Received: 9 October 2024 | Accepted: 8 April 2025

DOI: 10.1111/bmsp.12393

the british psychological society
promoting excellence in psychology

ARTICLE

A general diagnostic modelling framework for forced-choice assessments

Pablo Nájera[1] | **Rodrigo S. Kreitchmann**[2] | **Scarlett Escudero**[3] | **Francisco J. Abad**[3] | **Jimmy de la Torre**[4] | **Miguel A. Sorrel**[3]

[1]Department of Psychology, UNINPSI, Universidad Pontificia Comillas, Madrid, Spain

[2]Department of Methodology in Behavioral Sciences, Faculty of Psychology, Universidad Nacional de Educación a Distancia, Madrid, Spain

[3]Department of Social Psychology and Methodology, Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain

[4]Faculty of Education, The University of Hong Kong, Hong Kong City, Hong Kong

Correspondence
Miguel A. Sorrel, Department of Social Psychology and Methodology, Faculty of Psychology, Universidad Autónoma de Madrid, 6 Iván Pavlov St, Cantoblanco Campus, Madrid 28049, Spain.
Email: miguel.sorrel@uam.es

Funding information
Universidad Pontificia Comillas (2024 Call for Funding of Internal Research Projects: “Advancements in cognitive diagnosis models for formative assessments”); MICIU/AEI/10.13039/501100011033 and ERDF/EU under the project “Computerized adaptive tests based on new assessment formats” (reference: PID2022-137258NB-I00); UAM-IIC Chair Psychometric Models and Applications

Abstract
Diagnostic classification modelling (DCM) is a family of restricted latent class models often used in educational settings to assess students' strengths and weaknesses. Recently, there has been growing interest in applying DCM to noncognitive traits in fields such as clinical and organizational psychology, as well as personality profiling. To address common response biases in these assessments, such as social desirability, Huang (2023, *Educational and Psychological Measurement*, *83*, 146) adopted the forced-choice (FC) item format within the DCM framework, developing the FC-DCM. This model assumes that examinees with no clear preference for any statements in an FC block will choose completely at random. Additionally, the unique parametrization of the FC-DCM poses challenges for integration with established DCM frameworks in the literature. In the present study, we enhance the capabilities of DCM by introducing a general diagnostic framework for FC assessments. We present an adaptation of the G-DINA model to accommodate FC responses. Simulation results show that the G-DINA model provides accurate classifications, item parameter estimates and attribute correlations, outperforming the FC-DCM in realistic scenarios where item discrimination varies. A real FC assessment example further illustrates the better model fit of the G-DINA. Practical recommendations for using the FC format in diagnostic assessments of noncognitive traits are provided.

KEYWORDS
diagnostic classification, forced-choice assessments, latent class, noncognitive traits

1 | INTRODUCTION

Diagnostic classification modelling (DCM), also known as cognitive diagnosis modelling (CDM), is a family of restricted latent class models that classify respondents based on their levels in a series of discrete latent variables referred to as *attributes*. In the educational context, where most DCM research has been conducted, attributes are usually regarded as dichotomous; thus, examinees are classified into latent classes based on their mastery or non-mastery of these attributes (e.g., skills, competences, cognitive processes). The focus of DCM research on education arises from how well such diagnostic classifications align with formative assessments, as they help teachers design remedial instruction based on the identification of examinees' strengths and weaknesses (de la Torre & Minchen, 2014; Paulsen & Valdivia, 2021).

Despite the prominence of educational research in the field of DCM (for a literature review of current empirical applications, see Sessoms & Henson, 2018), the potential of these models transcends this context of application. Recently, there has been an increase in the number of studies that employ the DCM framework to facilitate diagnostic assessments in the areas of clinical psychology (de la Torre et al., 2015; Peng et al., 2019; Tan et al., 2023; Templin & Henson, 2006; Xi et al., 2019; Zhang et al., 2024), personality assessment (Liu & Shi, 2020; Revuelta et al., 2018) and organizational psychology (García et al., 2014; Sorrel et al., 2016). The application of DCM to these contexts opens the door to new or refined types of assessments, such as diagnostic clinical evaluations according to the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013), where symptom comorbidity and the interaction among disorders are properly captured and addressed (de la Torre et al., 2015; Templin & Henson, 2006), theoretically guided personality profiling for forensic purposes (Revuelta et al., 2018) and the construction and evaluation of situational judgement tests in staff selection or promotion processes (Sorrel et al., 2016).

These noncognitive assessments, however, are more susceptible to response biases that introduce nuisance variance to the constructs being measured. Especially in high-stakes evaluations, applicants may be motivated to exhibit socially desirable responses or may be affected by other response biases, such as acquiescence or extremity (Christiansen et al., 2005; Paulhus, 1991). Among the different procedures to mitigate the impact of such distorted responses (e.g., modelling response biases; Ferrando et al., 2009), using a forced-choice (FC) item format is an effective way of mitigating the pernicious effects of social desirability, acquiescence and extremity (Brown, 2016; Cheung & Chan, 2002; Christiansen et al., 2005; Kreitchmann et al., 2019). In the FC format, items (stimuli) with similar social desirability are paired in blocks, so that respondents must choose the item that best describes them.[1]

Recently, Huang (2023) proposed the first DCM specifically tailored to FC assessments. This model, referred to as the *forced-choice diagnostic classification model* (FC-DCM), marks a milestone in the diagnostic evaluation of noncognitive traits. Despite the significance of this methodological development, the FC-DCM is not without relevant theoretical and practical limitations. These limitations, which will be elaborated on below, include a restricted item response function and the inaccessibility of crucial additional features and analyses, such as classification accuracy assessment or model fit evaluation.

The main goal of this study is to enhance diagnostic assessments of noncognitive traits by introducing a general DCM framework for FC blocks. Namely, our proposal consists of integrating FC blocks into the *generalized deterministic input, noisy 'and' gate* (G-DINA) model (de la Torre, 2011), which addresses the previously mentioned limitations of the FC-DCM. The remainder of the paper is laid out as follows. First, the FC-DCM by Huang (2023) is reviewed. Second, an integrative approach for FC assessments via the G-DINA model is presented. Third, the viability of the G-DINA model is tested and compared

[1]This description pertains to the most-common item-pair FC blocks. A more exhaustive list of FC blocks format is provided by Hontangas et al. (2015).

to that of the FC-DCM through a Monte Carlo simulation study and a real data illustration. Finally, practical implications, limitations and future research lines are discussed.

1.1 | The forced-choice diagnostic classification model

Before introducing the FC-DCM, the notation used throughout the remainder of the paper is established. In the DCM literature, N, J and K often denote the number of respondents, items and attributes involved in an assessment, respectively. In the most common case of binary attributes (e.g., mastery vs. non-mastery of competences), there are $L = 2^K$ possible latent classes. Let $\boldsymbol{Y}$ represent a binary response matrix (e.g., correct vs. incorrect) with dimensions $N \times J$, and let $\boldsymbol{Q}$ denote the Q-matrix (Tatsuoka, 1983) with dimensions $J \times K$, where q_{jk} equals 1 or 0 depending on whether item j measures attribute k or not, respectively. Lastly, let $\boldsymbol{\alpha}_i$ represent the attribute profile of respondent i. For instance, $\boldsymbol{\alpha}_i = \{101\}$ means that, for an assessment measuring $K = 3$ attributes, respondent i masters the first and third attributes, but not the second. Furthermore, note that $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i'} = \boldsymbol{\alpha}_l$ implies that respondents i and i' belong to the same latent class l.

A few additional terms that are specific of FC assessment need to be established. To avoid ambiguity in language, we will reserve the term *statement* to individual items that are paired in FC *blocks*. Let F denote the number of FC blocks in a certain assessment. Each block compares two statements, referred to as A and B, which measure two different attributes, k_{fA} and k_{fB}, respectively. Under the FC-DCM, a response $y_{lf} = 1$ indicates that a respondent in latent class l has chosen statement A over B in block f. As an example, consider a FC block formed by two statements from the HEXACO Personality Inventory (Lee & Ashton, 2004) measuring Extraversion ($k_{f\text{A}}$) and Conscientiousness ($k_{f\text{B}}$), respectively: 'I feel that I have some likable qualities' and 'I often check my work repeatedly to find any mistakes'. Individuals that are extraverted but unconscientious are more likely to choose statement A over B, resulting in a response of $y_{lf} = 1$. Specifically, the block response function of the FC-DCM, which is the probability of endorsing statement A in block f as a function of latent class l, is given by:

$$P\left(y_{lf} = 1 \mid \boldsymbol{\alpha}_l\right) = \eta_{f0} + I\left(\alpha_{lk_{fA}} \geq \alpha_{lk_{fB}}\right)\left(.5 - \eta_{f0}\right) + \\ I\left(\alpha_{lk_{fA}} > \alpha_{lk_{fB}}\right)\left[\eta_{fA}\alpha_{lk_{fA}} + \eta_{fB}\left(1 - \alpha_{lk_{fB}}\right)\right], \tag{1}$$

where $I(\cdot)$ is the indicator function, and $\boldsymbol{\alpha}_{lk_{fA}}$ and $\boldsymbol{\alpha}_{lk_{fB}}$ denote the status for latent class l for the attributes involved in statements A and B, respectively. Here, η_{f0} represents the probability of choosing statement A for respondents for whom $\alpha_{lk_{fA}} < \alpha_{lk_{fB}}$, while η_{fA} and η_{fB} denote the effects of attributes k_{fA} and k_{fB}, respectively. Considering that each statement is unidimensional, and therefore each block is two-dimensional, there are four possible latent groups for each block. Table 1 displays the probability of endorsement for each latent group as a function of the FC-DCM parameters, which are also illustrated in Figure 1. Note that, out of the three parameters per block in the FC-DCM, one of them (η_{f0}) is a block-level parameter, while the other two (η_{fA} and η_{fB}) are statement-level parameters. Remarkably, according to Equation 1, whenever a respondent possesses neither (i.e., $\boldsymbol{\alpha}_{lk_{fA}} = \boldsymbol{\alpha}_{lk_{fB}} = 0$) or both (i.e., $\boldsymbol{\alpha}_{lk_{fA}} = \boldsymbol{\alpha}_{lk_{fB}} = 1$) attributes, the FC-DCM assumes there is no particular preference for either statement, resulting in a probability of endorsement equal to .5. In other words, the FC-DCM constraints $P(00) = P(11) = .5$.

In his study, Huang (2023) employed a higher-order structure to model attribute distributions (de la Torre & Douglas, 2004) and found that both measurement and structural parameters could be accurately estimated using Markov chain Monte Carlo (MCMC). Additionally, beyond between-statement multidimensionality (i.e., where each statement measures only one attribute), Huang also explored the scenario of within-statement multidimensionality (i.e., where each statement measures two attributes). In the latter case, a conjunctive or disjunctive rule must be adopted to represent respondents' latent

TABLE 1 Correspondence between latent groups' probabilities of endorsement and parameters for the FC-DCM, the MUPP and the G-DINA model.

Group	FC-DCM	MUPP	G-DINA
P (00)	.5	$\frac{g_{fA}\left(1-g_{fB}\right)}{g_{fA}\left(1-g_{fB}\right)+\left(1-g_{fA}\right)g_{fB}}$	δ_{f0}
P (10)	$.5+\eta_{fA}+\eta_{fB}$	$\frac{\left(1-s_{fA}\right)\left(1-g_{fB}\right)}{\left(1-s_{fA}\right)\left(1-g_{fB}\right)+s_{fA}g_{fB}}$	$\delta_{f0}+\delta_{fA}$
P (01)	η_{f0}	$\frac{g_{fA}s_{fB}}{g_{fA}s_{fB}+\left(1-g_{fA}\right)\left(1-s_{fB}\right)}$	$\delta_{f0}+\delta_{fB}$
P (11)	.5	$\frac{\left(1-s_{fA}\right)s_{fB}}{\left(1-s_{fA}\right)s_{fB}+s_{fA}\left(1-s_{fB}\right)}$	$\delta_{f0}+\delta_{fA}+\delta_{fB}+\delta_{fAB}$

Note: These parameters relate to the case where both statements are measuring their corresponding attributes in a direct direction.
Abbreviations: FC-DCM, forced choice diagnostic classification model; G-DINA, generalized deterministic, input, noisy 'and' gate; MUPP, multi-unidimensional pairwise preferences.

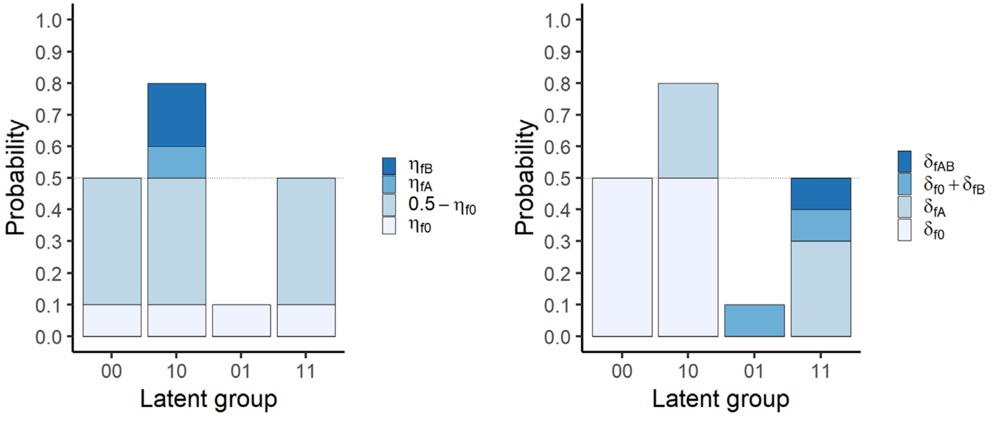

FIGURE 1 Correspondence between block parameters of endorsement for the FC-DCM and G-DINA model.

groups. For example, under a conjunctive rule, the status is redefined as $\xi_{lk_{fA}} = \prod_{k \in k_{fA}} \alpha_{lk_{fA}}^{q_{fAk}}$, resulting in $\xi_{lk_{fA}} = 1$ or 0, depending on whether respondents in latent class l possess all the attributes involved in statement A of block f or not, respectively. Thus, in that case, $\alpha_{lk_{fA}}$ and $\alpha_{lk_{fB}}$ in Equation 1 are substituted by $\xi_{lk_{fA}}$ and $\xi_{lk_{fB}}$, respectively.

The FC-DCM represents a significant milestone in the diagnostic assessment of noncognitive traits. From here, some areas of further developments can be identified. The most notable area for improvement relates to the fact that, as indicated by Equation 1 and illustrated in Table 1, the block response function of the FC-DCM imposes a fixed probability of .5 for the two latent groups where $\alpha_{lk_{fA}} = \alpha_{lk_{fB}}$. While this assumption might seem reasonable due to the lack of clear preference for either statement, it may not hold in various scenarios. Consider the previously mentioned FC block measuring *Extraversion* and *Conscientiousness*, respectively: 'I feel that I have some likable qualities' and 'I often check my work repeatedly to find any mistakes'. The first statement is relatively easy to agree with, as most extraverted individuals (and even some introverted ones) would likely agree that they possess at least some likable qualities. In contrast, the second statement is more extreme, with a slight association to obsessive behaviours, making it less likely to resonate with all conscientious individuals. In this scenario, one could expect the probability of choosing statement A for individuals who are both extraverted and conscientious, $P\left(y_{lf}=1 \mid \boldsymbol{\alpha}_l=\{11\}\right)$ to be >.5. Furthermore, even in FC blocks where the FC-DCM assumption might appear reasonable, the absence of alternative models to compare and evaluate the appropriateness of this constraint presents a significant challenge. As we will show later, a more general model would allow estimating those probabilities.

Beyond the constraints of the block response function, the higher-order structure used in the formulation of the FC-DCM also imposes restrictions on person parameters (de la Torre & Douglas, 2004). Namely, respondents' attributes are assumed to arise from a continuous second-order latent trait following a logistic model. This structure was designed to represent scenarios where attributes belong to the same domain (e.g., mathematical ability) and are therefore highly correlated (de la Torre & Douglas, 2004). However, this structure is less representative of many noncognitive traits, such as personality, where latent constructs are expected to be only mildly correlated (Booth & Hughes, 2014). Additionally, the FC-DCM is currently estimated using MCMC estimation in JAGS (Plummer, 2003), which is computationally intensive. In this paper, we aim to enhance diagnostic assessments with FC blocks by proposing an approach that: (a) incorporates a flexible block response function, allowing it to accommodate scenarios where $P\left(y_{lf}=1 \mid \boldsymbol{\alpha}_l=\{00\}\right) \neq P\left(y_{lf}=1 \mid \boldsymbol{\alpha}_l=\{11\}\right) \neq .5$; (b) permits the specification of different attribute structures beyond the higher-order model to better capture the theoretical relationships among attributes and (c) is efficiently estimated and easy to integrate with popular DCM R packages (R Core Team, 2024), such as CDM (George et al., 2016) or GDINA (Ma & de la Torre, 2020), facilitating complementary psychometric analyses including reliability assessment, model fit evaluation and *Q*-matrix validation, among others. As a result, we propose using the well-established G-DINA model (de la Torre, 2011) as an integrated framework for the analysis of FC blocks.

1.2 | The G-DINA framework for forced-choice assessments

1.2.1 | Review of the G-DINA model

The G-DINA model is a general, saturated DCM that encompasses various reduced models, such as the deterministic, input, noisy '*and*' gate (Junker & Sijtsma, 2001), the deterministic input, noisy '*or*' gate (Templin & Henson, 2006) or the *additive* cognitive diagnostic model (de la Torre, 2011). While these reduced models make specific assumptions about the item response function – thereby constraining some item parameters – the G-DINA model assigns a unique probability of success for each latent group in an unconstrained manner. It can be efficiently estimated using marginal maximum likelihood with the expectation–maximization algorithm (de la Torre, 2011) and supports a variety of attribute distributions, including the higher-order structure, attribute hierarchies (e.g., Tu et al., 2019) or a saturated structure. Additionally, the G-DINA model can be easily complemented with multiple analyses, including reliability assessment, model and item fit evaluation and comparison, empirical *Q*-matrix estimation, empirical *Q*-matrix validation, differential item functioning and person fit evaluation. All these analyses, available in well-known R packages such as 'CDM' and 'GDINA', enable comprehensive diagnostic assessments within a unified framework.

In its traditional notation, the item response function of the G-DINA model is defined as follows:

$$P\left(\boldsymbol{\alpha}_{lj}^*\right)=\delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{lk}+\sum_{k'=k+1}^{K_j^*}\sum_{k=1}^{K_j^*-1}\delta_{jkk'}\alpha_{lk}\alpha_{lk'}\ \ldots+\delta_{j12\ldots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk}, \quad (2)$$

where $P\left(\boldsymbol{\alpha}_{lj}^*\right)$ is the probability of correctly answering item j for latent group $\boldsymbol{\alpha}_{lj}^*$, which is the reduced attribute vector containing only the measured attributes by item j. Additionally, δ_{j0} denotes the baseline probability of success for item j for respondents who do not master any attribute, δ_{jk} represents the additive effect on the probability of success for item j due to mastering attribute k, $\delta_{jkk'}$ captures the interaction effect of mastering attributes k and k' and $\delta_{j12\ldots K_j^*}$ represents the interaction effect of mastering all K_j^* attributes required by item j. Adapted to the context of noncognitive factors, instead of discussing the probability of success, we would talk about the probability of endorsement, and instead of mastering an attribute, we would discuss whether the examinee possesses it or not.

1.2.2 | Adjusting the G-DINA model to FC blocks

Adapting the G-DINA model to FC assessments requires careful consideration of several key factors, which we outline below. We focus specifically on the case of pairs of unidimensional statements (i.e., between-statement multidimensionality). First, to maintain consistency in notation and considering that all blocks are measured by two attributes, the G-DINA response function can be reformulated to accommodate FC assessments as follows:

$$P\left(y_{lf}=1 \mid \boldsymbol{\alpha}_l\right)=\delta_{f0}+\delta_{fA}\alpha_{lk_{fA}}+\delta_{fB}\alpha_{lk_{fB}}+\delta_{fAB}\alpha_{lk_{fA}}\alpha_{lk_{fB}}, \tag{3}$$

where, given any block f, $\boldsymbol{\alpha}_{lk_{fA}}$ and $\boldsymbol{\alpha}_{lk_{fB}}$ indicate whether the person with attribute profile l possesses the specific attribute measured by the first item and second items, respectively, (i.e., $\boldsymbol{\alpha}_{lk_{fA}} \in \{0,1\}$ and $\boldsymbol{\alpha}_{lk_{fB}} \in \{0,1\}$), $\boldsymbol{\delta}_{f0}$ denotes the probability of endorsing block f for those respondents lacking the attributes measured by statements A and B, $\boldsymbol{\delta}_{fA}$ and $\boldsymbol{\delta}_{fB}$ represent the additive effect for respondents possessing the attribute measured by statement A or B, respectively and $\boldsymbol{\delta}_{fAB}$ captures the interaction effect of possessing both attributes. As detailed in Table 1 and exemplified in Figure 1, these components are estimated freely to yield a probability bounded between 0 and 1. Specifically, Equation 3 is illustrated in Figure 1 for a block formed by two direct statements, compared to the FC-DCM formulation (see Equation 1).

The second consideration relates to Q-matrix completeness and model identifiability. A Q-matrix is said to be complete if it contains an identity matrix (Köhn & Chiu, 2017), which is a necessary condition for the identifiability of many reduced diagnostic models. Model identifiability is crucial because it ensures that all latent classes can be distinguished (Xu & Zhang, 2016). In the case of FC assessments, where all blocks are measured by two attributes, this condition will not be satisfied. However, Gu and Xu (2020) demonstrated that the G-DINA model offers greater flexibility than reduced models. Therefore, while strict identifiability cannot be achieved without a complete Q-matrix, generic identifiability can be fulfilled, which is suitable for practical applications. To ensure generic identifiability, the Q-matrix must have the following structure:

$$Q=\begin{pmatrix} Q_1 \\ Q_2 \\ Q^\star \end{pmatrix}, \tag{4}$$

where Q_1 and Q_2 are square submatrices with ones on the diagonal (up to column permutation), while the remaining elements can be either 0 or 1. Additionally, $Q^\star$ must measure each attribute at least once (Gu & Xu, 2020).

The third consideration also pertains to model identifiability. In addition to the Q-matrix structure, a necessary condition for the G-DINA to be identifiable is the presence of monotonic probabilities of success (Gu & Xu, 2020, 2021). Monotonicity is achieved when higher probabilities of success are associated with latent groups that master a greater number of attributes. However, for FC blocks, where a response $y_{if}=1$ indicates that respondent i prefers statement A over statement B in block f, the probability of endorsement does not increase monotonically with the number of attributes possessed (see Figure 1). In fact, a respondent's expected response to a block depends not only on their attribute profile and the attributes measured by the block, but also on the polarity of the block. Table 2 summarizes the four possible configurations of an FC block based on the

TABLE 2 Statement preference, expected response and initial values for block parameter estimation depending on the direction of the statements.

Statement A	Statement B	{00}	{10}	{01}	{11}
		Statement preference			
Direct	Direct	—	*A*	*B*	—
Direct	Inverse	*B*	—	—	*A*
Inverse	Direct	*A*	—	—	*B*
Inverse	Inverse	—	*B*	*A*	—
		Expected response			
Direct	Direct	—	1	0	—
Direct	Inverse	0	—	—	1
Inverse	Direct	1	—	—	0
Inverse	Inverse	—	0	1	—
		Initial value for block parameter estimation			
Direct	Direct	.5000	.9999	.0001	.5000
Direct	Inverse	.0001	.5000	.5000	.9999
Inverse	Direct	.9999	.5000	.5000	.0001
Inverse	Inverse	.5000	.0001	.9999	.5000

Note: {00} = latent group formed by respondents that do not possess any attribute. The hyphen (–) indicates that the latent group does not have a clear preference for any of the statements.

direction (direct or inverse) of the statements and the expected preferred statement for each latent group. Based on these configurations, monotonicity is redefined as the alignment between expected responses and the magnitude of the probabilities of endorsement. The G-DINA model will be generically identified as long as the Q-matrix is constructed according to Equation 4 and 'polarity-specific monotonicity' is satisfied.

Given that the G-DINA model is originally designed for individual items rather than FC blocks, it does not inherently handle the scenario where $y_{if} = 1$ indicates that respondent i prefers statement A over statement B in block f. Additionally, the model does not account for the direction of the statements. To enable the G-DINA model to interpret the direction of responses and achieve 'polarity-specific monotonicity' aligned with the FC blocks design, the estimation process needs to be either constrained or guided appropriately. The first approach involves directly imposing monotonicity constraints during the estimation process (Hong et al., 2016; Ma & Jiang, 2021). This strategy simplifies estimation, especially under challenging conditions (e.g., small sample sizes), by preventing deviations from monotonicity caused by sampling error (i.e., overfitting; Ma & Jiang, 2021). However, this approach is effective only when the data-generating process is truly monotonic; otherwise, enforcing monotonicity constraints may result in model misspecification (de la Torre & Sorrel, 2017). This study adopts the second approach, which guides, rather than constrains, the probabilities of endorsement for latent groups, so that they align with the intended meaning of the responses based on the polarity of the block. Two different strategies to achieve this are explained below.

One effective approach is to use initial values in the estimation process. By assigning a high initial probability to the latent group expected to endorse the block (i.e., choose A over B), and a low initial probability to the latent group expected to reject the block (i.e., choose B over A), the model can be directed to match the FC assessment design. Table 2 summarizes the initial values applied in this study for the different latent groups based on the direction of the statements. As shown below in the simulation study and in a sensitivity analysis with real data (see Table S1), initial values between 1e-3 and 1e-9 help ensure that attributes are interpreted in the intended direction with minimal impact on attribute classification and parameter estimates, without the need of imposing constraints on parameter estimation.

Another option is to use Bayes modal (BM) estimation, which can guide block parameter estimation using prior distributions. Ma and Jiang (2021) proposed using BM estimation in small-scale scenarios to mitigate boundary problems sometimes encountered in DCM applications. In BM estimation, the probability of success for a latent group is calculated as follows:

$$P\left(\boldsymbol{\alpha}_{lj}^{*}\right)=\frac{r_{lj}+\left(\beta_1-1\right)}{n_l+\left(\beta_1+\beta_2-2\right)}, \tag{5}$$

where r_{lj} is the number of correct responses among respondents in latent group $\boldsymbol{\alpha}_{lj}^{*}$, n_l is the total number of respondents in that group and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the hyperparameters of the Beta distribution. The primary limitation of this approach is that it might introduce bias in parameter estimation, for example, if informative and inappropriate priors are used (Ma & Jiang, 2021).

1.2.3 | The G-DINA Model as a Generalized Version of the FC-DCM

Despite having a different parametrization, the FC-DCM can be framed as a restricted version of the G-DINA. Table 1 presents the probability of endorsement for each latent group as defined by the parameters of the FC-DCM (Equation 1) and the G-DINA model (Equation 2). Figure 1 also reflects the similarities between both models. Specifically, the G-DINA model can accommodate the probabilities of endorsement in the FC-DCM under the following constraints:

$$\delta_{f0}=.5, \tag{6}$$

$$\delta_{fA}=\eta_{fA}+\eta_{fB}, \tag{7}$$

$$\delta_{fB}=\eta_{f0}-.5, \tag{8}$$

$$\delta_{fAB}=.5-\eta_{fA}-\eta_{fB}-\eta_{f0}=-\left(\delta_{fA}+\delta_{fB}\right). \tag{9}$$

Thus, the FC-DCM can be expressed using the G-DINA parametrization. This reparameterization offers the immediate advantage of enabling more efficient estimation procedures compared to MCMC. Specifically, marginal maximum likelihood could be used, applying the aforementioned constraints. Additionally, BM estimation can be easily applied by using highly informative priors for the probabilities of endorsement of the indecisive latent groups, ensuring that these probabilities are approximately equal to .5. For instance, consider an FC block composed of two direct statements. Hence, the FC-DCM assumes $P(00)=P(11)=.5$, $P(10)\geq .5$ and $P(01)\leq .5$. These constraints can be effectively managed using BM estimation by setting the following priors: $P(00)$, $P(11)\sim Beta(10^8, 10^8)$ and $P(10)$, $P(01)$ $Beta(1, 1)$. As indicated in Equation 5, extreme priors for $P(00)$ and $P(11)$ will ensure that $P(00)\approx P(11)\approx .5$, while the noninformative priors for $P(10)$ and $P(01)$ will not impact the estimates, aligning them with those obtained through maximum likelihood estimation.

1.3 | The present study

The goal of the study is to show how it is possible to integrate the FC-DCM model within the G-DINA model framework. This will allow us to work without assuming that individuals who possess both traits (or neither) measured by an item block will choose randomly. Additionally, since the G-DINA model is being estimated, we can use the most computationally efficient and commonly used estimation algorithms, as well as implement other methods developed within this framework to study, for example, model fit. This integration is demonstrated in the remainder of the article through a simulation study and an empirical example.

2 | SIMULATION STUDY

2.1 | DCM implementation

The primary objective of this simulation study is to assess the G-DINA model for FC assessments and compare its performance with the FC-DCM under various conditions. We will utilize the reparametrized version of the FC-DCM, employing BM estimation with informative priors for constrained parameters and noninformative priors for free parameters, as detailed in the previous section. This estimation approach was implemented using the function provided by Ma and Jiang (2021). In contrast, the G-DINA model was estimated using the GDINA package, which applies marginal maximum likelihood estimation with the expectation–maximization algorithm. The only adjustment made in the estimation process was the incorporation of initial values for the probabilities of endorsement of the latent groups, as specified in Table 2. Neither model was constrained by a specific attribute distribution (e.g., higher-order model); instead, both used a saturated structure. Additionally, the true model was included as an upper baseline to compare performance, using the generating item parameters and attribute distribution. Attribute profile classifications were obtained using the expected a posteriori (EAP) estimator (Huebner & Wang, 2011).

2.2 | Design and data generation

Data were generated using a combination of the G-DINA model and the *multi-unidimensional pairwise-preference* (MUPP) model (Stark et al., 2005). The MUPP model is a well-established item response theory model for FC blocks formed by two statements, and it can be easily adapted to the DCM framework. Specifically, it models the probability of choosing one statement over the other as a ratio of the probabilities of agreeing with each statement independently. In the context of unidimensional statements, the probabilities of endorsement for different latent groups are shown in Table 1. For instance, in the case of two direct statements, the probability of choosing statement A over B for the latent group {10} is given by:

$$P\left(y_{lf}=1 \mid \boldsymbol{\alpha}_l=\{10\}\right)=\frac{\left(1-s_{fA}\right)\left(1-g_{fB}\right)}{\left(1-s_{fA}\right)\left(1-g_{fB}\right)+s_{fA}g_{fB}}, \tag{10}$$

where s_{fA} is the *slip* parameter for statement A (i.e., the probability of not agreeing with statement A despite possessing its attribute), and g_{fB} is the *guessing* parameter for statement B (i.e., the probability of agreeing with statement B despite not possessing its attribute). Similar to the G-DINA, the MUPP model allows for $P(00)$ and $P(11)$ to differ from .5, acknowledging that both statements may have different discriminating power. If the parameters for both statements are constrained to be equal (i.e., $g_{fA}=g_{fB}=s_{fA}=s_{fB}$), the assumption $P(00)=P(11)=.5$ of the FC-DCM is satisfied (see Table 1). Seven different factors were systematically manipulated in the simulation study: sample size ($N=500$, 1000), number of attributes ($K=5$, 10), number of statements per attribute (JK = 5, 10), number of FC blocks ($F=30$, 60), FC block polarity (100% homopolar/0% heteropolar, 70% homopolar/30% heteropolar), attribute correlations (AC = 0, .3, .6) and range of guessing and slip parameters to manipulate item quality ($R=0$, .2, .4). The levels for several factors, such as sample size, number of attributes, number of statements per attribute and number of blocks, were taken from Huang (2023) to facilitate comparison between the two studies. Additionally, while Huang (2023) only considered the case of homopolar blocks (i.e., both statements being direct), we included an additional condition where 30% of the blocks were heteropolar (i.e., one direct statement and one inverse statement). The inclusion of this condition was based on the relevance of this variable in previous FC literature concerning score ipsativity (e.g., Kreitchmann et al., 2023). Specifically, it is expected that the presence of heteropolar blocks will reduce ipsativity, which means, for example, that the correlation matrix between traits will be less likely to be biased towards a diagonal matrix.

Instead of using a higher-order model for the attribute structure, we employed the multivariate normal threshold model (Chiu et al., 2009) to generate respondents' attribute patterns. Specifically, K continuous latent variables were drawn from a multivariate normal distribution with a mean of 0 and correlations set according to the values of the AC factor (AC = 0, .3, .6). These continuous latent variables were then dichotomized, with each attribute $\boldsymbol{\alpha}_{ik}$ assigned a value of 0 or 1 based on whether the continuous score was below or above 0, respectively.

Lastly, the statement parameters used in the MUPP model were drawn from $g_{fA}, g_{fB}, s_{fA}, s_{fB} \sim U(.25 - R/2, .25 + R/2)$, where R represents the range of the guessing and slip parameters (R = 0, .2, .4). As previously noted, this is a critical simulation factor because when R = 0, it implies that $g_{fA} = g_{fB} = s_{fA} = s_{fB} = .25$, which upholds the FC-DCM assumption of $P(00) = P(11) = .5$. In other words, the FC-DCM is correctly specified under R = 0. Conversely, when R is > 0, the statement parameters will differ, violating this assumption. In these scenarios, the G-DINA model is expected to perform better than the FC-DCM due to its greater flexibility.

Given all this, the data generation process was carried out in five stages. Initially, a Q-matrix was constructed with $K \times$ JK statements, ensuring that all statements were unidimensional and each attribute was measured by the same number of statements. The guessing and slip parameters for each statement were generated according to the previously specified details. Second, statements were paired to form F blocks under several constraints: (1) each block had to include two statements measuring different attributes, (2) all statements needed to be used at least once before any were reused if the number of blocks exceeded twice the number of statements and (3) no two blocks could consist of the same pair of statements. Third, the probability of endorsement (i.e., choosing A over B) was calculated for each block using the MUPP model. For the condition with heteropolar blocks, the parameters were reversed for inverse statements (e.g., $g_{jA} = 1 - g_{jA}$). Fourth, N attribute profiles were generated based on the multivariate normal threshold model. Finally, N responses were generated using the G-DINA model, incorporating the computed probabilities of endorsement for each latent group and FC block based on the MUPP model along with the generated attribute profiles.

Data generation and analyses were conducted using the R package GDINA version 2.9.4 (Ma & de la Torre, 2020). All codes related to the simulation study are available at the Online Appendix https://osf.io/h6x9e/.

2.3 | Performance measures

The performance of the FC-DCM and G-DINA model was assessed in terms of classification accuracy, block parameter recovery, attribute correlations recovery and detection rates of relative model fit indices. Specifically, classification accuracy was evaluated using the proportion of correctly classified attribute vectors (PCV), defined as:

$$\mathrm{PCV} = \frac{\sum_{i=1}^{N} I\left(\hat{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i\right)}{N}, \tag{11}$$

where $\hat{\boldsymbol{\alpha}}_i$ and $\boldsymbol{\alpha}_i$ represent the estimated and true attribute profile for respondent i, respectively. Note that the PCV accounts for all attributes in the attribute profile, making it stricter with an increasing number of attributes (e.g., K = 10 vs. K = 5).

Block parameter recovery was measured by the root mean squared error (RMSE) of the probabilities of endorsement:

$$\mathrm{RMSE}\left[P\left(y_{lf}=1\mid\boldsymbol{\alpha}_l\right)\right]=\sqrt{\frac{1}{L\cdot N}\sum_{l=1}^{L}\sum_{i=1}^{N}\left(\hat{P}\left(y_{lf}=1\mid\boldsymbol{\alpha}_l\right)-P\left(y_{lf}=1\mid\boldsymbol{\alpha}_l\right)\right)^2}, \tag{12}$$

where $\hat{P}\left(y_{lf}=1\mid\boldsymbol{\alpha}_l\right)$ and $P\left(y_{lf}=1\mid\boldsymbol{\alpha}_l\right)$ denote the estimated and generating probabilities of endorsement for latent class l for item j, respectively. Attribute correlation recovery was evaluated by comparing the average tetrachoric correlation of the estimated attribute profiles with the generating attribute correlations (AC=0, .3, .6). To analyse the effects of simulation factors on the DCM procedures, repeated measures ANOVAs were performed. An effect was considered relevant if the partial eta-squared (η_p^2) was equal to or >.14 (Cohen, 2013).

Finally, the models were also compared in terms of relative fit using the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978).

2.4 | Results

2.4.1 | Classification accuracy

Table 3 displays the PCV for the true model, the G-DINA model and the FC-DCM across the different levels of range for guessing and slip parameters (within-method effect of $\eta_p^2=.24$), number of attributes (between-method effect of $\eta_p^2=.90$), number of FC blocks (between-method effect of $\eta_p^2=.76$) and polarity (between-method effect of $\eta_p^2=.14$). As could be anticipated, all DCM procedures showed higher classification accuracy with fewer attributes, longer tests and the presence of heteropolar blocks. The most significant difference between the G-DINA and FC-DCM was related to the range of guessing and slip parameters. Specifically, when these parameters were equal for all statements (i.e., $R=0$), the FC-DCM generally provided slightly more accurate classifications (average PCV=.623) compared to the G-DINA (average PCV=.608). In more realistic scenarios, with varying discriminatory power among statements, the G-DINA consistently outperformed the FC-DCM in terms of classification accuracy. The only exception was found in the scenario with a small range ($R=.2$), a large number of attributes ($K=10$), and shorter tests ($F=30$), where the FC-DCM yielded slightly better classifications. This finding could be explained by a balance between the condition of $R=.2$, which, although it contradicts the FC-DCM assumptions, is less extreme than $R=.4$, and the challenging conditions (e.g., large dimensionality and a short test) for parameter estimation, which have a greater impact on the estimation of the G-DINA model as it is more complex. Regardless, the G-DINA achieved overall average PCV values of .631 and .675 for $R=.2$ and .4, respectively, compared to .597 and .492 for the FC-DCM. Finally, the standard deviation of the PCV for both the G-DINA and FC-DCM models was generally similar to that of the true model. However, the FC-DCM exhibited greater instability under a wide range of statement parameters ($R=.4$).

2.4.2 | Recovery of block parameters

Table 4 presents the RMSE of block parameter estimates for the G-DINA model and the FC-DCM, considering different ranges the guessing and slip parameters (within-method effect of $\eta_p^2=.77$), the number of FC blocks (between-method effect of $\eta_p^2=.39$), attribute correlations (between-method effect of $\eta_p^2=.36$) and the number of attributes (between-method effect of $\eta_p^2=.27$). Overall, both models provided more accurate estimates with longer tests and fewer, less correlated attributes. Consistent with the PCV results, the FC-DCM yielded more accurate estimates (average RMSE=.050) compared to the G-DINA model (average RMSE=.079) only when its assumptions were satisfied ($R=0$). In contrast, the RMSE of the G-DINA model remained relatively stable across the different levels of variability in statement parameters (average RMSE=.075 for $R=.2$, and .070 for $R=.4$). Conversely,

TABLE 3 Classification accuracy.

R	*K*	*F*	*P*	True model		G-DINA		FC-DCM	
.4	5	60	70/30	.968	(.017)	**.964**	(.020)	.917	(.046)
			100/0	.965	(.019)	**.959**	(.022)	.777	(.114)
		30	70/30	.854	(.037)	**.837**	(.043)	.700	(.088)
			100/0	.840	(.040)	**.819**	(.048)	.527	(.136)
	10	60	70/30	.731	(.050)	**.663**	(.066)	.460	(.089)
			100/0	.719	(.052)	**.647**	(.068)	.323	(.121)
		30	70/30	.421	(.066)	**.270**	(.073)	.133	(.055)
			100/0	.403	(.059)	**.244**	(.068)	.095	(.052)
	Average			.738	(.212)	**.675**	(.271)	.492	(.294)
.2	5	60	70/30	.962	(.011)	**.957**	(.012)	.949	(.013)
			100/0	.945	(.014)	**.935**	(.017)	.845	(.062)
		30	70/30	.824	(.026)	**.806**	(.028)	.785	(.031)
			100/0	.783	(.025)	**.757**	(.032)	.674	(.063)
	10	60	70/30	.678	(.042)	**.608**	(.049)	.582	(.046)
			100/0	.654	(.036)	**.575**	(.046)	.512	(.043)
		30	70/30	.370	(.058)	.231	(.058)	**.244**	(.048)
			100/0	.333	(.039)	.177	(.044)	**.189**	(.037)
	Average			.694	(.225)	**.631**	(.280)	.597	(.260)
0	5	60	70/30	.960	(.008)	.954	(.009)	**.956**	(.009)
			100/0	.887	(.042)	**.895**	(.034)	.850	(.063)
		30	70/30	.814	(.022)	.794	(.024)	**.801**	(.022)
			100/0	.722	(.045)	**.723**	(.034)	.705	(.050)
	10	60	70/30	.660	(.040)	.589	(.045)	**.608**	(.041)
			100/0	.607	(.020)	.536	(.034)	**.553**	(.026)
		30	70/30	.353	(.056)	.218	(.056)	**.281**	(.051)
			100/0	.299	(.022)	.151	(.034)	**.235**	(.025)
	Average			.663	(.225)	.608	(.280)	**.623**	(.246)

Note: Average (and standard deviation) proportion of correctly classified attribute vectors (PCV) across conditions. Highest PCV between the G-DINA and FC-DCM are shown in bold.

Abbreviations: *F*, number of FC blocks; *K*, number of attributes; *P*, FC block polarity (homopolar/heteropolar); *R*, range of the guessing and slip parameters.

the accuracy of the FC-DCM was considerably impacted by this factor (average RMSE = .099 and .178 for R = .2 and .4, respectively). Lastly, estimation accuracy for both models was consistent across replicates, as evidenced by the small standard deviations reported in Table 4.

2.4.3 | Recovery of attribute correlations

Figure 2 shows the average correlations for the true model, the G-DINA model and the FC-DCM as a function of attribute correlations (within-method effect of $\eta_p^2 = .21$), block polarity (within-method effect of $\eta_p^2 = .18$) and the range for the guessing and slip parameters (within-method effect of $\eta_p^2 = .14$). Several notable trends can be noted in this figure. First, attribute correlation recovery was generally satisfactory (RMSE ≤ .130) across all methods when attributes were independent (AC = 0) or when heteropolar blocks were present (P = 70/30). Second, the FC-DCM tended to underestimate attribute correlations, particularly under the condition of homopolar blocks

TABLE 4 Recovery of block parameters.

R	*F*	*AC*	*K*	G-DINA		FC-DCM	
.4	60	0	5	**.043**	(.009)	.153	(.010)
			10	**.050**	(.011)	.160	(.010)
		.3	5	**.048**	(.011)	.158	(.013)
			10	**.054**	(.011)	.170	(.014)
		.6	5	**.066**	(.014)	.170	(.019)
			10	**.073**	(.014)	.191	(.021)
	30	0	5	**.051**	(.013)	.163	(.016)
			10	**.096**	(.018)	.180	(.016)
		.3	5	**.058**	(.015)	.178	(.022)
			10	**.100**	(.019)	.197	(.019)
		.6	5	**.083**	(.019)	.198	(.028)
			10	**.113**	(.020)	.217	(.023)
	Average			**.070**	(.027)	.178	(.026)
.2	60	0	5	**.045**	(.009)	.081	(.006)
			10	**.054**	(.011)	.084	(.005)
		.3	5	**.050**	(.010)	.084	(.006)
			10	**.059**	(.012)	.091	(.007)
		.6	5	**.068**	(.014)	.095	(.010)
			10	**.078**	(.014)	.107	(.013)
	30	0	5	**.057**	(.013)	.085	(.008)
			10	**.100**	(.017)	**.100**	(.010)
		.3	5	**.065**	(.016)	.094	(.013)
			10	**.106**	(.018)	.114	(.014)
		.6	5	**.090**	(.019)	.118	(.022)
			10	**.121**	(.019)	.139	(.021)
	Average			**.075**	(.028)	.099	(.021)
0	60	0	5	.047	(.009)	**.026**	(.006)
			10	.056	(.011)	**.032**	(.006)
		.3	5	.054	(.011)	**.034**	(.009)
			10	.062	(.012)	**.042**	(.010)
		.6	5	.075	(.014)	**.055**	(.015)
			10	.084	(.015)	**.064**	(.014)
	30	0	5	.060	(.014)	**.033**	(.008)
			10	.102	(.016)	**.052**	(.010)
		.3	5	.070	(.017)	**.045**	(.014)
			10	.110	(.017)	**.065**	(.015)
		.6	5	.100	(.020)	**.073**	(.021)
			10	.129	(.019)	**.087**	(.020)
	Average			.079	(.029)	**.050**	(.022)

Note: Average (and standard deviation) root mean square error (RMSE) of block parameter estimates across conditions. Lowest RMSE between the G-DINA and FC-DCM is shown in bold.

Abbreviations: AC, attribute correlations; *F*, number of FC blocks; *K*, number of attributes; *R*, range of the guessing and slip parameters.

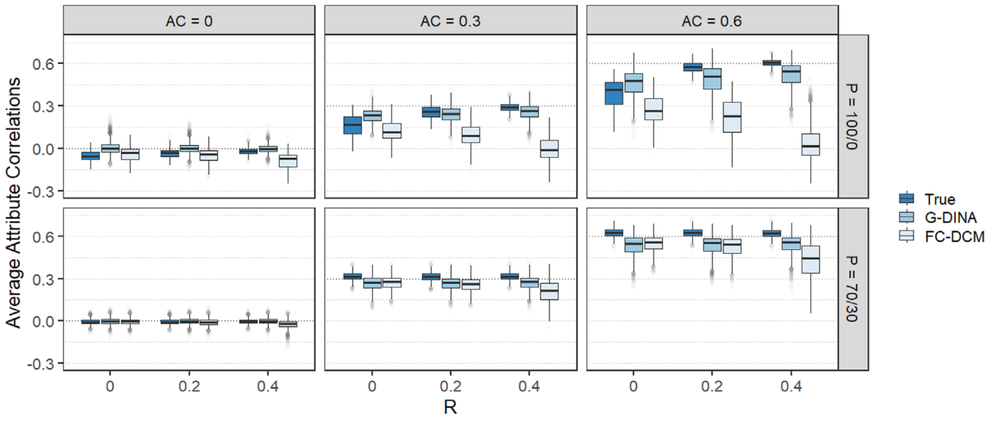

FIGURE 2 Average attribute correlations. The darker dotted line indicates the generating value for the attribute correlations (AC = 0, .3, .6). AC, attribute correlations; *P*, item block polarity (heteropolar/homopolar); *R*, range of the guessing and slip parameters.

(P = 100/0), high attribute correlations (e.g., AC = .6) and a broad range in the statement parameters (R = .4; RMSE = .650). Notably, even when the FC-DCM assumptions held (i.e., R = 0), it still underestimated attribute correlations when only homopolar blocks were present in the test (P = 100/0; .115 ≤ RMSE ≤ .371). In contrast, the G-DINA model generally provided accurate attribute correlations under most conditions (average RMSE = .120), though there was a slight underestimation tendency with highly correlated attributes (AC = .6), homopolar blocks (P = 100/0) and no variability in statement parameters (R = 0; RMSE = .204). Notably, this underestimation tendency was also observed in the true model (RMSE = .254), primarily due to the lower discriminant information provided by the blocks in the absence of variability in the statement parameters. When $R > 0$, the probabilities of endorsement become more informative, which helps in more accurately recovering attribute correlations.

2.4.4 | Relative model fit

Table 5 summarizes the proportion of replicas in which the G-DINA model was preferred over the FC-DCM based on the AIC and BIC, influenced by the range in the statement parameters, number of attributes and test length. The BIC always preferred the FC-DCM when it was correctly specified (with R = 0) and always chose the G-DINA model when the FC-DCM was clearly misspecified (with R = .4). It also tended to prefer the G-DINA model under R = .2, particularly with longer tests. The AIC proved to be a more useful relative model fit index in this study as it consistently selected the G-DINA model when the FC-DCM was misspecified (with $R > 0$) but tended to favour the simplicity of the FC-DCM when it was correctly specified (with R = 0). The exception to this was in the challenging condition of measuring a high number of attributes (K = 10) with a short test (F = 30), where the AIC preferred the G-DINA model 69.2% of the time although the true generating model was FC-DCM. This is the most challenging condition from an estimation perspective, as it corresponds to the smallest number of blocks for estimating the largest number of parameters due to the greater number of attributes. In the case of AIC, this result may reflect a case of overfitting, where a more complex model captures the idiosyncrasies of a dataset with limited information. It can be observed that, in this situation, BIC, which penalizes model complexity more, correctly identifies the generating model 100% of the time.

TABLE 5 Proportion of replicates under which AIC and BIC preferred the G-DINA model over the FC-DCM.

R	*K*	*F*	AIC	BIC
0	5	60	.000	.000
		30	.001	.000
	10	60	.000	.000
		30	.692	.000
.2	5	60	1.000	.835
		30	1.000	.662
	10	60	1.000	.785
		30	1.000	.567
.4	5	60	1.000	1.000
		30	1.000	1.000
	10	60	1.000	1.000
		30	1.000	1.000

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; *J*, number of FC blocks; *K*, number of attributes; *R*, range of the guessing and slip parameters.

3 | REAL DATA ILLUSTRATION

3.1 | Method

In this section, data from Bunji and Okada (2020) are used to illustrate both models and compare them across a series of analyses related to item and person parameter estimates and model fit. These data correspond to a two-alternative multidimensional FC format personality assessment. The sample size is 499 respondents, collected through a major online crowdsourcing service in Japan, resulting in a sample with an age range of 20 to 70 years, and 47% of the subjects are men. One case had 16 missing values, another had 2 missing values and three others had 1 missing value each. These cases were removed from the database, resulting in a final sample size of 494 cases. The questionnaire used was the Japanese version of the Big-Five factor marker questionnaire (Apple & Neff, 2012). This scale measures the Big-Five traits: emotional stability, extraversion, agreeableness, conscientiousness and intellect/imagination. The pairs were assembled to ensure an even distribution among trait pairs and to include pairs of statements that were aligned in both the same and opposite directions. The *Q*-matrix derived from the assignment of each statement to its theoretical domain is shown in Table 6. As in the simulation study, highly informative priors from the simulation study were used to set the indicated probabilities to .50 so that the G-DINA model corresponds to the FC-DCM. Additionally, the initial values for the remaining item parameters were set to ensure that the pole of the trait was kept in the desired direction. A sensibility analysis using different initial values (from .5 to 1e-9) was conducted to evaluate the impact of this decision on attribute classifications and parameter estimates.

Different approaches were used to compare the G-DINA and FC-DCM models. First, the item parameters estimated under each model were examined, with special attention given to those fixed at .50 in the FC-DCM model. Second, the models were also compared in relation to the person parameters, that is, the classifications made for each person and attribute. EAP was employed for attribute classifications. Finally, to determine which model is more appropriate for these data, the models were compared using relative fit statistics (AIC and BIC) and the congruence between the classifications obtained with these DCM models and the continuous trait scores, calculated as sum scores and using the two-parameter MUPP (MUPP-2PL) model by Morillo et al. (2016), a multidimensional item response theory (MIRT) model for FC data. The estimation of the MUPP-2PL model was performed using the R package mirt (Chalmers, 2012) following the guidelines provided in Kreitchmann et al. (2023). Specifically, when performing this estimation, we considered that in item response theory, the direction

TABLE 6 *Q*-matrix in the real data illustration.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
O	0	1	1	0	1	0	−1	−1	0	1	0	0	0	0	1	0	0	0	0	−1	0	0	1	0	1
C	0	0	0	−1	1	0	0	0	1	0	1	1	0	0	−1	0	1	−1	0	−1	0	1	0	0	0
E	−1	0	0	0	0	1	0	0	−1	−1	0	0	0	1	0	1	0	−1	1	0	0	0	0	−1	1
A	0	−1	0	0	0	0	−1	0	0	0	0	1	1	1	0	0	−1	0	1	0	−1	1	0	1	0
N	−1	0	1	−1	0	−1	0	−1	0	0	−1	0	−1	0	0	1	0	0	0	0	−1	0	−1	0	0

Note: Statements with a hyphen (−) are inverse statements.

Abbreviations: A, agreeableness; C, conscientiousness; E, extraversion; N, neuroticism; O, openness.

of traits is not identified (i.e., trait poles may flip during estimation, changing the sign of discrimination parameters without affecting response functions). Therefore, starting values are set to the block scale parameters to ensure traits are estimated in the intended direction, accounting for item polarity (e.g., extraversion instead of introversion). This is done to set initial values for the mirt function, where the online appendix code shows that assigning the expected sign of the discrimination parameter based on polarity and reversing the sign of the second item are both necessary to achieve the correct estimation. This is required since, under the MUPP-2PL model, the block scale parameter for the second item block is expected to have the opposite direction to the original item (i.e., the sign of the second item is reversed). For example, if both statements are positively keyed, the probability of endorsing statement A in a block is inversely related to the trait measured by statement B – higher levels of this trait correspond to a lower probability of endorsing A. The factor scores were calculated using the EAP estimator as in DCM, and following the recommendations of the R package, since it was a high-dimensional model (i.e., 5 factors), quasi-Monte Carlo integration was employed. Specifically, the point-biserial correlations between the DCM binary classifications and the MUPP-2PL continuous scores for each attribute were examined. The data and code are provided at the Online Appendix https://osf.io/h6x9e/.

3.2 | Results

The constraint of fixing probabilities at .50 affects 50 out of the 100 parameters estimated. In Figure 3, it can be observed that when these parameters are freely estimated within the G-DINA model, a wide variability is observed (min = .11, max = .96). When considering the 95% confidence interval, .50 is not included in 66% of the cases. That is, in some blocks, such as block 22, both parameters, when freely estimated, actually approach .50, whereas in others, like block 3, this does not happen. These probabilities are assumed to be .50 in FC-DCM under the assumption that when there is no predominant attribute, it is expected that the respondent will choose randomly. The information presented suggests that this may not be the case. The estimated probabilities for each block are presented in Table S2. Table 7 presents a selection of two items that allow for a clear understanding of the results. In Table 7, we can observe that for Block 22, the probabilities fixed at .50 remain close to this value when estimated freely, whereas for Block 3, they are clearly higher than .50. This may indicate a general tendency to agree more with some of the statements (e.g., 'Have a vivid imagination') even when neither of the attributes involved is present (.66) or when both are (.74). This can be related to the traits being measured (i.e., it is possible that the presence of Emotional Stability leads to a higher probability of agreement with 'Am relaxed most of the time', compared to the probability of agreement with 'Have a vivid imagination' if one demonstrates Openness) or response styles such as social desirability (i.e., in certain contexts, it may be preferable to choose a statement because it presents a more socially desirable image). This can be modelled using G-DINA but not with FC-DCM.

To compare the estimated parameters, we calculated the mean absolute difference between the estimates of both models. This allows us to observe that, although the estimated parameters generally appear similar, there is some variability in those probabilities fixed at .50. The mean absolute difference for the 50 parameters fixed at .50 is .18, while for the other 50 parameters, this value is slightly lower, reducing to .07. The next step was to assess the impact this has on the classification of individuals, a result that we summarize in the following text.

When comparing the classifications derived from each model, we found that, despite the relative similarity in some of the item parameters, the classifications were very dissimilar. Although 79.2% of the classifications are consistent on an attribute-by-attribute basis, the estimated attribute patterns from the two models match in just 30.6% of the cases. These differences in classification can be explained by the differences in the estimated parameters that were mentioned earlier. Given this, one might ask which model is preferable. To answer this question, we first examined the relative fit and found that, despite the larger number of parameters of the G-DINA model (131 vs. 81), it was preferred over the FC-DCM according to the AIC (14,434.46 vs. 14,926.91) and BIC (14,985.00 vs. 15,267.32). These

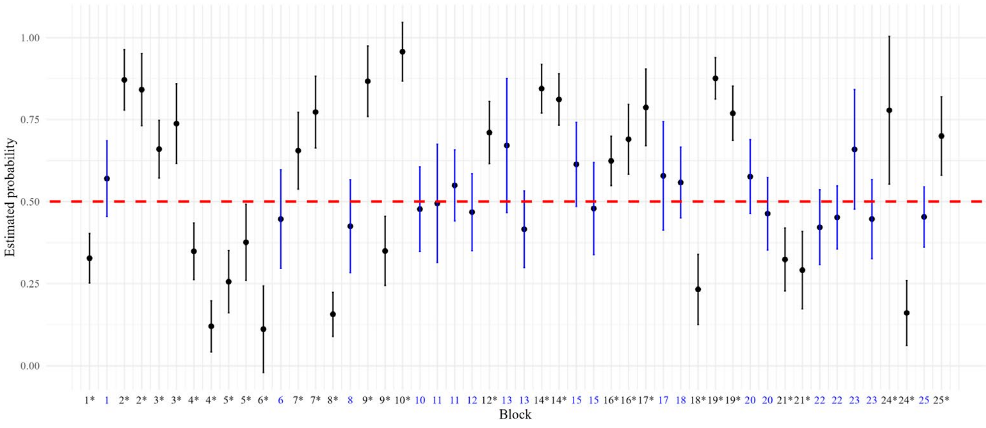

FIGURE 3 G-DINA estimated probabilities that are fixed at .50 in the FC-DCM. 95%-confidence interval containing .50 is highlighted in blue. Each of the 25 blocks contains two of these parameters. The block number appears in blue if the parameter's 95%-confidence interval contains .50, while an asterisk (*) is added when the confidence interval doest not include .50, indicating a statistically significant difference from this value.

results indicate that the constraint of fixing the probabilities of success at .50 leads to a considerably worse fit. Another piece of evidence in favour of this decision is found in the congruence between the direct sum score and the scores obtained with the MIRT model. These correlations, shown in Table 8, indicate that, with the sole exception of the sum scores in openness and extraversion, in the other eight cases the scores obtained with the G-DINA provide more favourable congruence evidence, with the average difference being .06 (min = −.06, max = .14).

Lastly, Table S1 shows that attribute classifications and parameter estimates were virtually identical when using initial values ranging from 1e-3 to 1e-9. Less extreme initial values led to more disparate results, particularly the noninformative initial value of .5, which was unable to capture the intended meaning of the attributes.

4 | DISCUSSION

The present paper has focused on enhancing diagnostic FC assessments. Recent attention has increasingly been directed towards evaluating noncognitive traits using the DCM framework (e.g., de la Torre et al., 2015; Liu & Shi, 2020), where measurement can be prone to certain biases such as acquiescence and social desirability. The FC item format represents a widely explored approach in the previous literature to mitigate this type of bias (Brown, 2016). Building on Huang's (2023) introduction of the first DCM specifically designed for FC blocks, we proposed adapting the well-established G-DINA model to these assessments. This adaptation offers several advantages, including a general block response function capable of handling block responses from statements with varying discriminatory power, and the possibility of easily integrating the plethora of analyses developed within the DCM framework for evaluating reliability, model fit and other psychometric features. Furthermore, we have shown that the FC-DCM can be considered a restricted version of the G-DINA model. Consequently, it can be reparametrized as the G-DINA model, facilitating more efficient estimation algorithms and enabling the exploration of additional analyses mentioned earlier.

To effectively adapt the G-DINA model for FC assessments, two major practical considerations must be addressed. First, the Q-matrix must adhere to the criteria established by Gu and Xu (2020) to ensure the generic identifiability of the G-DINA model when all blocks measure two attributes (see Equation 4). Second, 'polarity-specific monotonicity' needs to be addressed either by imposing monotonicity constraints or by using appropriate initial values to properly reflect the direction of the

TABLE 7 Estimated probabilities for two of the blocks under the G-DINA model and the FC-DCM in the real data illustration.

	P (00)	*P* (10)	*P* (01)	*P* (11)
Block 3: Prob. of choosing 'Have a vivid imagination' over 'Am relaxed most of the time'				
GDINA	.66*	.42	.99	.74*
FC-DCM	.50	.38	.94	.50
The first attribute refers to Neuroticism and the second one to Openness.				
Block 22: Prob. of choosing 'I am exacting in my work' over 'Feel other's emotions'				
GDINA	.42	.34	.53	.45
FC-DCM	.50	.32	.58	.50
The first attribute refers to Agreeableness and the second one to Conscientiousness.				

*This is a G-DINA estimated probability, fixed at 0.50 in the FC-DCM model, whose 95% confidence interval does not include 0.50.

TABLE 8 Point-biserial correlations between the DCM classifications and the sum and MUPP-2PL scores in the real data illustration.

	O	C	E	A	N
Sum scores					
G-DINA	.64	.77	.73	.68	.73
FC-DCM	.70	.68	.71	.61	.69
MUPP-2PL scores					
G-DINA	.76	.80	.80	.77	.78
FC-DCM	.71	.72	.70	.63	.67

Abbreviations: A, agreeableness; C, conscientiousness; E, extraversion; MUPP-2PL, multi-unidimensional pairwise preference two-parameter logistic model; N, neuroticism; O, openness.

statements and the attributes. MCMC estimation algorithms could also be employed, incorporating constrained sampling to ensure the identifiability of the model. Among the three options available for implementation within the G-DINA model for FC assessments, we opted for the second strategy due to its simplicity and straightforward implementation in marginal maximum likelihood estimation. This approach is considerably less computationally intensive than MCMC estimation. Future research could focus on comparing estimation procedures in these FC settings. However, based on previous studies (e.g., Sorrel et al., 2023) and the performance of the initial value-based strategy tested in the present research, we do not anticipate significant differences between the three approaches. Namely, in the simulation study, the use of initial values proved effective, with only 6 out of the 57,594 replicas exhibiting an inversion in direction of an attribute, and all of these occurring under the most challenging conditions (10 attributes and 30 blocks). Furthermore, the real data illustration showed the robustness of attribute classifications and parameter estimates when using initial values ranging from 1e-3 to 1e-9. To streamline the process of specifying these initial values and facilitate the application of FC assessments in R using the G-DINA model, we have included the 'FCGDINA' function in the cdmTools package (Nájera, Sorrel, & Abad, 2025).

The simulation study showed that the G-DINA model effectively recovers attribute profiles, block parameters and attribute correlations, consistently outperforming the FC-DCM, except in the scenario where all statements have identical guessing and slip parameters. This particular condition aligns with the constraints of the FC-DCM but is challenging to justify in most practical applications. Additionally, the simulation results suggest that, whenever feasible, incorporating heteropolar blocks in FC assessments has a beneficial overall effect, particularly for the recovery of attribute correlations. This is in line with the reduction of ipsativity in previous studies in the area of traditional IRT (e.g., Bürkner et al., 2019; Frick et al., 2023). The real data illustration further supported these findings, showing that the G-DINA model not only provided a better model fit but also yielded block parameters that considerably deviated from the constraints imposed by the FC-DCM. Notably, it is important to emphasize that the G-DINA model (and even the true generating model) benefits from pairing statements with unequal discrimination, as the probabilities of endorsement for latent groups with no clear preference for any statement deviate from .5, thus becoming more informative. It is common to observe that, among the statements included in a block, one is preferred over another due to the reasons outlined in the empirical example (e.g., related to trait level or social desirability). The G-DINA model provides a framework for modelling the data in such cases. The analysis of relative fit helps in determining when it becomes necessary to apply this more general model. The differences in the estimated parameters will allow for a more precise classification of the examinees.

This study concentrated on between-statement multidimensionality, where each block is composed of two unidimensional statements. Huang (2023) extended the FC-DCM to address within-statement multidimensionality, where each statement can measure multiple attributes. To achieve this purpose, a conjunctive or disjunctive rule was employed to determine the ideal response for each statement. Further research is needed to adapt this approach to the G-DINA model and explore whether rules beyond the conjunctive/disjunctive dichotomy can be incorporated. The measurement of noncognitive variables such as personality is often accompanied by an internal structure that is difficult to discern, frequently complex due to interstitiality and item wording, which can create overlaps between dimensions or hinder the clear identification of constructs (Abad et al., 2018; Nájera, Abad, & Sorrel, 2025). This makes it particularly interesting in this field to explore the application of exploratory models. There is a tradition of studies in CDM on this topic that can serve as a starting point for this line of work (e.g., Chen et al., 2021; Gu & Xu, 2021; Xiong et al., 2024). Finally, the correspondence between attribute profile classifications and MIRT scores, as demonstrated in the real data example, suggests that future research could explore ways to bridge these measurement frameworks. Such integration could enhance the interpretation of outcome information and support more informed decision-making in various fields, be it education, clinical psychology, organizational psychology and personality profiling, among others. To conclude, while the recovery of the proposed model is studied across a wide range of scenarios in the simulation study and an empirical illustration is provided, it will be crucial to test the model on different datasets, which will undoubtedly help identify additional aspects to be considered in future research.

AUTHOR CONTRIBUTIONS

Pablo Nájera: Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; writing – original draft; software; formal analysis; data curation. **Rodrigo S. Kreitchmann:** Writing – review and editing; conceptualization; investigation; methodology; formal analysis; software; data curation. **Scarlett Escudero:** Writing – review and editing; conceptualization; investigation; software; methodology; data curation. **Francisco J. Abad:** Funding acquisition; writing – review and editing; conceptualization; investigation; methodology; project administration; validation. **Jimmy de la Torre:** Validation; conceptualization; writing – review and editing. **Miguel A. Sorrel:** Funding acquisition; conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; project administration; supervision; formal analysis.

ACKNOWLEDGEMENTS

This work was funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU under the project 'Computerized adaptive tests based on new assessment formats' (reference: PID2022-137258NB-I00), Universidad Pontificia Comillas (2024 Call for Funding of Internal Research Projects: 'Advancements in cognitive diagnosis models for formative assessments') and the UAM-IIC Chair Psychometric Models and Applications. Portions of these findings were presented at the International Meeting of the Psychometric Association 2024 (July 2024; Prague, Czech Republic) and the XVIII Congreso de Metodología de las Ciencias Sociales y de la Salud (September 2024; Seville, Spain).

CONFLICT OF INTEREST STATEMENT

This study was not preregistered. We have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

All the code and data from this research are publicly available at https://osf.io/h6x9e/.

ORCID

Pablo Nájera https://orcid.org/0000-0001-7435-2744
Rodrigo S. Kreitchmann https://orcid.org/0000-0001-5199-9828
Scarlett Escudero https://orcid.org/0009-0008-8836-2933
Francisco J. Abad https://orcid.org/0000-0001-6728-2709
Jimmy de la Torre https://orcid.org/0000-0002-0893-3863
Miguel A. Sorrel https://orcid.org/0000-0002-5234-5217

REFERENCES

Abad, F. J., Sorrel, M. A., Garcia, L. F., & Aluja, A. (2018). Modeling general, specific, and method variance in personality measures: Results for ZKA-PQ and NEO-PI-R. *Assessment*, *25*(8), 959–977. https://doi.org/10.1177/1073191116667547

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Apple, M. T., & Neff, P. (2012). Using Rasch measurement to validate the Big Five factor marker questionnaire for a Japanese university population. *Journal of Applied Measurement*, *13*(3), 276–296.

Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment*, *21*(3), 260–271. https://doi.org/10.1177/1073191114528029

Brown, A. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, *81*(1), 135–160. https://doi.org/10.1007/s11336-014-9434-9

Bunji, K., & Okada, K. (2020). Joint modeling of the two-alternative multidimensional forced-choice personality measurement and its response time by a Thurstonian D-diffusion item response model. *Behavior Research Methods*, *52*(3), 1091–1107. https://doi.org/10.3758/s13428-019-01302-5

Bürkner, P.-C., Schulte, N., & Holling, H. (2019). On the statistical and practical limitations of Thurstonian IRT models. *Educational and Psychological Measurement*, *79*(5), 1–28. https://doi.org/10.1177/0013164419832063

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

Chen, Y., Liu, Y., Culpepper, S. A., & Chen, Y. (2021). Inferring the number of attributes for the exploratory DINA model. *Psychometrika*, *86*(1), 30–64. https://doi.org/10.1007/s11336-021-09750-9

Cheung, M. W.-L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling*, *9*, 55–77. https://doi.org/10.1207/S15328007SEM0901_4

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633–665. https://doi.org/10.1007/s11336-009-9125-0

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*(3), 267–307. https://doi.org/10.1207/s15327043hup1803_4

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353. https://doi.org/10.1007/BF02295640

de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, *20*, 89–97. https://doi.org/10.1016/j.pse.2014.11.001

de la Torre, J., & Sorrel, M. A. (2017). *Attribute classification accuracy improvement: Monotonicity constraints on the G-DINA model [Conference session]*. Annual Meeting of Psychometric Society, Zurich, Switzerland.

de la Torre, J., van der Ark, L. A., & Rossi, G. (2015). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296. https://doi.org/10.1080/07481756.2017.1327286

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling*, *16*, 364–381. https://doi.org/10.1080/10705510902751374

Frick, S., Brown, A., & Wetzel, E. (2023). Investigating the normativity of trait estimates from multidimensional forced choice data. *Multivariate Behavioral Research*, *58*(1), 1–29. https://doi.org/10.1080/00273171.2021.1938960

García, P. E., Olea, J., & de la Torre, J. (2014). Application of cognitive diagnosis models to competency-based situational judgment tests. *Psicothema*, *26*(3), 372–377. https://doi.org/10.7334/psicothema2013.322

George, A. C., Robitzsch, A., Kiefer, T., Gross, J., & Uenlue, A. (2016). The R Package CDM for cognitive diagnosis models. *Journal of Statistical Software*, *74*(2), 1–24. https://doi.org/10.18637/jss.v074.i02

Gu, Y., & Xu, G. (2020). Partial identifiability of restricted latent class models. *The Annals of Statistics*, *48*(4), 2082–2107. https://doi.org/10.1214/19-AOS1878

Gu, Y., & Xu, G. (2021). Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinnica*, *31*, 449–472. https://doi.org/10.5705/ss.202018.0410

Hong, C.-Y., Chang, Y.-W., & Tsai, R.-C. (2016). Estimation of generalized DINA model with order restrictions. *Journal of Classification*, *33*, 460–484. https://doi.org/10.1007/s00357-016-9215-5

Hontangas, P. M., de La Torre, J., Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, *39*(8), 598–612. https://doi.org/10.1177/0146621615585851

Huang, H.-Y. (2023). Diagnostic classification model for forced-choice items and noncognitive tests. *Educational and Psychological Measurement*, *83*(1), 146–180. https://doi.org/10.1177/00131644211069906

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, *71*(2), 407–419. https://doi.org/10.1177/0013164410388832

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258–272. https://doi.org/10.1177/01466210122032064

Köhn, H. F., & Chiu, C.-Y. (2017). A procedure for assessing the completeness of the Q-matrices of cognitively diagnostic tests. *Psychometrika*, *82*, 112–132. https://doi.org/10.1007/s11336-016-9536-7

Kreitchmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., & Morillo, D. (2019). Controlling for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items. *Frontiers in Psychology*, *10*, 2309. https://doi.org/10.3389/fpsyg.2019.02309

Kreitchmann, R. S., Sorrel, M. A., & Abad, F. J. (2023). On bank assembly and block selection in multidimensional forced-choice adaptive assessments. *Educational and Psychological Measurement*, *83*(2), 294–321. https://doi.org/10.1177/00131644221087986

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Liu, R., & Shi, D. (2020). Using diagnostic classification models in psychological rating scales. The Quantitative Methods for. *Psychology*, *16*(5), 442–456. https://doi.org/10.20982/tqmp.16.5.p442

Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, *93*(14), 1–26. https://doi.org/10.18637/jss.v093.i14

Ma, W., & Jiang, Z. (2021). Estimating cognitive diagnosis models in small samples: Bayes modal estimation and monotonic constraints. *Applied Psychological Measurement*, *45*(2), 95–111. https://doi.org/10.1177/0146621620977681

Morillo, D., Leenen, I., Abad, F. J., Hontangas, P., de la Torre, J., & Ponsoda, V. (2016). A dominance variant under the multi-unidimensional pairwise-preference framework model formulation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, *40*(7), 500–516. https://doi.org/10.1177/0146621616662226

Nájera, P., Abad, F. J., & Sorrel, M. A. (2025). Is exploratory factor analysis always to be preferred? A systematic comparison of factor analytic techniques throughout the confirmatory–exploratory continuum. *Psychological Methods*, *30*(1), 16–39. https://doi.org/10.1037/met0000579

Nájera, P., Sorrel, M. A., & Abad, F. A. (2025). cdmTools: Useful tools for cognitive diagnosis modeling. R package version 1.0.6. https://CRAN.R-project.org/package=cdmTools

Paulhus, D. (1991). Measurement and control of response bias. In J. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*, Vol. 1 (pp. 17–59). Academic Press. https://doi.org/10.1016/b978-0-12-590241-0.50006-x

Paulsen, J., & Valdivia, D. S. (2021). Examining cognitive diagnostic modeling in classroom assessment conditions. *The Journal of Experimental Education*, *90*(4), 916–933. https://doi.org/10.1080/00220973.2021.1891008

Peng, S., Wang, D., Gao, X., Cai, Y., & Tu, D. (2019). The CDA-BPD: retrofitting a traditional borderline personality questionnaire under the cognitive diagnosis model framework. *Journal of Pacific Rim Psychology*, *13*, e22. https://doi.org/10.1017/prp.2019.14

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing, 124(125.10, pp. 1–10).

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revuelta, J., Halty, L., & Ximénez, C. (2018). Validation of a questionnaire for personality profiling using cognitive diagnostic modeling. *The Spanish Journal of Psychology*, *21*, E63. https://doi.org/10.1017/sjp.2018.62

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 1–17. https://doi.org/10.1080/15366367.2018.1435104

Sorrel, M. A., Escudero, S., Nájera, P., Kreitchmann, R. S., & Vázquez-Lira, R. (2023). Exploring approaches for estimating parameters in cognitive diagnosis models with small sample sizes. *Psychiatry*, *5*(2), 336–349. https://doi.org/10.3390/1435104

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, *19*(3), 506–532. https://doi.org/10.1177/1094428116630065

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, *29*, 184–203. https://doi.org/10.1177/0146621604273988

Tan, Z., de la Torre, J., Ma, W., Huh, D., Larimer, M. E., & Mun, E.-Y. (2023). A tutorial on cognitive diagnosis modeling for characterizing mental health symptom profiles using existing item responses. *Prevention Science*, *24*, 480–492. https://doi.org/10.1007/s11121-022-01346-8

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H.-H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement*, *43*(4), 255–271. https://doi.org/10.1177/0146621618765721

Xi, C., Cai, Y., Peng, S., Lian, J., & Tu, D. (2019). A diagnostic classification version of Schizotypal Personality Questionnaire using diagnostic classification models. *International Journal of Methods in Psychiatric Research*, *29*(1), e1807. https://doi.org/10.1002/mpr.1807

Xiong, J., Luo, Z., Luo, G., Yu, X., & Li, Y. (2024). An exploratory Q-matrix estimation method based on sparse non-negative matrix factorization. *Behavior Research Methods*, *56*(7), 7647–7673. https://doi.org/10.3758/s13428-024-02442-z

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, *81*(3), 625–649. https://doi.org/10.1007/s11336-015-9471-z

Zhang, J., Cui, S., Xu, Y., Cui, T., Barnhart, W. R., Ji, F., Nagata, J. M., & He, J. (2024). Introducing diagnostic classification modeling as an unsupervised method for screening probable eating disorders. *Assessment*, *32*(3), 405–416. https://doi.org/10.1177/10731911241247483

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nájera, P., Kreitchmann, R. S., Escudero, S., Abad, F. J., de la Torre, J., & Sorrel, M. A. (2025). A general diagnostic modelling framework for forced-choice assessments. *British Journal of Mathematical and Statistical Psychology*, *00*, 1–23. https://doi.org/10.1111/bmsp.12393