes a "mimicker" of TOI, erroneously absorbing response variation that should be explained by TOI on nodes other than the directional-decision node, thereby undermining the substantive separation between the two (Merhof et al., 2024). For example, if in a depression scale most examinees have low levels of depression and the response set is concentrated at the low end, then the within-node unidimensional model, constrained by its structure, will attribute the response asymmetry caused by TOI to ERS, thereby overestimating ERS variance and distorting its covariance with TOI. The simulation study by Merhof et al. (2024) showed that an IRTree model with within-node multidimensional structure, by allowing TOI and ERS to act jointly within a node, enables TOI itself to explain the variation it induces, thus avoiding mimicry effects: under their simulation conditions, regardless of whether the data were driven solely by TOI or jointly by TOI and RS, and regardless of whether the response distribution was asymmetric, the IRTree model with within-node multidimensional structure could provide unbiased estimates of variances and covariances. Analyses based on PISA 2018 empirical data also showed that this model can, to a certain extent, correct the bias in TOI-ERS correlation coefficients produced by the within-node unidimensional model. Other simulation studies have found that extending from within-node unidimensional to within-node multidimensional structure can make fuller use of information, reduce the standard error of TOI estimates, and improve their estimation accuracy (Bolt & Meng, 2025).

Nevertheless, the above evidence of advantages comes from limited simulation studies and applied cases, and its generalizability and robustness still require further verification by more research. In addition, within-node multidimensional structure increases the complexity of the IRTree model, bringing challenges for model identification and estimation stability.

To ensure model identifiability, constraints usually need to be imposed on the slope parameters of pseudo-items. One approach is to set, for all items,

$\alpha_{2j}^{(\theta)} = \alpha_2^{(\theta)}, \alpha_{3j}^{(\theta)} = \alpha_3^{(\theta)}$,

freely estimate these two parameters, and fix the slope parameters of the other pseudo-items, such as $\alpha_{1j}, \alpha_{2j}^{(\eta)}, \alpha_{3j}^{(\eta)}$, to 1 (Hasselhorn et al., 2024). By comparing the magnitudes of $\alpha_2^{(\theta)}$ and $\alpha_3^{(\theta)}$ with the slopes of ERS on the same nodes (fixed to 1), one can examine the relative influence of TOI and ERS on extreme response decisions. Merhof et al. (2024) further restricted $\alpha_2^{(\theta)} = -\lambda, \alpha_3^{(\theta)} = \lambda$; λ likewise measures the relative magnitude of the effects of TOI and ERS (e.g., $\lambda > 1$ indicates that, in extreme response decisions, the influence of TOI is greater than that of ERS). Other studies have adopted a more flexible constraint scheme: at nodes 2 and 3 of the same item, ERS slopes are set equal ($\alpha_{2j}^{(\eta)} = \alpha_{3j}^{(\eta)}$), while TOI slopes are mutual opposites ($\alpha_{2j}^{(\theta)} = -\alpha_{3j}^{(\theta)}$), allowing the slopes at each node to vary across items and thereby capturing richer between-item difference information (Merhof & Meiser, 2023; Schoenmakers et al., 2024). At present, parameter-constraint schemes lack explicit theoretical grounding and, in practice, are mostly data-driven. In addition, directional invariance is also more complicated under a within-node multidimensional structure. In the empirical constraint approaches described above, researchers tend to set the absolute values of the slope parameters at the two extreme nodes to be correspondingly equal (i.e., the effect strengths of ERS and TOI are the same on both sides), while freely estimating the intercept parameters. Because the intercept in a within-node multidimensional structure no longer corresponds to an explicit meaning of "baseline difficulty," freely estimating the intercept is reasonable; as for whether the assumption that the absolute values of the slopes are equal across directions holds, it can be tested using the model-comparison method of Jeon and De Boeck (2019).

In addition to parameter constraints, incorporating a within-node multidimensional structure also faces challenges regarding the stability of parameter estimation. Systematic research is still lacking on parameter recovery for this model under different sample sizes, questionnaire lengths, and constraint conditions. The simulation results of Alarcon et al. (2023) on a within-node unidimensional IRTree indicate that the precision of parameter estimation decreases as node depth increases. It can therefore be inferred that when the deepest extreme-decision node simultaneously estimates the joint effects of TOI and ERS, estimation uncertainty may be further amplified under limited sample sizes.

In summary, although the within-node multidimensional structure increases model complexity, it shows relatively clear advantages in correcting bias, separating traits, and reducing the standard error of TOI estimation. In general, this structure is applicable in the following situations: researchers have clear theoretical expectations and believe that TOI has a global influence on the response process (rather than being limited to directional decisions); the data may have an asymmetric response distribution (e.g., clinical samples, high-stakes selection); and the research purpose emphasizes accurately separating RS and TOI rather than pursuing model parsimony. When the sample size is small or the theoretical assumptions are not yet clear, the within-node unidimensional model may be a more prudent choice.

2.2 Extension of the response function: Incorporating the ideal point process

The forms of item response functions in IRT are mainly divided into two types: the dominance process and the ideal point process (Drasgow et al., 2010). Unlike the monotonically increasing item response function corresponding to the dominance process, the ideal point process assumes that the item response function is single-peaked: the expected score reaches its peak when the TOI level equals the item’ s ideal point.

the value; and regardless of whether the TOI level deviates above or below the ideal point, the expected score decreases as the deviation distance increases.

Although most measurement tools are developed on the basis of dominance-process theory, existing research evidence suggests that, in noncognitive tests such as personality questionnaires, ideal-point processes often provide a better description of response mechanisms (Deng Wen-gen et al., 2010; Drasgow et al., 2010; Liu & Wang, 2016). Therefore, when analyzing forced-choice questionnaire data, it is also necessary to consider ideal-point process models, such as the commonly used generalized graded unfolding model (GGUM; Roberts et al., 2000). The GGUM describes the probability of choosing category $y \in \{0, 1, \cdots, K\}$ with the following formula:

$$P(Y_{ij} = y) = \frac{\exp\left[\lambda_j\left(y(\xi_i - \delta_j) - \sum_{s=0}^{y} \tau_{js}\right)\right] + \exp\left[\lambda_j\left((M-y)(\xi_i - \delta_j) - \sum_{s=0}^{y} \tau_{js}\right)\right]}{\sum_{k=0}^{K}\left\{\exp\left[\lambda_j\left(k(\xi_i - \delta_j) - \sum_{s=0}^{k} \tau_{js}\right)\right] + \exp\left[\lambda_j\left((M-k)(\xi_i - \delta_j) - \sum_{s=0}^{k} \tau_{js}\right)\right]\right\}}, (15)$$

where $M = 2K + 1$, ξ_i denotes the examinee parameter, λ_j is the item discrimination parameter, τ_{js} is the item-category threshold parameter, and δ_j is the item’ s ideal-point location parameter. Compared with the GPCM, which assumes a dominance process, the GGUM has a similar “divide-by-total” structure, but each category corresponds to two exponential terms, using two sets of scoring functions, increasing and decreasing, $(0, \cdots, K)$ and $(M, \cdots, K + 1)$, respectively, to characterize a single-peaked category response curve symmetric around $(\xi_i - \delta_j)$.

In recent years, researchers have begun to attempt to integrate ideal-point processes into the IRTree framework, thereby modeling attitude responses more realistically while controlling RS. Starting from the theoretical construction at the node level of IRTree, they examine which response function should be applied at each decision stage and have developed a series of new models.

Jin et al. (2022) were the first to propose the ideal-point–dominance two-stage tree model (IDtree). Its core argument is that, in directional decisions, the examinee’s response mechanism should follow an ideal-point process rather than a dominance process. They pointed out that, for constructs such as attitudes toward social inequality, individuals will “agree” only when their own construct level is close to the location described by the item; if it is too high or too low, they will "disagree." This is precisely the single-peaked characteristic of the ideal-point process. Therefore, the model first uses the GGUM to model the directional decision process, and in the subsequent intensity decision process it uses the GRM, which follows a dominance process, to characterize ERS. This pioneering work was the first to distinguish decision processes with different response functions at the node level, but it relied on computationally time-consuming Markov chain Monte Carlo (MCMC) estimation, and could handle only ERS, with limited applicability to scales with more than 4 points. On this basis, Li et al. (2025) proposed a more general unfolded IRTree (UIRTree) model and stated the node-level argument as follows: in an IRTree, all nodes affected by TOI should consider using an ideal-point process, whereas nodes clearly affected by ERS, MRS, and the like should retain a dominance process. The model shares with the basic IRTree a similar binary decision-tree structure, can accommodate various scale formats such as 4-point and 5-point scales, and the researchers further improved computational efficiency using the marginal maximum *a posteriori* (MMAP) estimation method based on the `mirt` package. Simulation and empirical analyses showed that, compared with the GGUM that ignores RS or the basic IRTree model that ignores ideal-point processes, the UIRTree model showed advantages in both parameter estimation and model fit. The study by Zeng et al. (2025) further explored a new theoretical assumption, arguing that directional decisions and intensity decisions should be governed by two different ideal-point

driven by TOI, and, through simulation and empirical studies, revealed that strength decisions are more likely related to the characteristics of one' s own "strength of endorsement" and the distance from the item' s ideal point, rather than to pure ERS. However, the discussion in that study is currently limited to a 4-point scale, and further validation is needed in other contexts.

Researchers' understanding of applying ideal-point processes in IRTree has continued to develop, highlighting the necessity of examining response-function specifications at the node level. From the perspective of theoretical assumptions, the core premise for incorporating an ideal-point process is that, when responding, individuals make "matching judgments" rather than "level judgments" ; their responses depend on the distance between their own TOI and the item location, rather than on the absolute level of TOI. In typical behavioral measurements such as attitudes and personality, the ideal-point assumption is usually more applicable than the dominance assumption. Before modeling, the response function of TOI in directional decisions may first be determined on theoretical grounds; if the theoretical basis is insufficient, it is not advisable to assume by default that TOI follows an ideal-point process. Instead, both dominance and ideal-point competing models should be fitted simultaneously, and judgments should be made by integrating factors such as fit indices and parameter interpretability.

2.3 Realization of a multidimensional structure with two response processes coexisting within a node

The within-node multidimensional structure introduced in Section 2.1 allows individuals' extreme responses to be influenced simultaneously by TOI and RS, where both TOI and RS use dominance processes; Section 2.2 further indicates that, to characterize the influence of TOI on directional decisions, one may consider adopting an ideal-point process.

At this point, a naturally arising question is: for a model that incorporates an ideal-point process, its extreme responses may be influenced both by ERS (a dominance process) and by the distance between TOI and the item location (an ideal-point process). How to allow dominance and ideal-point processes to coexist within the same node becomes a new challenge for modeling.

To address this issue, inspired by the GGUM structure, Merhof and Meiser (2024) proposed a more general multidimensional IRT model supporting the coexistence of dominance and ideal-point processes, allowing R complementary response processes to jointly influence responding. Its core structure is given by the following formula:

$$P(Y_{ij} = y) = \frac{\exp(\sum_{r=1}^{R} \psi_{1ijyr}) + \exp(\sum_{r=1}^{R} \psi_{2ijyr})}{\sum_{k=0}^{K} \left\{\exp(\sum_{r=1}^{R} \psi_{1ijkr}) + \exp(\sum_{r=1}^{R} \psi_{2ijkr})\right\}}, \tag{16}$$

where each exponential term is internally a linear combination of R response processes. In applications, the dominance process is usually modeled by the GPCM, and the ideal-point process by the GGUM. By setting an indicator variable m_r ($m_r = 1$ indicates that process r is a dominance process, and $m_r = 0$ indicates that it is an ideal-point process), the two exponential terms for category y can be expanded respectively as

$$\sum_{r=1}^{R} \psi_{1ijyr} = \sum_{r=1}^{R} \left[(\alpha_{jr}\xi_{ir})^{m_r} + (\lambda_{jr} y(\xi_{ir} - \delta_{jr}))^{(1-m_r)}\right] + \sum_{k=0}^{y} d_{jk}, \tag{17}$$

$$\sum_{r=1}^{R} \psi_{2ijyr} = \sum_{r=1}^{R} \left[(\alpha_{jr}\xi_{ir})^{m_r} + (\lambda_{jr} (M - y)(\xi_{ir} - \delta_{jr}))^{(1-m_r)}\right] + \sum_{k=0}^{y} d_{jk}, \tag{18}$$

To ensure model identifiability, the R processes for each category do not support separate estimation of threshold parameters; only common intercept parameters can be estimated.

the parameter d_{jk} (usually setting $d_{j0} = 0$).

Based on this new model, a node can simultaneously accommodate decision processes involving different latent variables and different response functions. Simulation studies show that this new structure not only fits data generated by itself well, but also maintains stable parameter estimates when fitting data generated by a unidimensional single-process model. Conversely, if a unidimensional single-process model is used to fit data generated by the new structure, substantial bias will occur (Merhof & Meiser, 2024).

This modeling framework has also promoted the refinement of within-node multidimensional structures in IRTree models for the trait process TOI, mainly by providing a more reasonable way to handle response nodes in five-point Likert scales. Taking the four-node decomposition shown in Table 2 as an example, suppose that TOI is the trait process in the directional decision. For nodes 2 to 4, the model can be specified as follows, influenced by TOI (θ_i) and ERS (η_{2i}):

$$Pr_{2ij} = logit^{-1}(\alpha_{2j}\theta_i + d_{2j}), \qquad (19)$$

$$Pr_{3ij} = logit^{-1}\left(\alpha_{3j}^{(\eta)}\eta_{2i} + \alpha_{3j}^{(\theta)}\theta_i + d_{3j}\right), \qquad (20)$$

$$Pr_{4ij} = logit^{-1}\left(\alpha_{4j}^{(\eta)}\eta_{2i} + \alpha_{4j}^{(\theta)}\theta_i + d_{4j}\right). \qquad (21)$$

For node 1, it is usually regarded as being influenced by MRS (see Equation (6)); however, with reference to the theoretical assumption that TOI exerts a global influence, a within-node multidimensional structure involving TOI and MRS can likewise be introduced. Considering that respondents with very high or very low TOI levels will not choose the middle option, Merhof and Meiser (2024) attempted to use an ideal-point process to characterize the effect of TOI on node 1. Because the directional decision is most ambiguous when the TOI level equals the difficulty of node 2 ($\beta_{2j} = -d_{2j}/\alpha_{2j}$), Merhof and Meiser (2024) set β_{2j} as the "location parameter" of TOI in node 1. Referring to the structures in Equations (16)-(18), the following model can be obtained:

$$Pr_{1ij} = \frac{\exp(\alpha_{1j}\eta_{1i} + \lambda_j(\theta_i - \beta_{2j}) + d_{1j}) + \exp(\alpha_{1j}\eta_{1i} + 2\lambda_j(\theta_i - \beta_{2j}) + d_{1j})}{1 + \exp(3\lambda_j(\theta_i - \beta_{2j})) + \exp(\alpha_{1j}\eta_{1i} + \lambda_j(\theta_i - \beta_{2j}) + d_{1j}) + \exp(\alpha_{1j}\eta_{1i} + 2\lambda_j(\theta_i - \beta_{2j}) + d_{1j})}, \qquad (22)$$

where η_{1i} is the MRS level, α_{1j} and λ_j are respectively the slope parameters of the MRS trait process and the TOI ideal-point process, and d_{1j} is the common intercept. Equations (19)-(22) constitute the complete IRTree model equations for five-point scoring items containing within-node multidimensional structures.

The within-node multidimensional structures with coexisting response-function processes introduced in this section may be regarded as the latest progress in refined modeling at the node level in IRTree models. However, research on these structures remains very limited. For five-point scales, there is still no systematic simulation study evaluating the parameter recovery of complex models such as those shown in Equations (19)-(22) under limited sample sizes, nor comparisons with existing simplified models. Moreover, for five-point scales in which the directional decision already adopts an ideal-point process, how to introduce within-node multidimensional structures at the compromise-response node remains a methodological gap. Overall, researchers should recognize the potential value of these frontier structures for theoretical modeling; however, given the current lack of clear model-selection criteria, sample-size requirements, software implementation guidance, and other supporting recommendations, caution is needed in practical applications.

3 Framework Evolution at the Process Level: Modeling Heterogeneity in the Response Process

This section raises the perspective from the interior of nodes to the entire response process, focusing on a more macro-level question: Do the different response reactions of different individuals, or even of the same individual, follow exactly the same psychological decision-making path? This question leads to the second main line in the evolution of the IRTree framework reviewed in this paper: heterogeneity modeling at the process level. The IRTree models discussed in the preceding sections, regardless of whether they incorporate multidimensional structure within nodes or ideal-point processes, all implicitly contain the assumption of a "homogeneous response process,"namely that all respondents follow the same decision path and differ only in their parameter levels. In practice, however, this assumption is often overly simplified. Taking a job-satisfaction survey as an example: first, different respondents may, when answering, not use RS, use only ERS, or use both ERS and MRS; this constitutes between-person heterogeneity. Second, even for the same respondent, a particular response (e.g., "somewhat agree") may be generated through different psychological pathways (the effect of ERS and TOI, or induced by ARS), reflecting heterogeneity at the response level. Furthermore, if the time dimension is added, the effect of RS may undergo systematic changes, reflecting dynamic within-person heterogeneity. These three levels of heterogeneity deepen layer by layer and jointly break through the assumption of a "homogeneous response process." The following sections review in turn progress in using the IRTree framework to model these three levels of heterogeneity.

3.1 Modeling Between-Person Heterogeneity

Common approaches to handling between-person heterogeneity are all reflected within the IRTree framework. When researchers have clear covariates (e.g., age, gender, region) or known groupings, explanatory IRTree models (Ames & Myers, 2021) or multigroup IRTree models (Plieninger, 2021) can be used to test parameter differences or explain sources of heterogeneity. At the same time, as person-centered research paradigms have become increasingly widespread, mixed IRTree models have also been developed and applied; through latent class structures, they identify subgroups in the data that have different response-strategy patterns. Within each subgroup, the model still treats RS as a continuous latent trait, but allows subgroups to differ in parameters or tree structure.

Early studies laid the foundation for the development of mixed IRTree models. Tijmstra et al. (2018) attempted to use a mixture model with one class being GPCM and another class being an IRTree structure, distinguishing two different cognitive mechanisms that respondents may follow when selecting intermediate options, and thus became a pioneer in mixed IRTree modeling. Khorramdel et al. (2019) proposed a practical three-step method for identifying ERS subgroups: first, use an IRTree model and a general IRT model to fit response data, and determine through model comparison whether ERS exists; if ERS is confirmed, then use a mixed IRT model to analyze pseudo-item data, combine multiple indices to select the optimal number of classes, and identify subgroups with different response patterns; finally, fit IRTree models separately to the response data of each identified subgroup, and test whether ERS exists only in specific classes. This method integrates the advantages of mixed IRT models and IRTree models, and can more effectively separate TOI from ERS.

Subsequently, researchers have successively developed different mixed IRTree models, assuming different IRTree structures under different latent categories. Table 3 reviews three representative studies of mixed IRTree models, from which a clear line of progression can be seen.

Kim and Bolt (2021) continued the exploration of Tijmstra et al. (2018) and took the lead in proposing a two-category mixed IRTree model, distinguishing two cognitive mechanisms behind extreme option selection and achieving separation between groups with ERS and those without ERS. Alagöz and Meiser (2024) extended the model to four categories in the context of a 5-point scale, while considering the combined use of MRS and ERS, and were able to identify multiple response-strategy types such as "ERS only," "MRS only," "both," and "neither." The study by Alagöz et al. (2025) went a step further: by leveraging the advantages of multidimensional structure within nodes, it allowed different categories to have different TOI slopes at the extreme-response nodes. This means that mixed IRTree models can not only identify whether a certain response strategy is present, but also quantify in finer detail the relative weights or dominance of different response strategies. The above studies also gradually expanded the application dimensions of the model, for example by treating latent categories as dependent variables and exploring factors that influence individuals' response-strategy membership (Alagöz et al., 2025; Khorramdel et al., 2019).

Table 3 Mixed IRTree models proposed in existing studies

Comparison dimension	Kim and Bolt (2021)	Alagöz and Meiser (2024)	Alagöz et al. (2025)
Analytic objective	Capture heterogeneity in the extreme-response process, distinguishing extreme choices driven by ERS versus TOI	Simultaneously capture individual differences in the use of MRS and ERS, and identify multiple response strategies	Capture weight differences among different response strategies driven respectively by ERS and TOI in the extreme-response process
Number and meaning of categories	Fixed 2 categories: non-ERS category (extreme response nodes are still driven by TOI) and ERS category	Fixed 4 categories: ERS only, MRS only, both ERS and MRS, neither ERS nor MRS	Classification is data-driven: allows the TOI slopes of extreme-response nodes to differ across categories
Estimation method	Bayesian MCMC	Bayesian MCMC	Three-step maximum likelihood estimation, incorporating covariates to predict category membership
Simulation-study performance	DIC can effectively select the model; classification accuracy > 90%; parameter estimation is superior to the single model	DIC can effectively select the model; classification accuracy > 90%; parameter estimation is superior to each single-response-strategy model	AIC has the highest accuracy in identifying the number of categories (> 80%); when separation is moderate or above, classification accuracy > 75%; parameter-estimation bias is small

Comparison dimension	Kim and Bolt (2021)	Alagöz and Meiser (2024)	Alagöz et al. (2025)
Main contribution	Simultaneously handles the presence/absence and intensity differences of ERS	Uses a mixed IRTree model to simultaneously examine individual differences in the use of ERS and MRS	Mixed IRTree modeling shifts from identifying different types of response strategies to identifying the allocation of weights across different response strategies

Note: DIC is the Deviance Information Criteria, and AIC is the Akaike Information Criteria.

Looking back over the developmental trajectory of the mixed IRTree model, its basic assumptions are that differences in response strategies among individuals can be represented by discrete latent categories, and that different categories internally follow different decision processes or parameter structures. Here we supplement the conditional settings for sample size and number of items in the simulation studies described above: Kim and Bolt (2021) used 1000 persons × 15 items; Alagöz and Meiser (2024) used 2000 persons × 20 items; Alagöz et al. (2025) used a full crossing of 3 sample sizes (1000, 2000, 3000) and 3 numbers of items (10, 20, 30). At present, there is still a lack of investigations targeting smaller sample sizes or shorter questionnaires; this limitation should be noted in practical applications. In terms of operational recommendations, the determination of the number of categories should be closely tied to the theoretical starting point: if there are clear theoretical expectations, one may follow Kim and Bolt (2021) or Alagöz and Meiser (2024) in prespecifying a fixed number of categories; if the theory is still unclear or one wishes to explore latent structures, one may follow Alagöz et al. (2025) in adopting a data-driven approach, using fit indices to select the number of categories and providing substantive interpretations for each retained category. In sum, the mixed IRTree model provides a flexible tool for understanding heterogeneity in response strategies among individuals, but its application should be built on adequate sample size, reasonable specification of the number of categories, and rigorous model validation.

3.2 Modeling Within-Response Heterogeneity

In the IRTree model, each response category usually corresponds to only one fixed decision path, and the outcomes at the terminal nodes of the branches in the tree diagram do not overlap. However, some studies have drawn on the ideas of the MPT model to design and examine special IRTree models that allow

Figure 3. Response-process assumptions of the 5-point-scored IRTree model incorporating ARS. The diagram contains Node 1, Node 2(b), Node 2(a), Node 3, and Node 4, with branches labeled by probabilities such as Pr_{1ij}, $1 - Pr_{1ij}$, $Pr^{(b)}_{2ij}$, $1 - Pr^{(b)}_{2ij}$, $Pr^{(a)}_{2ij}$, $1 - Pr^{(a)}_{2ij}$, Pr_{3ij}, $1 - Pr_{3ij}$, Pr_{4ij}, and $1 - Pr_{4ij}$, leading to responses $Y_{ij} = 5, 4, 3, 2, 1$.

Figure 3: Figure 3. Response-process assumptions of the 5-point-scored IRTree model incorporating ARS. The diagram contains Node 1, Node 2(b), Node 2(a), Node 3, and Node 4, with branches labeled by probabilities such as Pr_{1ij}, $1-Pr_{1ij}$, $Pr^{(b)}_{2ij}$, $1-Pr^{(b)}_{2ij}$, $Pr^{(a)}_{2ij}$, $1-Pr^{(a)}_{2ij}$, Pr_{3ij}, $1-Pr_{3ij}$, Pr_{4ij}, and $1-Pr_{4ij}$, leading to responses $Y_{ij} = 5, 4, 3, 2, 1$.

the same outcome to correspond to multiple paths (Plieninger, 2021; Plieninger & Heck, 2018). Such single-outcome, multiple-path models characterize process heterogeneity at the response level; that is, the same observed response may arise from different psychological decision processes, thereby extending the study of heterogeneity from stable individual differences to path uncertainty in a single response.

Taking the study by Plieninger and Heck (2018) as an example, the following explains how to incorporate ARS, which is often ignored by researchers, and analyzes how it jointly affects category selection together with MRS and ERS. As shown in Figure 3, this model adds a new node, "whether to endorse by default" (node 1 in Figure 3), before the basic IRTree structure, and from it derives another decision branch. Under this specification, the selection of categories "1," "2," and "3," in addition to being affected by the original person parameters, also additionally considers the effect of the respondent' s ARS level, which is reflected in the addition of a multiplicative term in the category-selection probability expression. The key point is that the selection of categories "4" and "5" has two possible paths: one is affected only by ARS and ERS, and the other passes through the complete ARS, MRS, TOI, and ERS decision processes. Therefore, the selection probabilities for these two categories are expressed as the sum of the products of the probabilities of the two influence paths:

$$P(Y_{ij} = 5) = Pr_{1ij}Pr^{(b)}_{2ij} + (1 - Pr_{1ij})(1 - Pr^{(a)}_{2ij})Pr_{3ij}Pr_{4ij}, \tag{23}$$

$$P(Y_{ij} = 4) = Pr_{1ij}(1-Pr^{(b)}_{2ij}) + (1-Pr_{1ij})(1-Pr^{(a)}_{2ij})Pr_{3ij}(1-Pr_{4ij}). \tag{24}$$

Figure 3 Response-process assumptions of the 5-point-scored IRTree model incorporating ARS

Note: Y_{ij} denotes the response of respondent i to item j; Pr_{1ij}, $Pr^{(a)}_{2ij}$, Pr_{3ij}, Pr_{4ij}, and $Pr^{(b)}_{2ij}$ respectively denote the probability of default endorsement at

Node 1 (choosing "4" or "5"), the probability of choosing the middle option "3" at Node 2(a), the probability of choosing the "agree" side at Node 3, the probability of choosing "5" or "1" rather than "4" or "2" at Node 4, and the probability of choosing "5" rather than "4" at Node 2(b).

In this model, Node 1, where ARS operates, is likewise parameterized through an IRT model. For this type of tree model that is mixed at the response level and has a probability-sum form, estimation must be carried out using Bayesian MCMC methods (Plieninger & Heck, 2018). The model can estimate each respondent' s ARS level and estimate the correlations between ARS and other person parameters. For categories "4" and "5," Pr_{1ij} reflects the uncertainty of the cognitive pathway behind them—namely, the extent to which the response stems from the automatic response of default endorsement, or is the result of integrated judgment involving TOI and multiple RS parameters. The main limitation of this model is that the scale must contain a sufficient number of positively worded and negatively worded items in order to successfully separate ARS from TOI.

3.3 Modeling Dynamic Heterogeneity Within Individuals

After systematically examining stable differences in cognitive processes across individuals and uncertainty within a single response, the research question further shifts to whether RS may undergo systematic changes within individuals as the task progresses or as time elapses. The application of IRTree models has also expanded from identifying relatively static group or individual patterns to capturing the dynamic heterogeneity of individuals' own response strategies. Modeling dynamic change not only challenges the traditional assumption that RS is stable across situations and across time, but also provides new possibilities for understanding response behavior more realistically and in greater detail.

Dynamic heterogeneity may first be manifested within a single measurement process. For example, during a questionnaire administration, when a respondent' s motivation to answer decreases as the questionnaire progresses, heuristic processing based on RS may gradually replace in-depth cognitive processing based on TOI, leading to a continuous evolution of response strategies. Merhof and Meiser (2023) proposed a dynamic IRTree model, revealing

indicates that, in long questionnaires, item position can systematically influence the discrimination parameters of TOI and ERS. This means that the same latent trait contributes differently to response decisions at different response stages. In this model, the slopes of the tested parameters at the nodes are all set as functions of item position, with the general form:

$$\alpha_j^{(p)} = \left(\gamma_1^{(p)} - \gamma_I^{(p)}\right)\left(1 - \left(\frac{j-1}{I-1}\right)^{\lambda^{(p)}}\right) + \gamma_I^{(p)}, \qquad (25)$$

where $\alpha_j^{(p)}$ denotes the slope of the tested parameter p (TOI, ERS, or MRS) on item j; $\gamma_1^{(p)}$ and $\gamma_I^{(p)}$ are the slopes on the first and last items, respectively; and $\lambda^{(p)}$ controls the shape of the trajectory. This parameterization allows the slope to change linearly or curvilinearly with item position, thereby capturing the continuous transition from in-depth cognitive processing to heuristic processing. The study further proposed an extended model incorporating random fluctuations, adding normally distributed disturbance terms with mean zero to the trajectory, so as to improve the flexibility and explanatory power of data fitting. In the empirical data, it was found that the effects of ERS and MRS both tended to strengthen gradually as item position moved later.

More explicit dynamic heterogeneity is reflected in longitudinal changes across time. Recent studies have begun to use tracking data to examine the stability and change of RS over time. For example, Ames and Leventhal (2022, 2023) proposed longitudinal IRTree models based on two-wave or multi-wave tracking data. These models assume that the discrimination and difficulty of pseudo-items do not change over time, while allowing respondent parameters and their distributional parameters to vary over time. Based on a within-node unidimensional structure, the model decomposes the respondent parameter at each decision node k at measurement time point t into the mean μ_{kt} at time point t (with the initial mean fixed at $\mu_{k1} = 0$) and the individual deviation θ_{ikt}. Then the probability of a pseudo-item response at time point t and decision node k is expressed as:

$$Pr\left(Y_{ijkt}^{*} = 1 \mid \theta_{ikt}, \mu_{kt}, a_{jk}, b_{jk}\right) = logit^{-1}\left(a_{jk}\left(\mu_{kt} + \theta_{ikt} + b_{jk}\right)\right), \qquad (26)$$

where a_{jk} and b_{jk} are, respectively, the time-invariant discrimination and easiness parameters of the pseudo-item (here following the definitions in the original literature, b_{jk} is the opposite of the traditional difficulty parameter). By analyzing mean drift and cross-time correlations, this model can clarify the change trends and stability of TOI and RS.

In addition, the IRTree framework has also been applied to data collected in intensive longitudinal studies (e.g., diary methods), and multilevel IRTree models have been developed (Hasselhorn et al., 2024). This model decomposes the variation in TOI and RS into two levels: between-person and within-person:

$$\eta_{it}^{(p)} = \mu_i^{(p)} + \varepsilon_{it}^{(p)} \qquad (27)$$

where $\eta_{it}^{(p)}$ is the level of respondent parameter p for respondent i at measurement time point t, $\mu_i^{(p)}$ is respondent i' s person-specific mean at the between-person level, and $\varepsilon_{it}^{(p)}$ is respondent i' s within-person deviation at measurement time point t. Correspondingly, the model estimates the variances and covariances of the respondent parameters separately at the between-person and within-person levels. Through multilevel decomposition, this model, for the first time in intensive longitudinal data, simultaneously

model interindividual stable tendencies and intraindividual dynamic fluctuations in TOI and RS, thereby estimating the intraindividual dynamics of TOI while controlling for potential interference from intraindividual variation in RS.

Research on the above dynamic IRTree models is still in its infancy, and empirical applications remain relatively limited. In particular, the parameter-estimation accuracy of multilevel IRTree models under different conditions has not yet been systematically evaluated through simulation studies. From the perspective of theoretical assumptions, dynamic heterogeneity within a single measurement occasion (e.g., item-position effects) only depicts fluctuations in the strength of RS influence as the task proceeds, while an individual' s RS level is still regarded as stable; by contrast, the longitudinal IRTree model of Ames and Leventhal (2022, 2023) and the multilevel IRTree model of Hasselhorn et al. (2024) break, from different angles, the traditional assumption that "RS is stable over time." Future research should clarify at the theoretical level the type of dynamics of concern and select appropriate modeling strategies accordingly. Overall, from item-position effects within a single measurement occasion, to longitudinal changes across time periods, and then to intraindividual fluctuations under intensive tracking, IRTree models continuously provide methodological support for understanding and correcting RS effects on increasingly fine-grained time scales, expanding the boundaries of interpretation of real response behavior; however, their methodological validation and applied dissemination still require further in-depth work.

4 Application of Item Response Tree Modeling Methods

4.1 Model Review

As noted above, IRTree models have developed rich model variants at both the node level and the process level. To help researchers systematically grasp the theoretical foundations, structural features, and applicable scenarios of various models, Table 4 summarizes and compares the main extended models of IRTree models.

From the perspective of implementation for parameter estimation, most model variants extended at the node level can be fitted using the frequentist methods supported by the `mirt` package, whereas process-level extensions usually require more complex estimation methods, such as Bayesian MCMC. Even at the node level, if within-node multidimensional structures are introduced, using the `mirt` package also requires researchers to define the response functions and parameter constraints of pseudo-items themselves, which is not user-friendly for applied researchers. To lower this technical threshold, this article has written a set of auxiliary R functions (see https://osf.io/bav7r). Among them, the `irtree4cat_wn_unidim` and `irtree4cat_wn_multidim` functions can quickly fit, respectively, IRTree models with a unidimensional within-node structure and a multidimensional within-node structure for 4-point scoring data; they support testing invariance constraints on the direction of extreme-response nodes (including unconstrained models and models with equal slopes; for within-node unidimensional models, equality constraints can also be imposed simultaneously on slopes and intercepts), and allow specification of the response-process model for TOI in directional decisions (dominance process or ideal-point process). Similarly, the `irtree5cat_wn_unidim` and `irtree5cat_wn_multidim` functions provide the same functionality for 5-point scoring data. With the help of these functions, researchers can focus more on theoretical assumptions and model specification, without becoming entangled in the technical details of code.

Table 4 Summary of Major Extended Models of the IRTree Framework

Model category	Core theoretical assumptions	Model-structure features	Typical application scenarios
Node-level extension	Multidimensional-structure IRTree within nodes	TOI affects all nodes, whereas RS affects only local nodes; the two coexist complementarily within nodes	Nodes for nondirectional decisions simultaneously load TOI and RS; cross-direction constraints usually need to be imposed on slopes
Node-level extension	IRTree with ideal-point process	The response mechanism of nodes (usually directional decisions) is based on matching judgments between the individual' s level and the item location	Introduces a unimodal response function at the corresponding node (e.g., a GGUM structure)

Model category	Core theoretical assumptions	Model-structure features	Typical application scenarios
Process-level extension	Mixture IRTree	There are subgroups that use response strategies; different subgroups follow different decision processes or parameter structures	Introduces latent categorical variables, allowing tree structures or within-node parameters to vary by class
Process-level extension	Single-outcome multi-path IRTree	The same observed response category can originate from multiple different psychological paths	Multiple paths point to the same outcome category, and its category probability is the sum of the probabilities of the paths
Process-level extension	Dynamic IRTree	The strength or level of RS varies systematically with item position, measurement waves, etc.	Allows model parameters (e.g., discrimination, latent-variable means) to change with position or time

4.2 Process Recommendations

Given the diversity of IRTree model variants, researchers analyzing Likert-scale data are often uncertain about "where to begin" and "how to choose." Integrating the framework developments reviewed above, this section proposes a progressive general analysis workflow for data scored on 4-point or 5-point scales, guiding researchers in building models with good interpretability. The workflow is divided into three stages. The first two stages focus on the establishment and optimization of homogeneous models (assuming that all respondents follow the same decision structure), whereas Stage 3 explores heterogeneity in the response process.

4.2.1 Stage 1: Determine the Necessity of IRTree Modeling and Specify the TOI Response Function

The core goal of Stage 1 is to determine whether IRTree modeling is necessary and to clarify whether TOI should follow a dominance process or an ideal-point process.

The response function for TOI should first be determined on theoretical grounds: if the construct has the characteristic that "higher levels are increasingly likely to agree," then

assume a dominance process; if the construct is closer to "agreeing when the person and item positions match and disagreeing when they diverge," then an ideal-point process is assumed. When the theoretical basis is insufficient, one may simultaneously fit ordinary IRT models corresponding to the two response mechanisms and make judgments by integrating convergence, fit indices, and parameter interpretability. Once the TOI response function has been determined, it is recommended that, in subsequent IRTree modeling, the directional decision in which the TOI is located should by default be kept consistent with the response function of the ordinary IRT model, whereas non-directional decisions reflecting RS effects should by default adopt a dominance process.

Subsequently, fit the corresponding ordinary IRT model without RS (denoted M_0), and then construct a unidimensional IRTree model within nodes (denoted M_1). Following the practice of Khorramdel et al. (2019), judge whether the process-decomposition modeling of the IRTree model is reasonable by comparing the fit of M_1 and M_0: if M_1 is superior to M_0, this indicates that the IRTree model fits the data better and provides support for the presence of RS in the data that can be captured by IRTree. In addition, parameter constraints under the assumption of directional invariance need to be debugged. Following the recommendations of Jeon and De Boeck (2019), one may begin with a model in which slopes and intercepts are all equal and then gradually relax constraints, determining the final form on the basis of changes in fit and parameter interpretability.

It should be emphasized that judging the necessity of IRTree modeling and determining the form of parameter constraints depend primarily on model-fit comparisons. However, the determination of the TOI response process must adhere to theoretical priority and integrate multiple factors; fit indices can serve only as one piece of reference evidence. If in Stage 1 it has been confirmed that an IRTree model is necessary, and theoretical expectations support a global influence of the TOI, then proceed to Stage 2 for further analysis.

4.2.2 Stage 2: Evaluate whether to incorporate within-node multidimensional structure The goal of this stage is, on the basis of Stage 1, to further evaluate whether the model is suitable for incorporating within-node multidimensional structure. The model in which the influence of the TOI is introduced into the extreme-response nodes is denoted M_2.

Although the assumption of a global TOI effect behind within-node multidimensional structure is usually reasonable, researchers should not rely solely on fit indices to judge whether M_2 is superior to M_1. They should also attend to the estimation and interpretability of parameters: whether M_2 converges and whether parameter estimates are reasonable (for example, whether the slope direction of the TOI on non-directional nodes conforms to theoretical expectations); and, compared with M_1, whether M_2 provides more meaningful substantive conclusions (e.g., whether the correlation between ERS and TOI decreases markedly, suggesting that some mimicry effect may have been corrected), and so forth. If M_2 is unstable in estimation or difficult to interpret, M_1 should be retained even if the fit is slightly better. In terms of parameter constraints, it is recommended first to follow convention by constraining the ERS slopes on the two extreme nodes to be equal and the TOI slopes to be additive inverses, in order to ensure model identification. Researchers may also try to relax this constraint and evaluate, through model comparison, whether slope constraints should be imposed.

In particular, for 5-point scoring data in which the TOI adopts a dominance process, if the extreme-response nodes have been extended to a multidimensional structure, one may further attempt also to incorporate the influence of the TOI at the unfolding-response node (see the preceding formula (22)), and decide whether to adopt it by comparing model interpretability and fit.

4.2.3 Stage 3: Explore heterogeneity in the response process After the optimal homogeneous model has been determined, researchers can further explore heterogeneity in the response process. Common approaches include: fitting mixed IRTtree models to identify subgroups with different response strategies, or using single-result multipath IRTtree models to analyze the mixing of ARS and traditional tree-like paths within responses. It should be noted that such heterogeneity modeling methods often face the challenges of numerous parameters and difficult estimation, and impose relatively high requirements on sample size, number of items, and model specification. It is recommended that they be attempted cautiously only when there is a clear theoretical motivation (e.g., the expected existence of a "no-RS respondent" subgroup) and sufficient sample size (generally 1000 or more respondents is recommended). Meanwhile, the homogeneous model established in stage two should always be used as the benchmark, and the necessity of the heterogeneous model should be evaluated by comparing whether it brings substantive theoretical gains and improved fit.

4.3 Application Example

4.3.1 Empirical Data and Analysis Method

This section demonstrates the IRTtree model analysis procedure using actual questionnaire data. The data come from the public dataset of the classic Big Five personality questionnaire (see https://www.kaggle.com/tunguz/big-five_personality-test). The questionnaire was administered online through the Open Psychometrics platform (https://openpsychometrics.org/tests/IPIP-BFFM/), and contains 50 5-point Likert rating items. As a mature personality measurement instrument, the Big Five personality questionnaire has been widely used in RS research; for example, Khorramdel and Von Davier (2014) analyzed ERS and MRS in responses to this questionnaire, and domestic RS research such as Guo Qingke et al. (2007) has also focused on this instrument. In the present study, data from 1000 respondents with no missing values or outliers were randomly selected as the sample; to facilitate demonstration of the IRTtree analysis procedure, only the 10 items of the Agreeableness subscale were selected, as a typical questionnaire data-analysis context with 5-point scoring and a unidimensional construct (scale content, data, and analysis code files are available at https://osf.io/bav7r). Before fitting the models, reverse-scored items were reverse-coded.

This example focuses on demonstrating stage one and stage two; stage three, the exploration of heterogeneity modeling, is not expanded upon[1]. In stage one, because the wording direction of the 10 items is relatively clear, it supports the TOI dominance process to some extent; however, considering that ideal-point process models are also commonly used in personality testing, the ideal-point process is not directly ruled out. Instead, IRT models for two response processes are fitted simultaneously (using GPCM and GGUM as example models for the dominance process and ideal-point process, respectively, without being limited to comparison of multiple IRT models under the same response process). Based on the multiple information criteria supported by the `mirt` package, together with parameter interpretability and model parsimony, model M_0 is selected. Then, with reference to the 4-node structure shown in Table 2, the response-process assumption of the IRT model is extended; a within-node unidimensional IRTtree model M_1 is built and direction-invariance constraints are determined. If M_1 is superior to M_0, stage two is entered; based on the assumption that TOI has a global influence, a multidimensional structure is introduced at the extreme-response node to obtain M_2. If, after integrating multiple factors, M_2 is selected, a multidimensional structure is further introduced at the folding-response node to obtain M_3, and similarly the model fit and parameters are integrated

the final model was selected by interpretability. In all models, the means of the latent variables were constrained to 0 and the variances to 1; for the IRTree model, the correlations between agreeableness and ERS and MRS were estimated. GPCM and GGUM used the `mirt` package to estimate model parameters; the IRTree model was analyzed using the self-written functions `irtree5cat_wn_unidim` and `irtree5cat_wn_multidim`, in which the `irtrees` package (De Boeck & Partchev, 2012) was used for pseudo-item coding, and the `mirt` package was used to estimate model parameters.

[1]The analysis-code file for this paper additionally provides example statements for establishing a single-result multipath IRTtree model, for reference by interested readers.

4.3.2 Analysis Results

Table 5 presents the fit-index results for the five models. In Stage 1, GPCM and GGUM had similar fit; only AIC favored GGUM, whereas the other information criteria, which penalize the number of parameters more heavily, all supported GPCM. Examining the parameter estimates showed that GGUM' s estimate of the location parameter for Item 3 was extremely unstable (point estimate and 95% confidence interval: $6.45\ [-1.81, 14.71]$), whereas GPCM did not show such a problem. Considering theoretical expectations, fit indices, and parameter-estimation performance together, the more concise GPCM was selected as M_0. For the within-node unidimensional IRTree model, three parameter constraints were attempted at the two extreme-response nodes: equal slopes and intercepts, equal slopes only, and unconstrained. Except for AIC, the information criteria all supported constraining only the slopes to be equal; therefore, the within-node unidimensional model with this constraint was selected as M_1. Across all information criteria examined, M_1 outperformed M_0, indicating that process-decomposition modeling with the IRTree model was more reasonable. Entering Stage 2, the "ERS+TOI" multidimensional structure was introduced at the extreme-response nodes. Because the model with freely estimated slopes failed to converge, the customary slope constraints (equal ERS slopes on both sides, and opposite TOI slopes) were used to obtain M_2. M_2 converged successfully; the direction of the slope estimates for agreeableness at the extreme-response nodes was as expected, and M_2 outperformed M_1 on all information criteria examined. After comprehensive consideration, M_2 was retained, and the influence of agreeableness was further introduced at Node 1, using the ideal-point process for representation, yielding M_3. M_3 likewise converged successfully, and there were no obvious problems with parameter estimation; from M_2 to M_3, all information criteria decreased further, and the decrease was larger than that from M_1 to M_2. Accordingly, the complete within-node multidimensional structure of M_3 can be considered more explanatory and was selected as the final model.

Table 5 Comparison of fit for different IRT or IRTree models

Model	Number of parameters	Log likelihood	AIC	BIC	SABIC	HQ
GPCM	50	−12833	25766	26012	25853	25859
GGUM	60	−12822	25764	26058	25868	25876
IRTree (M_1)	73	−12412	24969	25327	25096	25105
IRTree (M_2)	83	−12374	24913	25320	25057	25068
IRTree (M_3)	93	−12330	24845	25302	25006	25019

Note: M_1 is the IRTree model with a within-node unidimensional structure;

M_2 is the IRTree model in which a within-node multidimensional structure is introduced at the extreme-response nodes; M_3 is the IRTree model in which a within-node multidimensional structure is introduced at both the midpoint-response node and the extreme-response nodes. AIC is the Akaike Information

Criteria); BIC is the Bayesian Information Criterion; SABIC is the Sample-size Adjusted Bayesian Information Criterion; and HQ is the Hannan-Quinn Information Criterion. The above information criteria and log-likelihood are all model-evaluation indices supported as output by the mirt hybrid model.

The parameter-estimation analysis of the simulated effects can verify the rationality of the multidimensional structure within nodes. In the within-node unidimensional IRTree model, MRS and ERS, as well as MRS and acquiescence, were both significantly negatively correlated, whereas ERS and acquiescence were significantly positively correlated (the point estimates and 95% confidence intervals were, respectively: $r_{\eta_1,\eta_2} = -.62[-.70, -.54]$, $r_{\eta_1,\theta} = -.54[-.62, -.46]$, $r_{\eta_2,\theta} = .32[.23, .41]$). After the extreme response node was expanded into a multidimensional structure, the estimated correlation between ERS and acquiescence decreased ($r_{\eta_1,\eta_2} = -.51[-.61, -.41]$, $r_{\eta_1,\theta} = -.55[-.63, -.47]$, $r_{\eta_2,\theta} = .16[.04, .27]$). This pattern is consistent with the theoretical expectation for simulation effects (Merhof et al., 2024): that is, under a within-node unidimensional specification, the ERS node may contain variance that is partly explained by the acquiescence trait; after a multidimensional structure is introduced, this part of the variance is redistributed. In other words, changes in the estimated cross-model correlation coefficients reflect differences in the parameter-estimation patterns under different model specifications. After the middle-response node was further expanded into a multidimensional structure, the estimated correlation between MRS and acquiescence also became nonsignificant ($r_{\eta_1,\eta_2} = -.53[-.64, -.42]$, $r_{\eta_1,\theta} = -.07[-.32, .18]$, $r_{\eta_2,\theta} = .12[.01, .24]$). This trend suggests that simulation effects may also exist in the measurement of MRS. Within the range of candidate models examined in this paper, M_3 showed a relatively optimal fit, providing reference estimates for the relationships between acquiescence and the two types of RS: MRS was not significantly correlated with acquiescence, whereas ERS had a significant but relatively weak positive correlation with acquiescence.

The changes in the slope estimates for ERS and MRS also deserve attention. When the RS variance was fixed at 1, the slope measured the contribution of RS to local decisions on the test items. According to the theory of simulation effects, when the acquiescence trait is included in the model as a global factor, if the slope estimate of RS on an item decreases, this suggests that the original unidimensional model may have mixed into the estimation of that slope some variance explained by acquiescence. As shown in Figure 4, when the extreme-response node was changed from a unidimensional to a multidimensional structure, the slope estimates for MRS were basically stable, whereas the slope estimates for ERS decreased on some items (e.g., Items 1, 4, and 7). When the middle-response node was also changed to a multidimensional struc-

Figure 4. Slope estimates of RS on each item under the three IRTree models.

Figure 4: Figure 4. Slope estimates of RS on each item under the three IRTree models.

Scatterplot with x-axis "initial trait estimates" and y-axis "corrected trait estimates" ; legend: high ERS, high MRS, other.

Figure 5: Scatterplot with x-axis "initial trait estimates" and y-axis "corrected trait estimates" ; legend: high ERS, high MRS, other.

ture, the slope estimates for MRS decreased on more items (e.g., Items 1, 2, 4, 5, 7, and 8), whereas the slope estimates for ERS generally changed very little. The above comparison shows that, in the present sample data, under the within-node unidimensional model, the MRS node may be more seriously influenced by the acquiescence trait than the ERS node, thereby suggesting the potential significance of introducing a multidimensional structure in the middle-response node as well. Based on the slope estimates of the final model M_3, the degree to which each item was influenced by different RS can also be identified: Item 4 was most strongly influenced by ERS; Item 8 was most strongly influenced by MRS; whereas Item 3 was least influenced by both types of RS.

Figure 4 Slope estimates of RS on each item under the three IRTree models

Note: The error bars indicate 95% confidence intervals.

Figure 5 compares the estimates of the GPCM person trait parameters between the final model M_3 and the model ignoring RS, and separately colors examinees with high ERS and high MRS (both using the lenient criterion that the estimate is greater than 1; there were no examinees for whom the levels of both types of RS were on average higher than 1). As can be seen from the figure, the trait-parameter estimates are distributed around the diagonal, showing the characteristics of being dense in the middle and wide at both ends; that is, examinees whose trait levels lie at the two ends are corrected by the model to a greater extent. Specifically, for examinees with high MRS, those with higher initial trait estimates are further raised, whereas those with lower initial estimates are further lowered; conversely, for examinees with high ERS, those with lower initial estimates are corrected upward, whereas those with higher initial estimates are corrected downward. The critical points at which the correction directions change for the two types of examinees are both at a person trait level of approximately -0.9, and the magnitude of correction is smallest near this position.

Figure 5 Correction of GPCM latent-trait estimates based on the IRTree model

Note: The corrected trait estimates come from the model $D_{\eta_1}I_\theta - D_\theta - D_{\eta_2}D_\theta$. High ERS (MRS) adopts the lenient criterion that the ERS (MRS) estimate is higher than 1; there are no respondents whose two RS estimates are both higher than 1.

5 Discussion

5.1 Two Perspectives on Modeling Response Styles

This article focuses on reviewing the evolution of the IRTree framework at the node level and the process level, revealing the potential value of the IRTree model in RS research: namely, that it can, to a certain extent, simultaneously serve analyses from the two perspectives of measurement-bias control and response-process depiction. This gives IRTree distinctive advantages relative to traditional bias-control methods.

From the perspective of process understanding, the evolution of the IRTree framework provides an increasingly refined analytical tool for opening the black box of "how respondents answer." Bolt and Meng (2025) regard the IRTree model as a core tool serving the goal of "understanding psychological processes," arguing that its advantage lies in the clear correspondence between parameters and cognitive steps. Early IRTree models used tree-like

The structure decomposes ordered responding into several binary decision nodes, enabling researchers to enter into process-level inquiries into "how the respondent answers." Expansions at the node level—whether the introduction of multidimensional structures within nodes or the integration of ideal-point processes—are, in essence, refinements of assumptions about the psychological mechanisms behind individual decision nodes. Does TOI exert a global influence on all decisions? When judging agreement, is the respondent relying on level or on positional matching? These theoretical questions prompt researchers to confront, to an increasing extent, the cognitive logic of the response process. On the other hand, process-level extensions such as hybrid IRTree, multiple-path IRTree with a single outcome, and dynamic IRTree further help researchers examine in greater detail whether the entire response process is heterogeneous and how it can be better modeled. Together, these advances have pushed RS research from merely identifying whether bias exists toward a deeper understanding of how bias emerges and changes.

At the same time, from the perspective of bias control, the above developments have likewise brought about substantive improvements in measurement precision. A unidimensional IRTree model within nodes has already preliminarily achieved statistical separation of RS, while the introduction of multidimensional structures within nodes further reduces the risk of confounding and, to a certain extent, avoids possible bias in the estimation of respondent-parameter variances and covariances. In addition, combining ideal-point processes with IRTree can, while controlling RS, more comprehensively characterize the response mechanism of constructs, thereby improving the accuracy of TOI estimation. Hybrid modeling and dynamic modeling at the process level, by incorporating heterogeneity into the model structure, avoid estimation errors caused by aggregating different response strategies or responses at different time points. The fine-grained depiction of process-mechanism assumptions in the IRTree framework makes the model more closely approximate the actual response process and often yields a purer bias-separation effect.

It should be emphasized that although IRTree models can take both perspectives into account, their methodological core has always prioritized understanding cognitive processes. The evolution of the IRTree framework reflects the deepening trend in RS research: from the early view of RS as merely noise that needed to be removed, gradually shifting toward also regarding it as a component of cognitive strategy that needs to be understood. This deepening does not negate the necessity of bias control; rather, it advocates control on the basis of process understanding. Even in applications where bias control is the direct goal, the correction of trait estimates by IRTree must be grounded in explicit theoretical commitments about the response process—such as which nodes are affected by RS and whether TOI and RS jointly influence a node. The progressive analytical procedure proposed in this article is precisely based on this understanding.

By contrast, in recent years some new methods have also emerged with bias control as their core objective. For example, Grimmond et al. (2025), drawing on the idea of signal detection theory, defined RS operation as an individual's response threshold that remains stable across dimensions, thereby separating RS from the construct. Schoenmakers et al. (2025) introduced posterior predictive checks to indirectly identify ERS by detecting model misfit at the group and individual levels associated with ERS. These methods are more concise and direct in controlling bias, but the model specifications themselves do not intentionally reveal the cognitive processing that underlies RS.

The two lines of thought above serve different research objectives: if the research focus is on accurately measuring TOI, concise bias-control methods may be prioritized; if the aim is to explore the psychological process of responding or to test the substantive relationship between RS and TOI—such as group differences in strategies in cross-cultural research—then the process interpretability of IRTree is difficult to replace. In sum, researchers need to choose according to their own research objectives,

make prudent choices between the two perspectives. Especially for the development and validation of indigenous Chinese scales, this dual perspective has practical significance: researchers must attend to RS as a potential threat to scale validity in the form of response bias, while also treating cautiously the substantive psychological information that RS may carry in the Chinese cultural context. The IRTree model provides a practical analytic framework that accommodates both concerns.

5.2 Unresolved Issues in Modeling Response Styles with Item Response Trees

Although the IRTree model has made considerable progress at both theoretical and applied levels, existing research still contains several important but insufficiently explored issues that deserve focused attention from researchers.

First, existing studies have mainly focused on 4-point and 5-point scales, and the application of the IRTree model to scales with more response categories involves evident modeling branches. Although 4-point and 5-point scales are the most common in empirical research, 6-point and 7-point scales are also widely used (Böckenholt, 2017; Henninger et al., 2025). A 6-point scale lacks an explicit neutral category, which makes process assumptions related to MRS inherently ambiguous; when pseudo-items are coded for a 7-point scale, one must also face the choice of whether extreme responding should correspond only to the two endpoints ("1" and "7") or should include the next-to-extreme categories ("1," "2," and "6," "7"). More importantly, the intensity decision process in 6-point and 7-point scales is difficult to describe adequately with a single dichotomous node; thus, whether to use multiple dichotomous nodes or introduce polytomous scoring nodes has become a central controversy. If a sequence of dichotomous nodes is retained, different respondent parameters must be specified for different stages of intensity decisions; if polytomous scoring nodes are introduced (e.g., trichotomous pseudo-items coded as 0, 1, and 2), then polytomous IRT models must be used for parameterization. The study by Meiser et al. (2019) shows that these two modeling approaches are not equivalent and represent quite different theoretical perspectives, but both are further extensions of the dichotomous extreme-response process in existing IRTree models. At present, there is still no general recommendation for an IRTree application procedure for scales with more response categories. A relatively ideal approach is to fit different IRTree models and compare model fit and parameter interpretability. However, compared with 4-point and 5-point scales, there are more candidate models to compare and the process is more cumbersome; therefore, future research needs to provide clearer methodological guidance.

Second, even for the relatively mature 5-point scale, there remain issues worthy of deeper exploration. For example, for the IRTree model involving multidimensional structures within integrated nodes, the precision of parameter estimation has not yet been systematically evaluated through simulation studies; the respective advantageous conditions of Bayesian and frequentist estimation methods also lack clear conclusions; and the minimum sample size required to ensure parameter-estimation precision is still unclear. Increased model complexity also makes parameter estimation more time-consuming: in the example in Section 4.3, the estimation times for M_1, M_2, and M_3 on the same computer[3] were 127.55 seconds, 743.58 seconds, and 1961.69 seconds, respectively. For M_3, because its theoretical assumptions are more refined (the global influence of acquiescent personality traits), model fit is also markedly improved, and the parameter estimates are reasonable (consistent with

[3] The operating system was Windows 11 (version 25H2), the processor was AMD Ryzen 7 H 260 w/ Radeon 780M Graphics (3.80 GHz), memory was 32.0 GB RAM, and the analytic software was R 4.4.0 (x64).

the theoretical expectation of the simulation effect); therefore, the overall benefit exceeds the time cost of a single fit. However, for applied researchers who emphasize rapid analysis, or for methodological researchers who need a large number of repeated simulations, the time cost caused by complexity remains a major obstacle to application. In this regard, on the one hand, optimization of estimation tools needs to be considered; on the other hand, as in Alagöz and Meiser (2024) or Plieninger (2021), one may also try to explore new tree structures and coding methods. These new models have received less attention, but may have stronger potential for extension. For example, Alagöz and Meiser (2024) proposed another within-node multidimensional structure for 5-point scales, which does not require using an ideal-point process to depict the TOI on the folded response node and may have higher estimation efficiency. These models still await future research for comparison and evaluation, and their applicability in specific contexts needs to be explored.

Finally, within the IRTree framework, parameter equivalence presents a unique complexity, and the scope of investigation may extend from within the tree structure to cross-group contexts. The preceding discussion has noted that the IRTree model faces a distinctive issue of directional invariance when measuring ERS, and there is currently no consensus on how to constrain the parameters on the two polar extreme-response nodes. Broader parameter equivalence also involves the systematic influence of item characteristics (such as keying direction) on the parameters of each decision node. For example, Böckenholt (2019) found that item keying direction affects the parameters of the directional decision node but does not involve the folded and extreme-response nodes. This indicates that item characteristics may undermine the equivalence of IRTree parameters across different conditions. Furthermore, when research involves multi-group comparisons (such as gender, culture, and time points), the issue of parameter equivalence naturally extends to the detection of differential item functioning (DIF). Within the IRTree framework, the sources of DIF are more diverse, and DIF at different nodes may correspond to quite different psychological mechanisms. For example, when common instruments such as Big Five personality questionnaires are administered to groups from Chinese and Anglo-American Western cultural backgrounds, DIF at the directional decision node reflects differences in content understanding between Chinese and Western samples, whereas DIF at the extreme-response node reflects differences in the use of ERS between Chinese and Western samples. Jin et al. (2019) were the first to try applying logistic regression and the odds-ratio method to DIF detection for pseudo-items, and found that the odds-ratio method performed more robustly at nodes containing missing data. However, that study focused only on the detection of uniform DIF and did not involve more complex DIF patterns (such as non-uniform DIF) or DIF detection in multidimensional node contexts. Future research may attempt to develop DIF detection methods suitable for IRTree models and provide operational procedural recommendations for applied studies, thereby helping researchers more precisely locate the sources of parameter non-equivalence.

5.3 Prospects for Application Trends of Item Response Trees: Expansion and Potential Integration Directions

With the evolution of IRTree models, their application scenarios are also expanding from traditional RS measurement to the broader domain of response behavior. Among them, missing-data modeling and careless-responding modeling have become two active research directions and show potential for integration with RS modeling.

In the handling of missing data, the IRTree model provides a new analytic approach. Debeer et al. (2017) first applied it to modeling omission behavior in cognitive tests, finding that under a non-random missingness mechanism, ignoring omission behavior leads to biased item-difficulty parameters, whereas IRTree can obtain unbiased estimates. Soğuksu and Demir (2025) systematically compared IRTree with expected maxim⋯

The performance of the maximization algorithm and multiple imputation was compared, and it was found that under conditions of random and nonrandom missingness, IRTree estimates were more accurate than those obtained with traditional methods. Moreover, the relationship between the estimated omission tendency and ability provided a basis for diagnosing the mechanism of data missingness.

In the modeling of non-effortful responding, Leventhal and Pastor (2024) proposed an IRTree model for rapid-guessing behavior in low-stakes tests that includes two nodes: "whether the response is non-effortful"and "whether the answer is correct." This model allows estimation of the association between non-effortful responding tendency and ability, and does not require deleting examinees who respond non-effortfully. Empirical analyses showed that non-effortful responding tendency was moderately negatively correlated with self-reported effort scores, providing preliminary evidence for the validity of the method. Leventhal and Hunsberger (2025) further verified the robustness of this model through simulation studies. The flexibility of this approach lies in the fact that, in the first stage, only a dichotomous indicator of non-effortful responding is required; this means that it can accommodate diverse definitions of non-effortful responding across different testing contexts, whether the indicator is based on response-time thresholds, confidence ratings, click behavior, or eye-tracking data.

The above extensions suggest that, in theory, the IRTree model has the potential to simultaneously model multiple sources of bias in test or questionnaire responses. For example, some studies have begun to combine IRTree with cognitive diagnostic models, using a tree structure to simultaneously model missing responses and random-guessing behavior in cognitive diagnostic tests (Li Manqi et al., 2026). These modeling ideas and RS modeling share essentially the same methodology: response behavior is decomposed into interpretable decision processes, and the examinee parameters behind different processes, as well as their interrelations, are estimated. Accordingly, one direction worth exploring is to attempt to incorporate missing-data nodes, non-effortful-responding nodes, and RS nodes jointly into a tree structure, thereby constructing an IRTree framework capable of simultaneously modeling multiple biases. This idea has not yet been discussed in the literature, and its feasibility and conditions of applicability still require subsequent empirical testing. If this direction can be advanced, it will help to more comprehensively correct trait estimates and examine the interrelations among different biases, providing more clues for understanding the cognitive processes involved in questionnaire responding. In addition, incorporating process data (such as response time) may further improve the accuracy of modeling the response process; if cross-sectional data are extended to longitudinal designs, the dynamic changes in these response processes can also be explored. Overall, the IRTree model is evolving from a specialized tool focused on RS measurement into a flexible framework capable of handling multiple response processes, which also confirms its potential for further development as a core tool for “understanding psychological processes.”

References

Deng Wen-gen, Dai Haiqi, & Dai Huiqun. (2010). A review of item response process and unfolding model research in personality measurement. *Advances in Psychological Science, 18*(12), 1986–1990.

Guo Qingke, Wang Zhao, Han Dan, & Shi Kan. (2007). Response styles among Chinese high school students and their effects. *Acta Psychologica Sinica, 39*(1), 176–183.

Li Manqi, Peng Siji, Wang Qin, & Cai Yan. (2026). Development and application of a cognitive diagnosis model for handling missing responses and random guessing. *Psychological Science, 49*(1), 207–224.

Liu Yuanyuan, Ding Yi, Peng Kaiping, & Hu Chuanpeng. (2019). Application of the polytomous additive tree model in social psychology. *Psychological Science, 42*(2), 422–

429.

Zhang, Y.-B., & Wang, H.-H. (2019). Measurement and statistical control of response styles. *Psychological Science*, 42(3), 747–754.

Adams, D. J., Bolt, D. M., Deng, S., Smith, S. S., & Baker, T. B. (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of Mathematical and Statistical Psychology*, 72(3), 466–485. https://doi.org/10.1111/bmsp.12169

Alagöz, Ö. E. C., & Meiser, T. (2024). Investigating heterogeneity in response strategies: A mixture multidimensional IRTree approach. *Educational and Psychological Measurement*, 84(5), 957-993. https://doi.org/10.1177/00131644231206765

Alagöz, Ö. E. C., Meiser, T., & Khorramdel, L. (2025). Disentangling individual differences in cognitive response mechanisms for rating scale items: A flexible-mixture multidimensional IRTree approach. *Behavior Research Methods*, 57(9), 256. https://doi.org/10.3758/s13428-025-02778-0

Alarcon, G. M., Lee, M. A., & Johnson, D. (2023). A Monte Carlo study of IRTree models' ability to recover item parameters. *Frontiers in Psychology*, 14, 1003756. https://doi.org/10.3389/fpsyg.2023.1003756

Ames, A. J., & Leventhal, B. C. (2022). Modeling changes in response style with longitudinal IRTree models. *Multivariate Behavioral Research*, 57(5), 859–878. https://doi.org/10.1080/00273171.2021.1920361

Ames, A. J., & Leventhal, B. C. (2023). Application of a longitudinal IRTree model: Response style changes over time. *Assessment*, 30(2), 332–347. https://doi.org/10.1177/10731911211042932

Ames, A. J., & Myers, A. J. (2021). Explaining variability in response style traits: A covariate-adjusted IRTree. *Educational and Psychological Measurement*, 81(4), 756–780. https://doi.org/10.1177/0013164420969780

Arias, V. B., Ponce, F. P., Garrido, L. E., Nieto-Cañaveras, M. D., Martínez-Molina, A., & Arias, B. (2023). Detecting non-content-based response styles in survey data: An application of mixture factor analysis. *Behavior Research Methods*, 56(4), 3242–3258. https://doi.org/10.3758/s13428-023-02308-w

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(6), 1235–1245. https://doi.org/10.1016/j.paid.2005.10.018

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods, 17*(4), 665–678. https://doi.org/10.1037/a0028111

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods, 22*(1), 69–83. https://doi.org/10.1037/met0000106

Böckenholt, U. (2019). Assessing item-feature effects with item response tree models. *British Journal of Mathematical and Statistical Psychology, 72*(3), 486–

500. https://doi.org/10.1111/bmsp.12163

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*(1), 159-181. https://doi.org/10.1111/bmsp.12086

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*(5), 335-352. https://doi.org/10.1177/0146621608329891

Bolt, D. M., & Meng, L. (2025). IRT-based response style models and related methodology: Review and commentary. *British Journal of Mathematical and Statistical Psychology.* Advance online publication. https://doi.org/10.1111/bmsp.70006

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212. https://doi.org/10.1177/0022022100031002003

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

Debeer, D., Janssen, R., & De Boeck, P. (2017). Modeling skipped and not-reached items using IRTrees. *Journal of Educational Measurement, 54*(3), 333-363. https://doi.org/10.1111/jedm.12147

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*(4), 465-476. https://doi.org/10.1111/j.1754-9434.2010.01273.x

Grimmond, J., Brown, S. D., & Hawkins, G. E. (2025). A solution to the pervasive problem of response bias in self-reports. *Proceedings of the National Academy of Sciences, 122*(3), e2412807122.

https://doi.org/10.1073/pnas.2412807122

Guo, T., & Spina, R. (2019). Cross-Cultural Variations in Extreme Rejecting and Extreme Affirming Response Styles. *Journal of Cross-Cultural Psychology, 50*(8), 955-971. https://doi.org/10.1177/0022022119873072

Harzing, A. W., Köster, K., & Zhao, S. (2012). Response style differences in cross-national research. *Management International Review, 52*, 341-363. https://doi.org/10.1007/s11575-011-0111-2

Hasselhorn, K., Ottenstein, C., Meiser, T., & Lischetzke, T. (2024). The effects of questionnaire length on the relative impact of response styles in ambulatory assessment. *Multivariate Behavioral Research*, 1-15. https://doi.org/10.1080/00273171.2024.2354233

Henninger, M., & Meiser, T. (2020). Different approaches to modeling response styles in divide-by-total item response theory models (part 1): A model integration. *Psychological Methods, 25*(5), 560-576. https://doi.org/10.1037/met0000249

Henninger, M., Vanhasbroeck, N., & Tuerlinckx, F. (2025). Affect dynamics or response bias? The relationship between extreme response style and affect dynamics in a controlled experiment. *Psychological Assessment, 37*(11), 639-655. https://doi.org/10.1037/pas0001370

Huang, H.-Y. (2016). Mixture random-effect IRT models for controlling extreme response style on rating scales. *Frontiers in Psychology, 7*, Article 1706. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01706

Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods, 48*(3), 1070-1085. https://doi.org/10.3758/s13428-015-0631-y

Jeon, M., & De Boeck, P. (2019). Evaluation on types of invariance in studying extreme response bias with an IRTree approach. *British Journal of Mathematical and Statistical Psychology, 72*(3), 517-537. https://doi.org/10.1111/bmsp.12182

Jin, K.-Y., Wu, Y.-J., & Chen, H.-F. (2019). Adopting the multi-process approach to detect differential item functioning in Likert scales. *Quantitative Psychology, Springer Proceedings in Mathematics & Statistics, 265*, 307-317. https://doi.org/10.1007/978-3-030-01310-3_27

Jin, K.-Y., Wu, Y.-J., & Chen, H.-F. (2022). A new multiprocess IRT model with ideal points for Likert-type items. *Journal of Educational and Behavioral Statistics, 47*(3), 297-321. https://doi.org/10.3102/10769986211057160

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*(2), 161-177. https://doi.org/10.1080/00273171.2013.866536

Khorramdel, L., Von Davier, M., & Pokropek, A. (2019). Combining mixture distribution and multidimensional IRTree models for the measurement of extreme response styles. *British Journal of Mathematical and Statistical Psychology, 72*(3), 538-559. https://doi.org/10.1111/bmsp.12179

Kim, N., & Bolt, D. M. (2021). A mixture IRTree model for extreme response style: Accounting for response process uncertainty. *Educational and Psychological Measurement, 81*(1), 131-154. https://doi.org/10.1177/0013164420913915

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Leventhal, B. C., & Pastor, D. (2024). An illustration of an IRTree model for disengagement. *Educational and Psychological Measurement, 84*(4), 810–834. https://doi.org/10.1177/00131644231185533

Leventhal, B.C., & Hunsberger, J. (2025). Branching out: How the IRTree model can root out disengagement in a variety of low-stakes contexts. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 16*(3), 179–202. https://doi.org/10.21031/epod.1662875

Li, Z. (2022). *Filling gaps in the research on IRTree theoretically and practically: A comprehensive taxonomy and a user-friendly R package* [Unpublished doctoral dissertation]. The Ohio State University.

Li, Z., Li, L., Zhang, B., Cao, M., & Tay, L. (2025). Killing two birds with one stone: Accounting for unfolding item response process and response styles using unfolding item response tree models. *Multivariate Behavioral Research, 60*(2), 161–183. https://doi.org/10.1080/00273171.2024.2394607

Liu, C. W., & Wang, W. C. (2016). Unfolding IRT models for Likert-type items with a don' t know option. *Applied Psychological Measurement, 40*(7), 517–533. https://doi.org/10.1177/0146621616664047

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*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

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*(3), 501–516. https://doi.org/10.1111/bmsp.12158

Merhof, V., Böhm, C. M., & Meiser, T. (2024). Separation of traits and extreme response style in IRTree models:

The role of mimicry effects for the meaningful interpretation of estimates. *Educational and Psychological Measurement, 84*(5), 927–956. https://doi.org/10.1177/00131644231213319

Merhof, V., & Meiser, T. (2023). Dynamic response strategies: Accounting for response process heterogeneity in IRTree decision nodes. *Psychometrika, 88*(4), 1354–1380. https://doi.org/10.1007/s11336-023-09901-0

Merhof, V., & Meiser, T. (2024). Co-occurring dominance and ideal point processes: A general IRTree framework for multidimensional item responding. *Behavior Research Methods, 56*(7), 7005–7025. https://doi.org/10.3758/s13428-024-02405-4

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012). Response strategies and response styles in cross-cultural surveys. *Cross-Cultural Research, 46*(3), 255–279. https://doi.org/10.1177/1069397112440939

Muthén, L. K., & Muthén, B. O. (1998–2022). *Mplus user' s guide* (8th ed.). Muthén & Muthén.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Plieninger, H. (2021). Developing and applying IR-tree models: Guidelines, caveats, and an extension to multiple groups. *Organizational Research Methods, 24*(3), 654–670. https://doi.org/10.1177/1094428120911096

Plieninger, H., & Heck, D. W. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research, 53*(5), 633–654. https://doi.org/10.1080/00273171.2018.1469966

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3–32. https://doi.org/10.1177/01466216000241001

Schoenmakers, M., Tijmstra, J., Vermunt, J., & Bolsinova, M. (2024). Correcting for extreme response style: Model choice matters. *Educational and Psychological Measurement, 84*(1), 145–170. https://doi.org/10.1177/00131644231155838

Schoenmakers, M., Tijmstra, J., Vermunt, J., & Bolsinova, M. (2025). Posterior predictive checks for the detection of extreme response style. *Behavior Research Methods, 57*(9), 234. https://doi.org/10.3758/s13428-025-02756-

Soğuksu, Y. B., & Demir, E. (2025). The effect of modeling missing data with IRTree approach on parameter estimates under different simulation conditions. *Educational and Psychological Measurement, 85*(3), 507–526. https://doi.org/10.1177/00131644241306024

Thissen-Roe, A., & Thissen, D. (2013). A two-decision model for responses to likert-type items. *Journal of Educational and Behavioral Statistics, 38*(5), 522–547. https://doi.org/10.3102/1076998613481500

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*(6), 2325–2344. https://doi.org/10.3758/s13428-017-0997-0

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*(2), 195–217. https://doi.org/10.1093/ijpor/eds021

Xue, M., & Chen, Y. (2023). A Stan tutorial on Bayesian IRTree models: Conventional models and explanatory extension. *Behavior Research Methods. 56*, 1817–1837. https://doi.org/10.3758/s13428-023-02121-5

Zeng, B., Wen, H., & Jeon, M. (2025). Ideal point or dominance process? Unfolding tree approaches to Likert scale data with multi-process models. *Multivariate Behavioral Research, 60*(5), 898–929. https://doi.org/10.1080/00273171.2025.2496505

Zhang, Y., Yang, Z., & Wang, Y. (2022). The impact of extreme response style on the mean comparison of two independent samples. *Sage Open, 12*(2), 21582440221108168. https://doi.org/10.1177/21582440221108168

Żółtak, T., Pokropek, A., & Muszyński, M. (2025). Response styles and omitted variable bias. Detection and mitigation. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–10. https://doi.org/10.1080/10705511.2025.2555203

The evolution of item response tree modeling for response styles:

Framework advancements and applications

DONG Jieyuan, LIU Hongyun

(Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Faculty of Psychology, Beijing Normal University, Beijing 100875, China)

Abstract: Item response tree (IRTree) models offer a refined framework for disentangling and understanding response styles by decomposing Likert-type item responses into multiple decision nodes. The ongoing evolution of the IRTree framework in recent years has greatly advanced response style research. At the node level, the IRTree model has been extended from within-node unidimensional structures to within-node multidimensional structures, and from assuming dominance processes at all nodes to incorporating ideal point processes, thereby providing a richer representation of the theoretical mechanisms underlying item responding. At the response process level, a growing body of research has explored heterogeneity in the response process, including mixture modeling across respondents, modeling of within-response pathway uncertainty, and intraindividual dynamic modeling, which enables increasingly nuanced investigations. Building on a review of these key advances, this article further summarizes a stepwise analytical workflow for IRTree modeling, and provides supporting R functions along with an empirical demonstration. Finally, this article discusses two perspectives on response style modeling—controlling for measurement bias and gaining insight into the response process—and outlines promising directions for future development of IRTree models.

Keywords: response styles, item response tree, Likert scale, item response theory

Note: Figure translations are in progress. See original paper for figures.