Psychometrika (2025), **90**, 1319–1345
doi:10.1017/psy.2025.10036

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Teamwork Cognitive Diagnostic Modeling

Peida Zhan, Zhimou Wang, Gaohong Chu and Haixin Qiao

School of Psychology, Zhejiang Normal University, Jinhua City, China

Corresponding author: Peida Zhan; Email: pdzhan@gmail.com

(Received 31 December 2024; revised 9 July 2025; accepted 16 July 2025; published online 8 August 2025)

Abstract

Teamwork relies on collaboration to achieve goals that exceed individual capabilities, with team cognition playing a key role by integrating individual expertise and shared understanding. Identifying the causes of inefficiencies or poor team performance is critical for implementing targeted interventions and fostering the development of team cognition. This study proposes a teamwork cognitive diagnostic modeling framework comprising 12 specific models—collectively referred to as Team-CDMs—which are designed to capture the interdependence among team members through emergent team cognitions by jointly modeling individual cognitive attributes and a team-level construct, termed *teamwork quality*, which reflects the social dimension of collaboration. The models can be used to identify strengths and weaknesses in team cognition and determine whether poor performance arises from cognitive deficiencies or social issues. Two simulation studies were conducted to assess the psychometric properties of the models under diverse conditions, followed by a teamwork reasoning task to demonstrate their application. The results showed that Team-CDMs achieve robust parameter estimation, effectively diagnose individual attributes, and assess teamwork quality while pinpointing the causes of poor performance. These findings underscore the utility of Team-CDMs in understanding, diagnosing, and improving team cognition, offering a foundation for future research and practical applications in teamwork-based assessments.

Keywords: cognitive diagnosis; collaborative problem-solving; diagnostic classification model; team cognition; teamwork

1. Introduction

Teamwork involves the collaboration of individuals to achieve shared goals that are difficult to accomplish alone (Salas et al., 1992). Studying teamwork explores the cognitive and behavioral dynamics of interpersonal interactions, enhancing our understanding of cooperation (Salas et al., 2017). This shift in focus from individual cognition to teamwork is evident in emerging research on collaborative problem-solving (Li et al., 2023; OECD, 2017), group intelligence (Burton et al., 2024; Woolley et al., 2010), social learning (Pan et al., 2022), and team creativity (Lu et al., 2023).

The effectiveness of teamwork is essential for teams to outperform individuals. Effective teamwork enhances efficiency, expands knowledge, fosters innovation, reduces errors, and boosts satisfaction, while ineffective teamwork often underperforms individual efforts (Austin, 2003; Hao et al., 2017). Achieving effective teamwork is challenging, as it relies heavily on team cognition—team-level constructs like shared understanding and the collective organization of knowledge—that emerge from

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

cognitive interactions among members and play a critical role in shaping performance (Cannon-Bowers et al., 1993; DeChurch & Mesmer-Magnus, 2010; Klimoski & Mohammed, 1994; Salas et al., 2017). Even teams of highly skilled individuals may struggle to perform well, highlighting the importance of considering team cognition alongside individual cognition (Devine & Philips, 2001; Woolley et al., 2015). Thus, promoting effective teamwork requires objective measures of team performance and the ability to identify strengths and weaknesses in team cognition, including team members' individual cognitions and the interactions or interdependencies among them (Salas et al., 2007, 2017).

In addition to using total scores from self- and expert-report questionnaires to evaluate teamwork performance (Andersson et al., 2017), large-scale assessment programs such as PISA and ATC21S have adopted computer-based collaborative problem-solving tasks to assess team performance. This shift has prompted researchers to explore new psychometric models for analyzing teamwork response data (e.g., Andrews et al., 2017; Li et al., 2023; von Davier & Halpin, 2013; Wilson et al., 2017; Yuan et al., 2019). For instance, Wilson et al. (2017) incorporated the team-level random effect into the Rasch model to characterize interdependence in members' response probabilities, assuming that both individual member ability and team-level random effect jointly influence performance in collaborative tasks; Yuan et al. (2019) decomposed collaborative problem-solving behaviors into independent actions (solely influenced by single-member ability) and collaborative actions (jointly determined by both members' abilities), analyzing them through a two-dimensional Rasch model; Li et al. (2023) employed a two-dimensional item response theory model to assess individual cognitive ability and social competence, revealing a moderate positive correlation between them. One of the central premises of these models is that the evaluation of teamwork performance should integrate both cognitive and social dimensions. This dual-dimensional framework recognizes that effective teamwork depends not only on individuals' cognitive abilities but also on the social interactions and the resulting interdependencies among team members. However, most, if not all, existing models are constructed within the item response theory framework and provide only an overall measurement of individuals' teamwork abilities (e.g., the collaborative problem-solving ability) without offering a fine-grained diagnosis of individual cognitive attributes and assessing corresponding team cognitions. Consequently, these models fall short in identifying specific causes of ineffective teamwork and poor performance, such as whether issues stem from a lack of knowledge or skills among team members or problems within the collaborative process. This limitation makes it challenging to design targeted team-building interventions or provide support for developing the team cognition necessary for effective teamwork (Lacerenza et al., 2018).

In recent decades, cognitive diagnosis or diagnostic classification (Rupp et al., 2010) has become a widely used assessment paradigm in psychology, education, and behavioral sciences. Designed to identify individuals' mastery of cognitive attributes (e.g., understanding, knowledge, and skills) and provide targeted feedback to support cognitive development (Tang & Zhan, 2021), numerous cognitive diagnostic models (CDMs; von Davier & Lee, 2019) have been developed and applied in diagnosing attributes, such as spatial rotation (Wang et al., 2018), pathological gambling (Templin & Henson, 2006), numerical literacy (Liang et al., 2021), rational number operations (Tang & Zhan, 2021), and problem-solving skills (Zhan & Qiao, 2022). However, traditional CDMs are limited to diagnosing individual cognitive attributes and cannot analyze teamwork response data, falling short of addressing the assessment needs of team cognition in collaborative settings. This limitation restricts their application in real-world scenarios where teamwork is a fundamental organizational form.

To the best of our knowledge, no CDM has been applied to teamwork-based diagnostic assessments, making this a novel and unexplored area in psychometrics. This study aims to propose a teamwork CDM framework comprising 12 specific models—collectively referred to as Team-CDMs—which are designed to capture the interdependence among team members through emergent team cognitions by jointly modeling individual cognitive attributes and a team-level construct, termed *teamwork quality*, which reflects the social dimension of collaboration. These Team-CDMs not only diagnose individual team members' cognitive attributes but also assess teamwork quality, offering valuable insights into the strengths and weaknesses of team cognition.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

This paper begins with a review of the log-linear model (LLM; Maris, 1999), a representative CDM for diagnosing individual cognitive attributes, which also serves as the foundation for the construction of the proposed Team-CDMs. Subsequently, 12 Team-CDMs are constructed by fully crossing three design factors: (1) teamwork response mode (teamwork-separated and teamwork-unified), (2) teamwork interaction mechanism (disjunctive, additive, and conjunctive), and (3) teamwork-effect specification (random-effect and covariate-modulated). A Bayesian approach is employed for model parameter estimation. The psychometric performance of the Team-CDMs is evaluated through two simulation studies, followed by an empirical study illustrating the practical application of the proposed models. This paper concludes with key findings and a discussion of directions for future research.

2. Teamwork cognitive diagnostic modeling

2.1. Modeling foundation: log-linear model

Existing CDMs for individual cognitive attributes are typically categorized into four types based on the condensation rules that define how latent attributes influence observed item responses: conjunctive, disjunctive, additive (or compensatory), and saturated models (Henson et al., 2009; Maris, 1999). Conjunctive models, such as the deterministic input, noisy "and" gate (DINA) model (Junker & Sijtsma, 2001), assume that all required attributes must be mastered to produce a correct response. Disjunctive models, such as the deterministic input, noisy "or" gate model (Templin & Henson, 2006), assume that mastery of at least one required attribute is sufficient. Additive models, such as the LLM (Maris, 1999), allow for compensatory relationships among attributes, enabling partial mastery to contribute to the response probability. Saturated models, such as the log-linear CDM (Henson et al., 2009) and the generalized DINA model (de la Torre, 2011), include all possible main effects and interactions among attributes, without imposing constraints related to specific condensation rules.

In this study, the LLM is taken as an example to illustrate the conceptualization of the proposed Team-CDMs. There are three primary reasons for choosing the LLM as the foundation instead of a saturated model. First, the LLM strikes a balance between the excessive complexity of saturated models and the simplicity of the DINA model.[1] It also serves as a commonly used restricted form of both the general diagnostic model (von Davier, 2008) and the log-linear CDM. This flexibility allows the proposed Team-CDM to be theoretically extended to a more general model if needed. Second, saturated models require large sample sizes for robust parameter estimation (Chiu et al., 2018), which is often impractical in small-scale teamwork projects. Third, the primary focus of this study is to extend traditional CDMs for analyzing teamwork response, rather than on defining how multiple cognitive attributes combine to determine the probability of a correct item response (i.e., condensation rules) within traditional CDMs. Therefore, the selection of the foundational model is not the central objective of the study. While we acknowledge that saturated CDMs can capture complex attribute interactions—particularly in complex cognitive tasks—we intentionally adopt simpler CDM forms to better isolate and emphasize the role of interdependence in teamwork settings.

Let Y_{ni} represent the observed response of participant n ($n = 1, \ldots, N$) to item i ($i = 1, \ldots, I$). The item response function of the LLM can be expressed as

$$logit\left(P\left(Y_{ni}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\gamma_{1ki}q_{ik}\alpha_{nk}, \tag{1}$$

where $P\left(Y_{ni}=1\right)$ is the probability that participant n correctly responds to item i and γ_{0i} and γ_{1ki} represent the intercept and the kth main effect of item i, respectively. Here, $\alpha_{nk} \in \{0,1\}$ denotes the cognitive attribute status of participant n on attribute k ($k = 1, \ldots, K$), with $\alpha_{nk} = 1$ if mastered and

[1]The simplicity of the DINA model implies that, for any given item, it can only classify respondents into two distinct groups: those who have mastered all requisite attributes for the item versus those who have not yet achieved mastery.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

$\alpha_{nk} = 0$ otherwise. $q_{ik} \in \{0,1\}$ is an element of the Q-matrix (Tatsuoka, 1983), with $q_{ik} = 1$ if attribute k is required for a correct response to item i, and $q_{ik} = 0$ otherwise.

As a compensatory model, the LLM assumes that, for the probability of a correct response, mastery of one required attribute may compensate for the lack of mastery in another required attribute. In the LLM, the lowest correct response probability of item i is

$$g_i = \frac{\exp(\gamma_{0i})}{1+\exp(\gamma_{0i})}, \tag{1a}$$

which represents the guessing probability of a correct response to item i when no required attributes are mastered. The correct response probability increases progressively with participants' mastery of the required attributes, reaching a highest probability defined as

$$h_i = \frac{\exp\left(\gamma_{0i} + \sum_{k=1}^{K} \gamma_{1ki} q_{ik}\right)}{1+\exp\left(\gamma_{0i} + \sum_{k=1}^{K} \gamma_{1ki} q_{ik}\right)}, \tag{1b}$$

which denotes the ideal probability of a correct response to item i when all required attributes are mastered. Typically, g_i and h_i are referred to as the item guessing and non-slipping parameters, respectively, with $s_i = 1 - h_i$ commonly known as the slipping parameter. Finally, the likelihood function of the LLM can be found in Section A3 in the Supplementary Material.

2.2. Teamwork cognitive diagnostic models

Before presenting the new models, we highlight several key points of this study. First, we focus on the dyadic team structure (i.e., either team t [$t = 1, 2, \ldots, T$] contains two members A(t) and B(t)), as it represents a foundational form of team configurations, such as "one-with-many" and leaderless teams (Kenny et al., 2006).

Second, teamwork tasks typically include individual response items (answered independently by each member) alongside teamwork response items (answered collaboratively by all team members). Since individual response items and teamwork response items measure the same latent constructs, this combination provides a benchmark for assessing the advantages or disadvantages of teamwork (Hao et al., 2017, 2019). Based on this, the observed responses to teamwork response items can take two primary modes: (1) the teamwork-unified response mode, where team members collaborate to produce a single unified response (e.g., Woolley et al., 2010; Zhang et al., 2024), and (2) the teamwork-separate response mode, where team members work together but submit individual responses separately (e.g., Li et al., 2023; Wilson et al., 2017; Yuan et al., 2019). Typically, the teamwork-separate response mode serves as a precursor to the teamwork-unified response mode, as members must first generate individual responses before synthesizing them into a unified submission. Note that this study does not aim to evaluate the relative advantages or disadvantages of the two response modes, but rather treats them as two distinct testing scenarios that may arise in practice. In such cases, for tasks adopting the teamwork-unified response mode, the complete response data comprise three item response matrices: two individual response matrices for each team member (i.e., $\mathbf{Y}_{\mathrm{A}(t)}$ and $\mathbf{Y}_{\mathrm{B}(t)}$) and one matrix for the teamwork-unified items (i.e., $\boldsymbol{Y}^{*}_{AB(t)}$). In contrast, for tasks using the teamwork-separate response mode, the response data consist of four matrices: two for individual responses (i.e., $\mathbf{Y}_{\mathrm{A}(t)}$ and $\mathbf{Y}_{\mathrm{B}(t)}$) and two for member-specific teamwork response matrices (i.e., $\boldsymbol{Y}^{*}_{A(t)}$ and $\boldsymbol{Y}^{*}_{B(t)}$). Separate models were constructed for each teamwork response mode.

2.2.1. Models for tasks adopting teamwork-separate response mode

The proposed model contains two components: one for individual item responses and another for teamwork-separate item responses. Let $Y_{A(t)i}$ and $Y_{B(t)i}$ be the observed individual responses of team members A and B of team t to individual response item i ($i = 1, \ldots, I_1$), respectively, and let $Y^{*}_{A(t)i}$ and

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

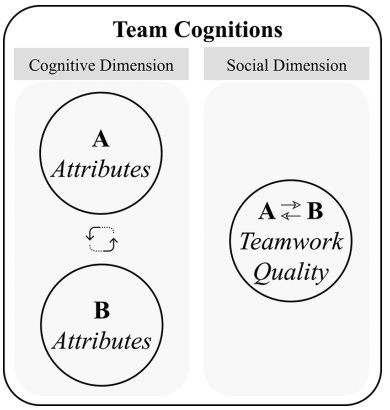


Figure 1. The structural diagram of team cognitions. *Note*: A and B are two team members.

$Y^*_{B(t)i}$ be the observed teamwork-separate responses of team members A and B of team t to teamwork response item i ($i = 1, \ldots, I_2$), respectively.

For individual response items, the LLM can be applied to each team member as

$$logit\left(P\left(Y_{\#(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\gamma_{1ki}q_{ik}\alpha_{\#(t)k}, \tag{2}$$

namely,

$$for\ teammate\ A, logit\left(P\left(Y_{A(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\gamma_{1ki}q_{ik}\alpha_{A(t)k}, \tag{2a}$$

$$for\ teammate\ B, logit\left(P\left(Y_{B(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\gamma_{1ki}q_{ik}\alpha_{B(t)k}, \tag{2b}$$

where # denotes team member A or B, $P\left(Y_{A(t)i}=1\right)$ and $P\left(Y_{B(t)i}=1\right)$ are the probabilities that team members A and B of team t correctly respond to item i, respectively, and $\alpha_{A(t)k}$ and $\alpha_{B(t)k}$ denote the individual cognitive attribute status of team members A and B on attribute k, respectively. This is equivalent to dividing the N participants equally into two groups of $T = N/2$ participants each, with each group responding the same number of I_1 items; q_{ik} is the element of the sub-Q-matrix corresponding to individual response items.

For teamwork response items, a major challenge in modeling teamwork responses lies in accurately representing the interaction or interdependence between team members. Drawing on the theoretical framework of team cognition (e.g., Cannon-Bowers et al., 1993; Salas et al., 2017), the present study conceptualizes *team cognition as a team-level construct emerging from the interaction of team members' cognitive attribute, capturing both team's collective cognitive capacity, such as shared knowledge and skills, and the quality of teamwork, such as the alignment, coordination, and communication among team members*. In other words, team cognitions are continuous, team-level constructs that correspond to individual cognitive attributes. Hence, the emergence of a specific team cognition is contingent upon the mastery of the relevant cognitive attribute by at least one team member, thereby establishing the foundation for shared understanding, and the quality of teamwork further shapes the level of team cognition, with high-quality teamwork enhancing it and low-quality teamwork diminishing it (DeChurch & Mesmer-Magnus, 2010; Lewis, 2003). Figure 1 displays the structural diagram of team cognitions in this study.

In such a case, a team cognition parameter can be introduced into Equation (2) to account for the interaction or interdependence between team members. Accordingly, the item response function for teamwork-separate responses can be specified as follows:

$$logit\left(P\left(Y^*_{\#(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\left(\gamma_{1ki}\alpha_{\#(t)k}+\Theta_{tk}\right)q^*_{ik}, \tag{3}$$

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

namely,

$$\text{for teammate A: } logit\left(P\left(Y^*_{A(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\left(\gamma_{1ki}\alpha_{A(t)k}+\Theta_{tk}\right)q^*_{ik}, \tag{3a}$$

$$\text{for teammate B: } logit\left(P\left(Y^*_{B(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\left(\gamma_{1ki}\alpha_{B(t)k}+\Theta_{tk}\right)q^*_{ik}, \tag{3b}$$

where Θ_{tk} represents the team cognition formed by team t on cognitive attribute k, q^*_{ik} is an element of the sub-Q-matrix corresponding to teamwork response items, and other parameters have been defined above. This model assumes that the probability of a correct teamwork-separated response is jointly determined by the individual member's mastery of the required cognitive attributes and the corresponding team cognition formed through member interactions.

Furthermore, this study hypothesizes that each team member's cognitive attributes contribute to teamwork responses indirectly by shaping shared team cognition through three distinct types of interaction mechanisms. The first interaction type, termed the *disjunctive interaction*, assumes that if any team member has mastered a given cognitive attribute, the team as a whole possesses that attribute as the basis for team cognition (cf. DeChurch & Mesmer-Magnus, 2010; Xue et al., 2018), with the degree of team cognition influenced by the quality of teamwork. The second type, termed the *additive interaction*, assumes that the degree of team cognition is influenced not only by teamwork quality but also by the cumulative cognitive attributes jointly mastered by team members (cf. Hao et al., 2017; Salas et al., 2017). The third type is termed the *conjunctive interaction*, which requires both team members to collectively master a specific cognitive attribute to establish the necessary foundation for developing corresponding team cognition, where the degree of team cognition is further influenced by the quality of teamwork (cf. Cooke, 2015).

Specifically, given these three distinct interaction mechanisms, the team cognition parameter can be further defined as

$$\Theta_{tk}=\begin{cases}\tau_t\left[1-\left(1-\alpha_{A(t)k}\right)\left(1-\alpha_{B(t)k}\right)\right], & \text{for disjunctive interaction}\\ \tau_t\left(\alpha_{A(t)k}+\alpha_{B(t)k}\right), & \text{for additive interaction}\\ \tau_t\alpha_{A(t)k}\alpha_{B(t)k}, & \text{for conjunctive interaction}\end{cases} \tag{4}$$

where $\tau_t \sim N\left(0,\sigma^2_\tau\right)$ is a team-level random effect parameter, referred to as the teamwork-effect parameter, which captures the quality of teamwork between the members of team t. Higher values of this parameter indicate that the team exhibits relatively higher teamwork quality compared with other teams in the sample population of participants. It reflects the relative positioning of a team's teamwork quality on a latent continuum, rather than a deterministic evaluation of good or bad performance. Specifically, when $\tau_t > 0$, it indicates that teamwork provides benefits, enhances team cognition, and increases the probability of correct teamwork responses. A value of $\tau_t = 0$ indicates ineffective collaboration among team members, meaning that no team cognition is formed. In this case, the probability of teamwork response is separately contributed by each team member, resembling a division of labor without actual collaboration. When $\tau_t < 0$, teamwork impairs team cognition, reducing the probability of correct teamwork responses.

For ease of understanding, consider a hypothetical example in which the individual cognitive attribute patterns of two members within team t are $(1, 0, 0)'$ and $(1, 1, 1)'$, respectively. Based on Equation (4), the resulting team cognition pattern varies by the type of team interaction mechanism: $(\tau_t, \tau_t, \tau_t)'$ for disjunctive interaction, $(2\tau_t, \tau_t, \tau_t)'$ for additive interaction, and $(\tau_t, 0, 0)'$ for conjunctive interaction, respectively. Then, the team cognition pattern, together with the individual-level cognitive patterns, jointly influences the probability of a correct teamwork response, as defined in Equation (3).

2.2.2. Models for tasks adopting teamwork-unified response mode

When the influence of team cognition is excluded, unlike the teamwork-separate response mode—in which each member's response probability is solely determined by their own mastery of the required cognitive attributes—the teamwork-unified response mode assumes that the team's unified response

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

probability is jointly influenced by the combined attribute mastery of both members. Therefore, the item response function for teamwork-unified responses can be specified as follows:

$$logit\left(P\left(Y^*_{AB(t)i}=1\right)\right)=\gamma_{0i}+\sum_{k=1}^{K}\left(\gamma_{1ki}\alpha_{A(t)k}+\gamma_{1ki}\alpha_{B(t)k}+\Theta_{tk}\right)q^*_{ik}, \tag{5}$$

where $P\left(Y^*_{AB(t)i}=1\right)$ denotes the correct teamwork-unified response probability of team t to item i. Equation (4) can be incorporated accordingly, and all other parameters are as previously defined.

This model illustrates how the team cognition formed from the two members, along with their individual mastery of cognitive attributes, jointly influences the probability of a correct teamwork-unified response. In such a case, both members will no longer have their own probability of a correct teamwork response. Noting that, given the interpersonal equivalence of cognitive attributes, the model does not differentiate the relative contributions of each member's attribute mastery to the probability of teamwork-unified responses. That is, it assumes equal contributions from both members (Yuan et al., 2019).

2.2.3. Explanatory models with covariates

In the proposed models above, the teamwork-effect parameter (τ_t) is a team-level random effect parameter that is freely estimated (cf. Wilson et al., 2017). Typically, information such as communications between team members can serve as an indicator of teamwork quality (e.g., Li et al., 2023; Woolley et al., 2010). Consequently, it is reasonable to assume that the teamwork-effect parameter depends on covariates reflecting teamwork quality, such as the frequency of communication between team members and the number of teamwork items answered consistently by both members.[2] Inspired by the explanatory models (Park et al., 2018; Wilson & De Boeck, 2004), the teamwork-effect parameter in Equation (4) can be further deconstructed as

$$\tau_t=\sum_{h=1}^{H}\tau_{1h}z_{ht}+\varepsilon_t, \tag{6}$$

where z_{ht} represents the teamwork covariate h ($h=1,2,\ldots,H$) of team t weighted by coefficient τ_{1h} and $\varepsilon_t\sim N\left(0,\sigma^2_\varepsilon\right)$ is the residual term.

Overall, this study proposes 12 Team-CDMs based on the teamwork response modes (separate and unified), the teamwork interaction mechanism (disjunctive, additive, and conjunctive), and the teamwork-effect specification (random-effect and covariate-modulated). Specifically, combining Equations (2)–(4) results in three models for teamwork-separate responses with different teamwork interaction mechanisms:

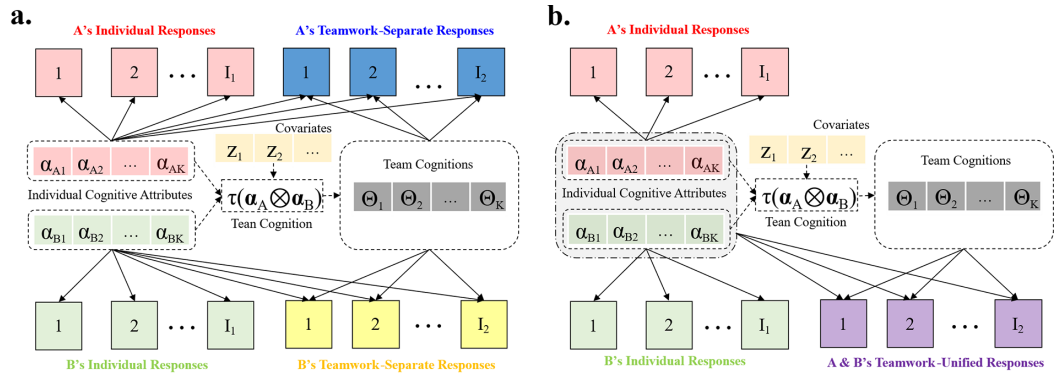
- *M1: Disjunctive model for teamwork-separate responses*;
- *M2: Additive model for teamwork-separate responses*;
- *M3: Conjunctive model for teamwork-separate responses.*

Combining Equations (2), (4), and (5) results in another three models for teamwork-unified responses with different teamwork interaction mechanisms:

- *M4: Disjunctive model for teamwork-unified responses*;
- *M5: Additive model for teamwork-unified responses*;
- *M6: Conjunctive model for teamwork-unified responses.*

[2]Before the two team members collaborate to submit a final team response, they may hold different individual responses. To capture this information, team members can be allowed to submit their own individual responses rather than being required to agree on the same response. The decision to submit a unified answer or not is left to the team members. This approach was applied in the empirical study conducted in this research.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press



c.

Model	Teamwork-Effect	Teamwork Response Mode	Interaction Mechanism
M1	Random-effect	Separate	Disjunctive
M2	Random-effect	Separate	Additive
M3	Random-effect	Separate	Conjunctive
M4	Random-effect	Unified	Disjunctive
M5	Random-effect	Unified	Additive
M6	Random-effect	Unified	Conjunctive
M7	Covariates	Separate	Disjunctive
M8	Covariates	Separate	Additive
M9	Covariates	Separate	Conjunctive
M10	Covariates	Unified	Disjunctive
M11	Covariates	Unified	Additive
M12	Covariates	Unified	Conjunctive

Figure 2. Graphical representations and summarization of teamwork cognitive diagnostic models (CDMs). *Note:* (a) Models for tasks adopting teamwork-separate response mode; (b) models for tasks adopting teamwork-unified response mode; (c) summarization of 12 teamwork CDMs with three modeling factors. M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; Θ = team cognition; A and B = team members A and B; *z* = covariate; ⊗ = interaction mechanism; I_1 = the number of individual response items; I_2 = the number of teamwork response items.

Deconstructing the random effect teamwork-effect parameter in the above six models through the inclusion of teamwork covariates yields six corresponding explanatory models:

- *M7: Disjunctive model for teamwork-separate responses with covariates*;
- *M8: Additive model for teamwork-separate responses with covariates*;
- *M9: Conjunctive model for teamwork-separate responses with covariates*;
- *M10: Disjunctive model for teamwork-unified responses with covariates*;
- *M11: Additive model for teamwork-unified responses with covariates*;
- *M12: Conjunctive model for teamwork-unified responses with covariates.*

To facilitate the presentation, the initial models (M1 ~ M6) are referred to as standard models collectively, whereas the subsequent models (M7 ~ M12) are designated as explanatory models. Figure 2 provides graphical representations and a summary of these proposed Team-CDMs. The likelihood function of the proposed models can be found in Section A3 in the Supplementary Material.

2.3. Bayesian parameter estimation

The fully Bayesian Markov Chain Monte Carlo (MCMC) algorithm is employed to estimate the parameters of the proposed models. This approach streamlines the estimation of intricate model parameters

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

by simulating the posterior distributions of item parameters and latent variables, thus enabling the subsequent inference of model parameters. The JAGS (Version 4.3.2; Plummer, 2015) is invoked from within the R software environment to implement the Bayesian MCMC algorithm. For detailed guidance on using JAGS for Bayesian MCMC estimation in CDMs, please refer to Zhan, Jiao, Man, et al. (2019).

Except for the teamwork-effect parameter, the prior distributions of the other parameters refer to those outlined in Zhan, Jiao, Man, et al. (2019). To begin with, item responses are assumed to be independently distributed following a Bernoulli distribution: $Y_{\#(t)i} \sim Bernoulli\left(P\left(Y_{\#(t)i}=1\right)\right)$ and $Y_{ti} \sim Bernoulli\left(P\left(Y_{ti}=1\right)\right)$. Imposing the monotonicity restriction, the priors of item parameters are specified $\gamma_{0i} \sim N\left(-2.197,1\right)$ and $\gamma_{1ki} \sim N^{+}\left(0,2^{2}\right)$. Let $\pi_c = P\left(\boldsymbol{\alpha}_c\right)$ be the marginal probability of attribute pattern c in the population (i.e., the mixing proportion), then $\boldsymbol{\pi} = \left(\pi_1,\ldots,\pi_C\right)'$ be a C-dimensional probabilistic vector of attribute patterns, and $\sum_{c=1}^{C}\pi_c = 1$, where $C = 2^K$ is the number of possible attribute patterns. The prior distribution of the attributes is set as: $c \sim Categorical\left(\boldsymbol{\pi}\right)$, $\boldsymbol{\pi} \sim Dirichlet\left(\boldsymbol{\lambda}\right)$, and $\boldsymbol{\lambda} = \left(1,\ldots,1\right)'$, indicating that participant n's (i.e., A(t)'s or B(t)'s) attribute pattern is assumed to follow a categorical distribution, with the probability of membership in the cth pattern, then $\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_c$. For standard Team-CDMs, the prior distribution of the teamwork-effect parameter is set as $\tau_t \sim N\left(0,\sigma_\tau^2\right)$ with hyper-prior $\sigma_\tau^2 \sim InvGamma\left(1,1\right)$. In contrast, for explanatory Team-CDMs, the prior distribution of the hth coefficient is set as $\tau_{1h} \sim N\left(0,1\right)$, and the prior distribution of the residual term is set as $\varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)$ with hyper-prior $\sigma_\varepsilon^2 \sim InvGamma\left(2,0.5\right)$.

Finally, the posterior mean is used as the estimated value of continuous parameters, such as item parameters and teamwork-effect parameters. The posterior mode is used as the estimated value of individual cognitive attributes.

3. Simulation study

We conducted two simulation studies to evaluate the proposed models. The first study assessed the psychometric performance of the 12 Team-CDMs under various simulated testing scenarios. The second study examined potential risks arising from omitting the teamwork-effect parameter or the whole team cognitions in the analysis of teamwork response data. It also investigated the negative consequences of analyzing only individual response items in a teamwork-based assessment and evaluated each model's performance when analyzing only collaborative response items.

3.1. Simulation Study 1

3.1.1. Design and data generation

In addition to using each of the 12 Team-CDMs as data generating models, two factors were manipulated: first, the number of teams (T) at four levels of 50, 100, 200, and 400, corresponding to a total of 100, 200, 400, and 800 participants; second, the number of individual and teamwork items ($I_1 = I_2$) at two levels: 15 and 30, corresponding to a total of 30 and 60 items. Consistent with the empirical study (see Section 4), five personal latent attributes ($K = 5$) were diagnosed, with the Q-matrices shown in Figure 3. Each Q-matrix for the task comprises two submatrices: one corresponding to individual response items and the other to teamwork response items. Each submatrix included at least one identity matrix for completeness, and each attribute was measured at least three times to further ensure model identifiability (Gu & Xu, 2020; Köhn & Chiu, 2017; Xu & Zhang, 2016). In addition, each attribute was measured approximately the same number of times to mitigate the potential impact of unbalanced attribute settings on parameter estimation results (Cheng, 2010).

Other parameters were generated based on specific distributions. Since the standard and explanatory models define the teamwork-effect parameters differently, we used two distinct methods to generate these parameters accordingly. For standard models (M1 ~ M6), a multivariate normal distribution was first used to generate two members' general abilities (e.g., the fluid intelligence in the empirical study)

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

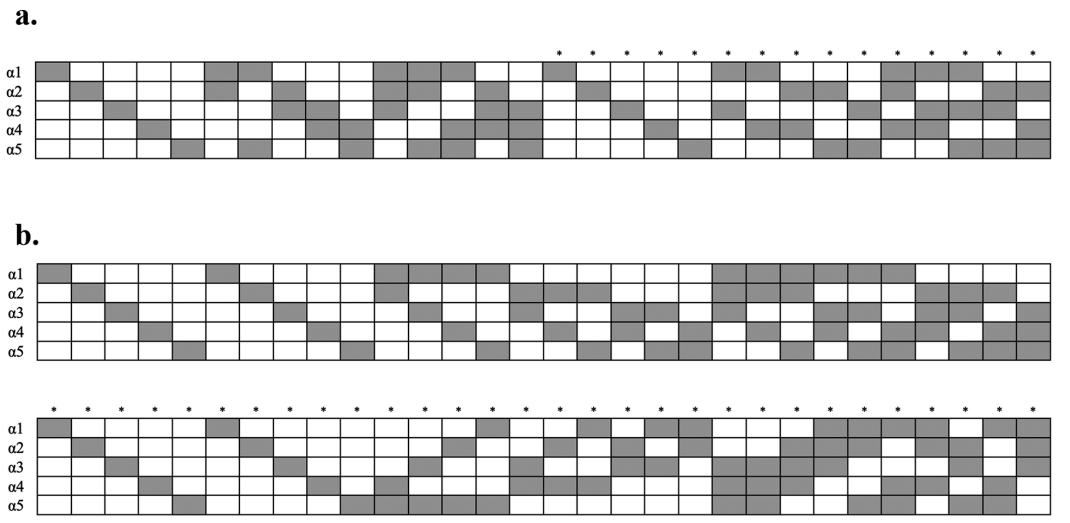


Figure 3. *K*-by-*I* Q' matrix for simulation study. *Note*: Gray means "1" and blank means "0". "*" denotes teamwork response items. (a) Both individual and teamwork response items contain 15 items. (b) Both individual and teamwork response items contain 30 items.

and their teamwork effect:

$$\begin{pmatrix} \boldsymbol{\theta}_A \\ \boldsymbol{\theta}_B \\ \boldsymbol{\tau} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 & 0.3 \\ 0.6 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{pmatrix} \right). \tag{7}$$

This setup assumes a moderate positive correlation between the general abilities of the two members, and a low positive correlation between their respective general abilities and the teamwork effect. For explanatory models (M7 ~ M12), $H = 3$ covariates (z_1, z_2, and z_3) were considered, all standardized for simplicity. Specifically, each covariate was generated from a standard normal distribution, with equal weights represented as $\boldsymbol{\tau}_1 = (0.5, 0.5, 0.5)'$. Then, a multivariate normal distribution was used to generate two members' general abilities and the residual term of teamwork effect:

$$\begin{pmatrix} \boldsymbol{\theta}_A \\ \boldsymbol{\theta}_B \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.5^2 \end{pmatrix} \right). \tag{8}$$

Under these settings, the three covariates account for 75% of the variance in the teamwork effect, with the variance of the teamwork-effect parameter still constrained to 1 (i.e., $\sigma_\tau^2 = 1$). Additionally, the correlation between the teamwork effect and each general ability is fixed at 0.3, consistent with the specifications in Equation (6).

Subsequently, a higher-order latent structural model was applied to both standard and explanatory models to generate individual cognitive attributes, enabling the incorporation of correlations among these attributes (de la Torre & Douglas, 2004), as

$$logit\left(P\left(\alpha_{\#(t)k} = 1\right)\right) = \lambda_{1k}\theta_{\#(t)} + \lambda_{0k}, \tag{9}$$

where λ_{0k} is the attribute intercept parameter (fixed at $\boldsymbol{\lambda}_0 = (-1.5, -1, 0, 1, 1.5)'$) and λ_{1k} is the attribute slope parameter (fixed at $\boldsymbol{\lambda}_1 = (1.5, 1.5, 1.5, 1.5, 1.5)'$). Then, the true mastery of each member on each attribute was generated from a Bernoulli distribution: $\alpha_{\#(t)k} \sim Bernoulli\left(P\left(\alpha_{\#(t)k} = 1\right)\right)$.

For item parameters, item intercept parameters were generated from $\gamma_{0i} \sim N\left(-2.197, 0.5^2\right)$ and item main-effect parameters were generated from $\gamma_{1ki} \sim N\left(\frac{4.394}{\sum_{k=1}^{K} q_{ik}}, 0.5^2\right)$. In this setting, g_i for all items follow

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

a positively skewed distribution (mean ≈ 0.1, minimum ≈ 0.05, maximum ≈ 0.45), and h_i for all items follow a negatively skewed distribution (mean ≈ 0.9, minimum ≈ 0.55, maximum ≈ 0.95). Consequently, the correlation between g_i and $1 - h_i$ follows a negative correlation, which is more realistic (Zhan, Jiao, Liao, et al., 2019).

Finally, the observed individual responses were generated as $Y_{\#(t)i} \sim Bernoulli\left(P\left(Y_{\#(t)i}=1\right)\right)$, where $P\left(Y_{\#(t)i}=1\right)$ is respectively defined in Equations (2); the observed teamwork-separate responses were generated as $Y^*_{\#(t)i} \sim Bernoulli\left(P\left(Y^*_{\#(t)i}=1\right)\right)$, where $P\left(Y^*_{\#(t)i}=1\right)$ is, respectively, defined in Equations (3); and the observed teamwork-unified responses were generated as $Y^*_{AB(t)i} \sim Bernoulli\left(P\left(Y^*_{AB(t)i}=1\right)\right)$, where $P\left(Y^*_{AB(t)i}=1\right)$ is defined in Equation (5). A total of 50 datasets were generated in each simulated condition.

3.1.2. Analysis

A total of 12 Team-CDMs were used to analyze the data generated under each model. For each simulated condition, 50 replications were conducted using two Markov chains with random starting points. After discarding the first 3,000 iterations of 5,000 as burn-in, 4,000 iterations (2,000 per chain) were retained for parameter inference. To assess convergence, the potential scale reduction factor (PSRF; Brooks & Gelman, 1998) was computed for each parameter. In this study, PSRF values were generally below 1.01 and all below 1.2, indicating good convergence under the specified settings.

The bias and root-mean-square error (RMSE) were computed to evaluate the recovery of continuous parameters (e.g., item parameters and the teamwork-effect parameter) as $bias\left(\widehat{x}\right) = \sum_{r=1}^{50}\frac{\widehat{x}_r - x_r}{50}$ and $RMSE\left(\widehat{x}\right) = \sqrt{\sum_{r=1}^{50}\frac{\left(\widehat{x}_r - x_r\right)^2}{50}}$, respectively, where x_r and $\widehat{x}_r$ is the generated and estimated values of the model parameter in replication r. In addition, the correlations between the generated and estimated values (denoted as Cor) of the model parameters were computed. The classification accuracy of attributes (CAA) and the classification accuracy of attribute patterns (CAP) were calculated to evaluate diagnostic classification accuracy. For CAA, the accuracy for each attribute was computed as $CAA_k = \frac{\sum_{r=1}^{50}\sum_{n=1}^{N} I\left(\widehat{\boldsymbol{\alpha}}_{nkr}=\boldsymbol{\alpha}_{nkr}\right)}{NR}$, where N represents the total number of participants (i.e., $2T$). Two types of CAPs were calculated: one based on 5 attributes for individual members and another based on 10 attributes for the two members of teams. The former evaluates the classification accuracy of cognitive attribute patterns for individual participants, whereas the latter treats the team as a whole, assessing the proportion of teams in which the cognitive attribute patterns of both members are accurately classified. Specifically, the CAP for individuals was computed as $CAP_{individuals} = \frac{\sum_{r=1}^{50}\sum_{n=1}^{N} I\left(\widehat{\boldsymbol{\alpha}}_{nk}=\boldsymbol{\alpha}_{nk}\right)}{NR}$, and for teams as $CAP_{Team} = \frac{\sum_{r=1}^{50}\sum_{t=1}^{T} I\left(\left(\widehat{\boldsymbol{\alpha}}_{A(t),}\ \widehat{\boldsymbol{\alpha}}_{B(t)}\right)'=\left(\boldsymbol{\alpha}_{A(t),}\ \boldsymbol{\alpha}_{B(t)}\right)'\right)}{TR}$.

3.1.3. Results

Table 1 summarizes the computation time of the proposed model across all simulated conditions. Computation time remains highly consistent across the 50 iterations within each condition, suggesting stable algorithmic performance. Statistical analysis indicates that the primary factors influencing computation time are data volume—specifically, the number of teams, the number of items, and the teamwork response mode. In contrast, the other manipulated variables have minimal impact on computational cost.

Figure 4 displays the RMSE values for the item guessing and non-slipping parameters, with detailed metrics for bias, RMSE, and correlation (Cor) provided in Table A1 in the Supplementary Material. Overall, item parameter recovery was satisfactory across all models and simulation conditions. Nearly all RMSE values were below 0.1, with the exception of a few non-slipping parameters associated with teamwork response items under small sample size conditions. The explanatory models (M7–M12) slightly outperformed the standard models (M1–M6) in parameter recovery. Similarly, models using the teamwork-separate mode (M1–M3 and M7–M9) demonstrated marginally better recovery than those

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

Table 1. Computation time of proposed models in Simulation Study 1

Teamwork response mode	*T*	*I*	Time (in minutes)
Unified	50	15	1
		30	2
	100	15	2
		30	3
	200	15	3.5
		30	7
	400	15	7
		30	14
Separate	50	15	1.5
		30	2.5
	100	15	3
		30	5
	200	15	5
		30	9
	400	15	10
		30	19

Note: T: the number of dyadic teams; I: the number of individual or teamwork response items. All computations were conducted on a Mac mini (Apple M4 Pro, 64 GB Memory) using multi-program parallel processing.

employing the teamwork-unified mode (M4–M6 and M10–M12). No notable differences were found in parameter recovery between models using different interaction mechanisms. Finally, increasing the number of teams and items, thereby augmenting the amount of available data, significantly improved the accuracy of item parameter recovery.

Figure 5 presents the classification accuracy of attributes (CAA) and attribute patterns (CAP). Detailed values are available in Table A2 in the Supplementary Material. First, CAA for each attribute was consistently high (above 0.9) across all models and simulated conditions. Second, the standard models and explanatory models exhibited comparable classification accuracy, indicating equivalent performance. Third, models using the teamwork-separate mode (M1–M3 and M7–M9) achieved better classification accuracy than those employing the teamwork-unified mode (M4–M6 and M10–M12). Fourth, no notable differences were found in classification accuracy between models using different interaction mechanisms. Finally, increasing the number of items significantly enhanced classification accuracy for both attributes and attribute patterns.

Figure 6 illustrates the recovery of the teamwork-effect parameter, with detailed values for this parameter and other related parameters (e.g., weights of covariates in the explanatory models) provided in Table A3 in the Supplementary Material. First, recovery accuracy improves as the number of items increases, whereas the number of teams has minimal impact. Second, explanatory models (M7–M12) show superior recovery compared with standard models (M1–M6), emphasizing the advantage of integrating covariates. Third, models using the teamwork-separate mode (M1–M3 and M7–M9) achieve better recovery than those employing the teamwork-unified mode (M4–M6 and M10–M12). Fourth, models based on the additive interaction mechanism (M1, M4, M7, and M10) exhibit slightly better recovery performance compared with those based on disjunctive and conjunctive interaction

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

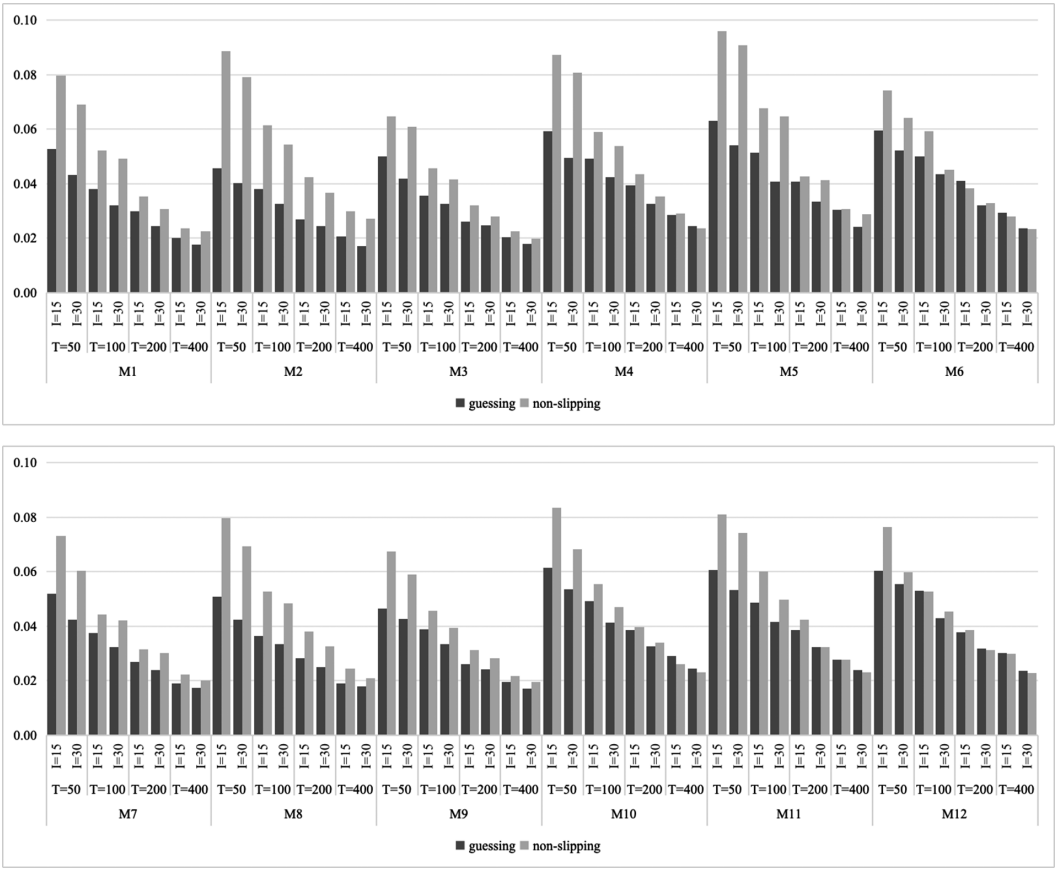


Figure 4. Root-mean-square error of item parameters in Simulation Study 1.
Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; *T*: the number of dyadic teams; *I*: the number of individual or teamwork response items; guessing: item guessing parameter, namely, minimum correct response probability of an item; non-slipping: item non-slipping parameter, namely, maximum correct response probability of an item.

mechanisms. Notably, the recovery of the teamwork quality parameters is poorest under the conjunctive interaction mechanism. One possible explanation is that, in this framework, these parameters affect the response probability only when two members simultaneously master specific required attributes—that is, the prerequisite for social behaviors is relatively stringent. Consequently, fewer teamwork-effect parameters have a meaningful impact on the response data, which restricts their estimation and results in poorer recovery performance.

Overall, the results of Simulation Study 1 demonstrated that the setup of the 12 proposed Team-CDMs was well justified, and the parameter estimates were accurate enough, indicating good psychometric performance across various testing scenarios.

3.2. Simulation Study 2

3.2.1. Design, data generation, and analysis

This study investigated the risks of omitting team cognitions or teamwork effects in analyzing collaborative response data. It also assessed the drawbacks of focusing solely on individual responses in

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

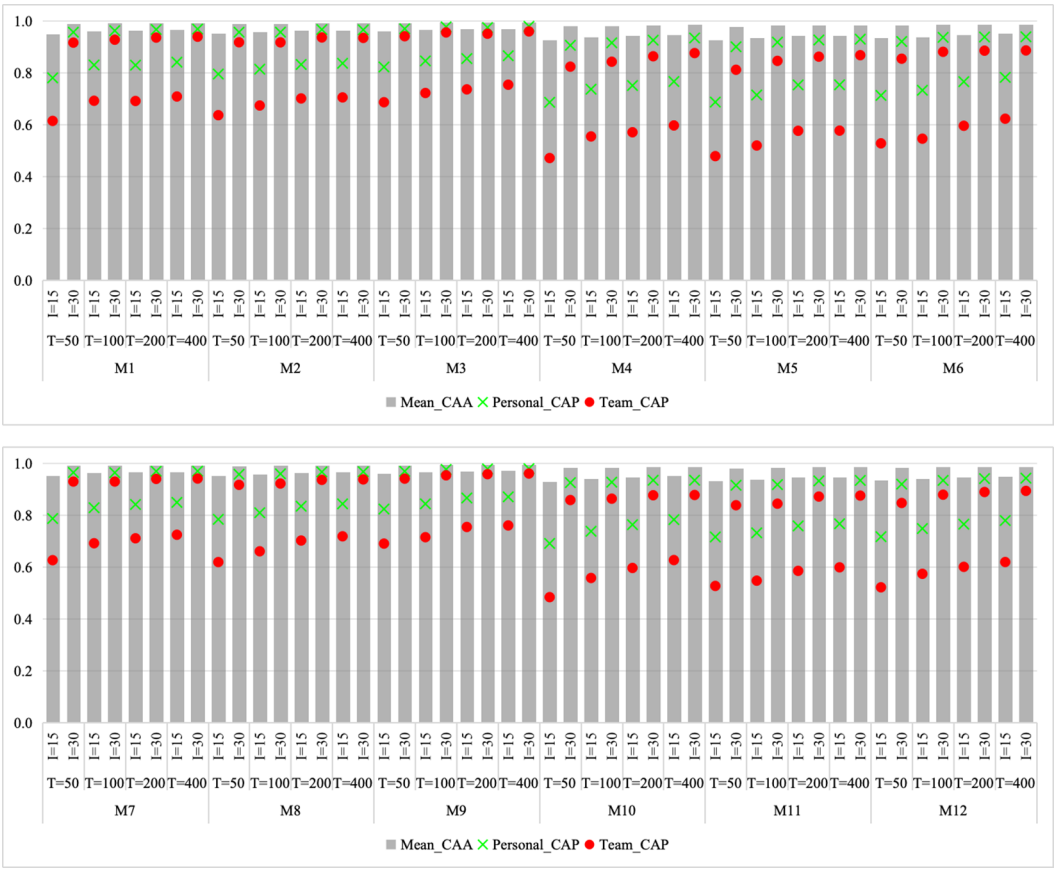


Figure 5. Classification accuracy of attributes and attribute patterns in simulation 1.
Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; *T*: the number of dyadic teams; *I*: the number of individual or teamwork response items; Mean_CAA: mean classification accuracy of five attributes; CAP Individuals: classification accuracy of attribute pattern for individual participants (five attributes); CAP team: classification accuracy of attribute pattern for teams (10 attributes).

teamwork assessments and evaluated model performance when analyzing only collaborative responses. Each of the 12 Team-CDMs was used to generate data, with the number of teams fixed at 100 and the number of individual or teamwork items set at 15. For simplicity, this study directly used 50 datasets generated under such specific simulated conditions in Simulation Study 1. This relatively limited testing condition was chosen to minimize the ceiling effect on model performance that could arise from an overabundance of information in the data.

For each dataset, five analytical models were applied: (1) the Team-CDM itself (i.e., M1–M8); (2) a constrained model with the teamwork-effect parameter fixed at 1, namely, ignoring the effect of teamwork quality on the team cognition and the probability of correct teamwork responses, and only the interaction between team members' attributes is considered (denoted as M1a–M8a); (3) a constrained model excluding the whole team cognition on the probability of correct teamwork responses (denoted as M1b–M8b), in which the interdependence between team members is ignored; (4) the Team-CDM

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

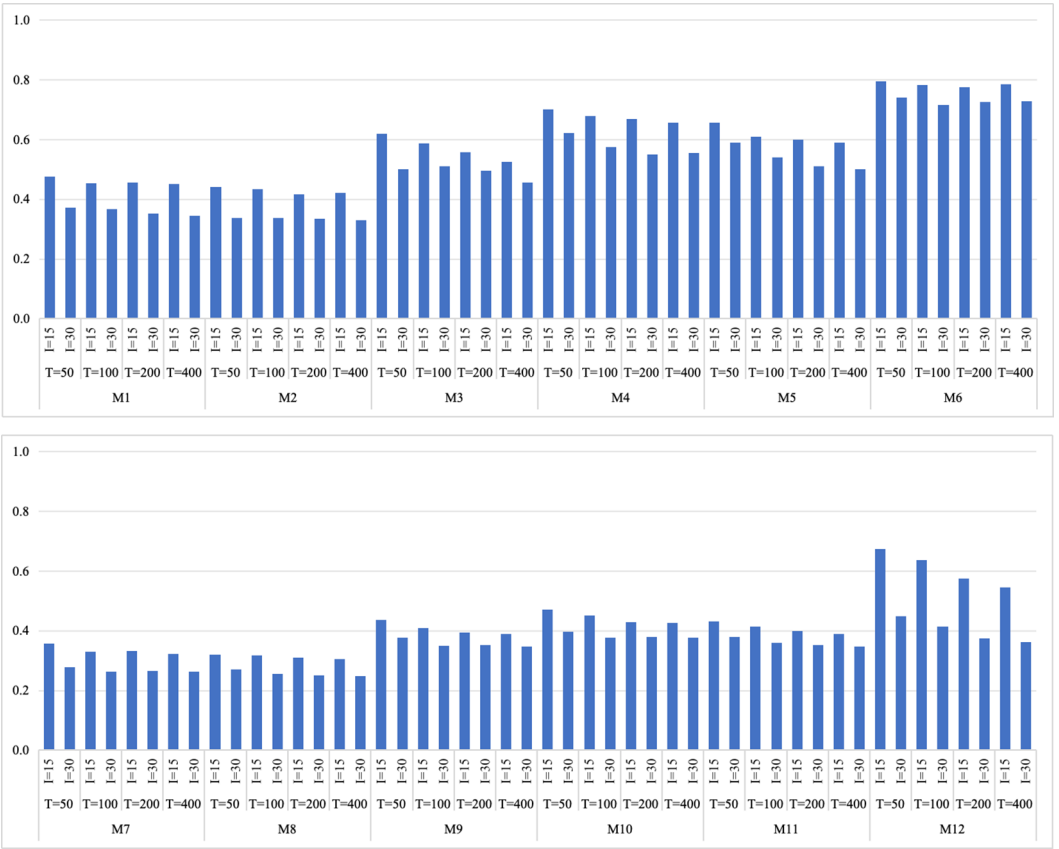


Figure 6. Root-mean-square error of the teamwork-effect parameter in Simulation Study 1.
Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; *T*: the number of dyadic teams; *I*: the number of individual or teamwork response items.

itself applied only to individual response data (denoted as M1-1 to M8-1; equivalent to LLM), and (5) the Team-CDM itself applied only to collaborative response data (denoted as M1-2 to M8-2).

Notably, fixing the teamwork-effect parameter makes the interaction mechanism the only factor shaping the model's structure. This effectively eliminates the distinction between the standard models and the corresponding explanatory models, resulting in equivalences such as M1a being equivalent to M7a, M2a–M8a, and so on. Furthermore, when team cognitions are excluded, all models with the same teamwork response mode become identical. For instance, M1b = M2b = M3b = M7b = M8b = M9b (models employing the teamwork-separate mode are also equivalent to the LLM) and M4b = M5b = M6b = M10b = M11b = M12b.

To compare the fit of the original and constrained models to the data, three model comparison indices were computed: deviance information criterion (DIC), widely applicable information criterion (WAIC), and leave-one-out cross-validation (LOO). Lower values of these indices indicate a better model fit. The remaining analytical procedures and evaluation metrics were consistent with those used in Simulation Study 1.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

3.2.2. Results

Table 2 presents comparative analyses of model-data fit statistics between the original and constrained models. The results demonstrate that the original parameterization (M1–M12) exhibits a consistently superior fit relative to both constrained variants (M1a–M12a and M1b–M12b) across all conditions. Moreover, all three fit indices successfully identified the theoretically optimal model.

Table A4 in the Supplementary Material reports the recovery of item parameters. First, the original models (M1–M12) achieve the lowest RMSE values across almost all conditions. Second, parameter recovery is relatively poor when analyzing individual response items or collaborative response items alone, though slightly better for the former. This highlights the limitations of focusing solely on individual response items in a collaborative task and the necessity of including them. Accurate diagnosis of individual cognitive attribute mastery—ensured through individual response items—is essential for isolating and evaluating the unique contribution of team cognitions in the teamwork response items. Third, fixing the teamwork-effect parameter (M1a–M12a) results in a significant increase in RMSE for collaborative response items, underscoring the negative impact of neglecting teamwork quality on parameter estimation, particularly for collaborative tasks. Fourth, excluding the whole team cognitions (M1b–M12b) produces RMSE values for individual response items comparable to the unconstrained models. However, for collaborative response items, RMSE values increase moderately. Interestingly, ignoring team cognitions entirely yields better item parameter recovery than considering interaction while neglecting teamwork quality. This finding underscores the importance of accounting for both interaction mechanisms and teamwork quality to accurately model team cognition.

Figure 7 summarizes the CAA and CAP across various models, with detailed CAA values for each attribute provided in Table A5 in the Supplementary Material. First, the Team-CDMs (M1–M12) achieve the highest classification accuracies. Second, classification accuracy is relatively poor when analyzing individual response items or collaborative response items alone. Third, models that fail to fully account for team cognition—whether by ignoring teamwork quality or excluding the whole team cognition—show reduced classification accuracy. Notably, models that exclude the whole team cognition entirely perform slightly better than those that neglect teamwork quality alone, consistent with the findings for item parameter recovery.

Overall, the results of Simulation Study 2 highlight that accurate modeling of team cognition requires integrating both the teamwork-effect and interaction mechanisms. Models that failed to fully account for these factors—either by neglecting teamwork quality or excluding interaction terms—exhibited poor parameter recovery and classification accuracy, particularly for collaborative response items. Additionally, analyzing individual or collaborative response items in isolation yielded similarly suboptimal results, emphasizing the importance of incorporating both individual contributions and collaborative processes into the modeling of teamwork.

4. Empirical example

4.1. Instrument

A teamwork reasoning task for measuring collective intelligence was implemented to demonstrate the application of the proposed Team-CDMs. The collective intelligence can be regarded as an important aspect that deeply influences a group's collaborative problem-solving ability. Similar to how individual intelligence equips an individual with the cognitive means to handle various tasks independently, collective intelligence emerges from the interactions and synergy among group members, endowing the group with a unique capacity to approach and resolve complex problems collaboratively (Woolley et al., 2010).

A shortened version of Raven's Advanced Progressive Matrices (APM; Zhan, Chen, Man, et al., 2024) was used as the foundation for this teamwork task to reduce administration time and alleviate

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

Table 2. Summary of model-data fits in Simulation Study 2

Data generation model	Analysis model	DIC	WAIC	LOO
M1	M1	5,282.313	5,248.170	5,272.791
	M1a	5,757.498	5,796.819	5,811.091
	M1b	5,714.603	5,763.838	5,777.082
M2	M2	5,132.905	5,086.911	5,114.825
	M2a	5,727.021	5,748.058	5,763.111
	M2b	5,739.723	5,778.573	5,791.836
M3	M3	5,437.219	5,428.611	5,446.490
	M3a	5,591.219	5,665.758	5,679.015
	M3b	5,581.275	5,629.759	5,641.352
M4	M4	4,173.722	3,951.738	3,992.763
	M4a	4,199.353	4,090.436	4,121.418
	M4b	4,201.097	4,074.465	4,103.756
M5	M5	4,162.845	3,950.756	3,991.876
	M5a	4,289.650	4,185.163	4,213.495
	M5b	4,305.781	4,178.402	4,205.636
M6	M6	4,079.387	3,978.165	4,010.149
	M6a	4,100.843	4,008.529	4,038.988
	M6b	4,109.181	3,997.218	4,026.268
M7	M7	5,281.971	5,261.924	5,281.793
	M7a	5,803.273	5,849.306	5,863.448
	M7b	5,754.713	5,822.062	5,833.450
M8	M8	5,151.633	5,126.801	5,146.621
	M8a	5,823.476	5,839.695	5,854.930
	M8b	5,817.223	5,863.535	5,875.192
M9	M9	5,431.320	5,431.473	5,447.864
	M9a	5,644.422	5,724.501	5,736.458
	M9b	5,628.931	5,684.492	5,696.326
M10	M10	4,124.650	3,940.747	3,974.252
	M10a	4,225.158	4,115.770	4,143.175
	M10b	4,218.895	4,103.296	4,130.398
M11	M11	4,131.842	3,944.951	3,981.360
	M11a	4,358.809	4,265.441	4,294.619
	M11b	4,371.080	4,244.254	4,271.990
M12	M12	4,043.172	3,925.996	3,962.287
	M12a	4,058.285	3,965.110	3,995.077
	M12b	4,069.117	3,952.930	3,981.518

Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; M1a–M12a: M1–M12 with fixed teamwork effect; M1b–M12b: M1–M12 without team cognitions; DIC: deviance information criterion; WAIC: widely available information criterion; LOO: leave-one-out cross-validation.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

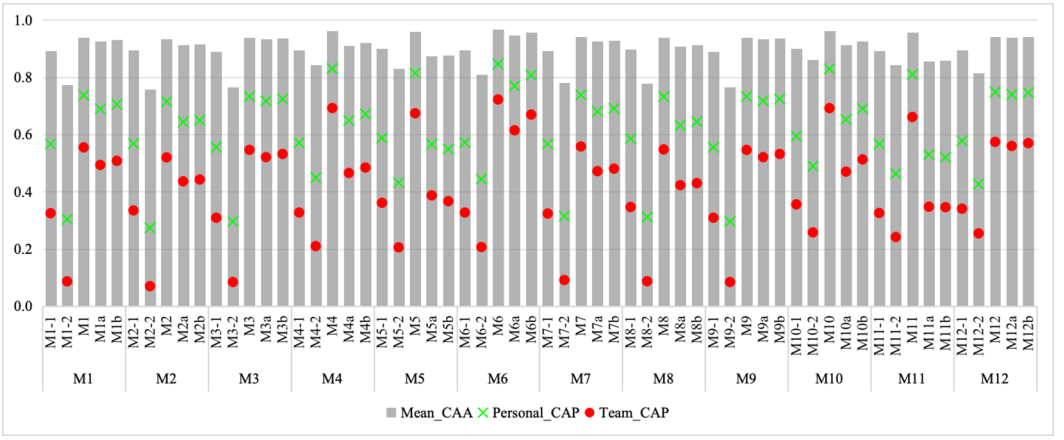


Figure 7. Classification accuracy of attributes and attribute patterns in simulation 2.
Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; M1-1 to M12-1: M1–M12 for individual response items; M1-2 to M12-2: M1–M12 for collaborative response items; M1a–M12a: M1–M12 with fixed teamwork effect; M1b–M12b: M1–M12 without team cognitions; CAA: classification accuracy of attribute; CAP Individuals: classification accuracy of attribute pattern for individual participant (5 attributes); CAP team: classification accuracy of attribute pattern for team (10 attributes).

participants' cognitive load. This shortened APM includes 18 items selected from the original 36-item APM using the revised maximum priority index algorithm (Liu et al., 2018). The selection ensured alignment with the cognitive component distribution of the original APM while maximizing test information under cognitive component constraints. Five cognitive components (Carpenter et al., 1990) were treated as cognitive attributes in this study: (α1) constant in a row, (α2) quantitative pairwise progression, (α3) addition/subtraction, (α4) distribution of three, and (α5) distribution of two.

To incorporate the teamwork aspect, the 18 items were divided into two sub-tests with balanced difficulty. One set was randomly assigned as individual response items, and the other as collaborative response items. The original APM's item serial numbers and their corresponding cognitive attributes (Q-matrix) for individual and teamwork response items are shown in Figure 8d. It should be noted that due to the absence of an identity submatrix, the Q-matrix is incomplete and may not fully satisfy the strict identifiability conditions for certain model parameters (Köhn & Chiu, 2017; Xu & Zhang, 2016). This limitation may affect the accuracy of diagnostic classification to some extent. Sample individual and teamwork items are presented in Figure 8a,b. For individual response items, both team members viewed the same item on their screens and responded independently on separate computers without communication. For collaborative response items, resource interdependence was introduced by creating information asymmetry: specific graphs in the matrix (e.g., three in the left or middle column) were masked to encourage verbal interaction. Team members were seated back-to-back to prevent nonverbal communication (e.g., visual or physical cues), as shown in Figure 8c. To ensure simultaneous responses, the system is configured so that when one member submits an answer first, he/she must wait until the other member also submits his/her answer before moving on to the next item.

Finally, five team cognition constructs corresponding to individual cognitive attributes are formed as (Θ_1) shared understanding of constant in a row, (Θ_2) shared understanding of quantitative pairwise progression, (Θ_3) shared understanding of addition/subtraction, (Θ_4) shared understanding of the distribution of three, and (Θ_5) shared understanding of the distribution of two.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

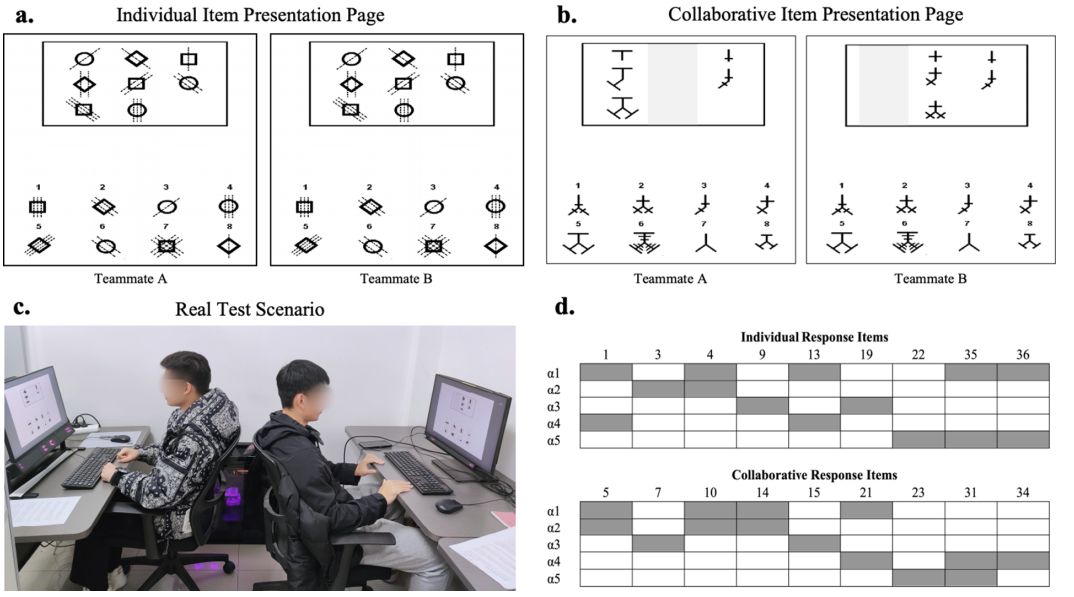


Figure 8. Sample individual and collaborative response items, real test scenario, and Q-matrices.
Note: (a) An individual response item consists of a 3 × 3 matrix with figural elements in the matrix area and 12 choices in the response options area; one cell in the matrix is missing and must be selected from the response options; both team members view the same information for the item on their screens. (b) A collaborative response item builds on the individual response item by blocking out part of the graphs, creating information asymmetry. In this case, each team member sees different but complementary information about the item on their screens. (c) Sitting back-to-back prevents communication, such as visual and physical movement. (d) Q-matrices for individual and collaborative response items; gray indicates "1," whereas white indicates "0."

4.2. Participants

A total of 67 dyadic teams, comprising 134 students, participated in this study (Gender: male = 26, female = 108; Age: $M = 20.310$, $SD = 1.952$). All participants were voluntarily recruited from a university in a coastal province in China, and none had previously taken the APM.

4.3. Data analysis

In this study, for collaborative response items, team members were permitted to submit their own individual responses rather than being required to agree on a unified response. This approach enabled the collection of data under two teamwork response modes. First, by recording the responses of each participant for the teamwork items, we naturally obtained data for the teamwork-separate response mode. Second, for the teamwork-unified response mode, a teamwork-unified response was considered correct only if both team members' answers were consistent and correct (i.e., $Y_{ti} = 1$ if $Y_{A(t)i} = Y_{B(t)i} = 1$). In all other cases, teamwork-unified responses were treated as incorrect (i.e., $Y_{ti} = 0$ if $Y_{A(t)i} = Y_{B(t)i} = 0$, $Y_{A(t)i} = 1$ and $Y_{B(t)i} = 0$, or $Y_{A(t)i} = 0$ and $Y_{B(t)i} = 1$).

Three teamwork covariates were used in this study: the number of teamwork items answered consistently by two team members (z_1; $M = 8.611$, $SD = 1.044$), the frequency of communication between team members (z_2; $M = 172.642$, $SD = 101.720$), and the time difference between two team members submitting their responses to teamwork items (z_3; in seconds, $M = 24.843$, $SD = 37.958$). For z_1, consistency was recorded only when the options submitted separately by both members were identical, regardless of whether they were correct. For z_2, the speaking frequency of each team member was recorded individually, and the communication frequency for the pair was defined as the lower frequency between the two. This approach excluded cases where one member spoke frequently without receiving a response. For z_3, the total time difference between the two members' answer submissions across all

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

teamwork items was calculated, without differentiating who waited for whom. Logically, higher values for the first two covariates indicate higher teamwork quality, while a higher value for the third covariate suggests lower teamwork quality. To explore the relative importance of these covariates, all three were standardized before being incorporated into the explanatory models.

All 12 Team-CDMs, the LLM, and a Rasch model with teamwork effect (denoted as Team-Rasch, details can be found in Section A2 in the Supplementary Material) (cf. Wilson et al., 2017) were applied to fit the data using two Markov chains per model. Each chain included 20,000 iterations, with the first 10,000 iterations discarded as burn-in. The remaining 20,000 iterations (10,000 per chain) were used for model parameter inference. We checked for convergence using the PSRF. The LLM can only be used to analyze the teamwork-separate response mode, which was employed to investigate the impact of using traditional CDM while disregarding team cognition in teamwork tasks. The Team-Rasch also applies only to the teamwork-separate response mode, which was used to showcase the performance of existing teamwork models to demonstrate the relative advantages of proposed models. Posterior predictive model checking (Gelman et al., 2014) was employed to evaluate the absolute model-data fit. A posterior predictive probability (*ppp*) value between 0.05 and 0.95 was considered indicative of no systematic differences between the observed and predicted data, suggesting an adequate model-data fit. The *ppp* values were computed as $ppp = \sum_{e=1}^{E}(Sum(\boldsymbol{X}^{postpred(e)}) \geq Sum(\boldsymbol{X}))/E$, where E is the total number of MCMC iterations, $\boldsymbol{X}$ is the observed data, and $\boldsymbol{X}^{postpred(e)}$ represents the posterior predicted data in the eth iteration; the posterior predicted data were generated using the item response function of each Team-CDM, based on parameter samples from the posterior distributions (Levy & Mislevy, 2016; Zhan, Chen, Wang, et al., 2024). For model selection, the DIC, WAIC, and LOO were calculated, with lower values indicating a better fit between the model and the data.

4.4. Results

The PSRF values for all parameters across models were below 1.1, indicating good convergence under the specified settings. Table 3 summarizes the model-data fit statistics for the 12 proposed models, the LLM, and the Team-Rasch. All *ppp* values fell within the range of 0.05–0.95, indicating an acceptable fit to the data. Furthermore, based on the DIC, WAIC, and LOO indices, the following findings can be observed: (1) the LLM exhibited poorer data fit compared with the models using the teamwork-separate mode, highlighting the significance of incorporating team cognition in teamwork tasks. (2) The Team-Rasch model exhibited poorer data fit compared with the models employing the teamwork-separate response mode, but outperformed the LLM. These results suggest that CDMs provide a more appropriate analytical framework for this task and underscore the importance of accounting for interdependence among team members. (3) The models incorporating the additive interaction mechanism generally exhibited the best fit among the three interaction mechanisms. However, in rare instances, the models employing the conjunctive interaction mechanism demonstrated a slightly better fit to the data. (4) The explanatory models consistently outperformed their corresponding standard models. However, because the data analyzed by models using the teamwork-separate and teamwork-unified modes are not identical, their relative fit indices are not directly comparable. For simplicity, the subsequent discussion focuses on parameter estimates from M11, which demonstrated the best relative fit among the proposed models.

Figure 9 summarizes the estimates of model parameters, including item parameters, attribute patterns, and teamwork effects. First, the overall quality of the items was not satisfactory, particularly due to the high guessing probabilities for the first four individual response items, likely caused by their low difficulty (cf. Liu et al., 2023). Second, the diagnostic results revealed that the 134 participants exhibited 24 of the 32 possible attribute patterns. Among these, the largest group of participants (31.34%) mastered all five attributes, followed by those who were missing the first and fifth attributes. This distribution suggests that the reasoning abilities (or fluid intelligence) of the participants were generally high. As shown in Table A6 in the Supplementary Material, the tetrachoric correlations among the five attributes reveal a low-to-moderate degree of association, suggesting that the attributes are relatively distinct yet

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

Table 3. Summary of model-data fits in empirical study

				Absolute fit				
	Relative fit			Individual responses		Separate responses		Unified responses
Model	DIC	WAIC	LOO	*ppp*_A	*ppp*_B	*ppp*_A	*ppp*_B	*ppp*_AB
M1	1,855.611	1,625.199	1,677.418	0.191	0.283	0.320	0.210	
M2	1,848.604	1,621.187	1,662.124	0.206	0.259	0.398	0.275	
M3	1,823.113	1,631.739	1,655.721	0.302	0.314	0.194	0.160	
M4	1,821.533	1,500.235	1,536.896	0.205	0.355			0.252
M5	1,777.569	1,494.596	1,531.770	0.188	0.348			0.251
M6	1,791.912	1,494.392	1,529.811	0.191	0.369			0.275
M7	1,855.004	1,609.804	1,662.263	0.170	0.273	0.267	0.169	
M8	1,851.862	1,617.215	1,660.853	0.195	0.268	0.362	0.238	
M9	1,820.375	1,628.343	1,652.488	0.301	0.300	0.186	0.145	
M10	1,792.779	1,494.120	1527.546	0.206	0.337			0.209
M11	1,752.376	1,488.108	1,526.886	0.192	0.347			0.214
M12	1,753.634	1,494.854	1533.783	0.193	0.322			0.197
LLM	2,259.624	1,928.356	1,947.566	0.246	0.296	0.133	0.252	
TR	1,983.906	1,868.442	1,870.830	0.813	0.425	0.178	0.129	

Note: M1: disjunctive model for teamwork-separate responses; M2: additive model for teamwork-separate responses; M3: conjunctive model for teamwork-separate responses; M4: disjunctive model for teamwork-unified responses; M5: additive model for teamwork-unified responses; M6: conjunctive model for teamwork-unified responses; M7: disjunctive model for teamwork-separate responses with covariates; M8: additive model for teamwork-separate responses with covariates; M9: conjunctive model for teamwork-separate responses with covariates; M10: disjunctive model for teamwork-unified responses with covariates; M11: additive model for teamwork-unified responses with covariates; M12: conjunctive model for teamwork-unified responses with covariates; LLM: log-linear model; TR: Team-Rasch model; DIC: deviance information criterion; WAIC: widely available information criterion; LOO: leave-one-out cross-validation; *ppp*: posterior predictive probability.

not entirely independent. Third, the teamwork-effect parameter estimates were significantly positively correlated with the total number of attributes mastered by the two team members ($r(\tau, \alpha) = 0.274$, $p = 0.025$), the consistency of answers submitted by the two members ($r(\tau, z_1) = 0.836$, $p < 0.001$), and the total frequency of communication ($r(\tau, z_2) = 0.404$, $p = 0.001$). Conversely, the teamwork effect was significantly negatively correlated with the total waiting time between the two members ($r(\tau, z_3) = -0.493$, $p < 0.001$). However, the partial correlation coefficient indicates that the relationship between the teamwork effect and the total waiting time was not significant when controlling for other variables ($pr(\tau, z_3) = 0.081$, $p = 0.511$), suggesting that the observed correlation depends on the influence of other variables. This finding was further supported by the w12 parameter estimates of the three covariates, with the estimates of $\widehat{\tau}_{11} = 0.624$, $\widehat{\tau}_{12} = 0.206$, and $\widehat{\tau}_{13} = 0.046$. Hence, the consistency of answers submitted by team members and the frequency of communication between them positively predicted the quality of teamwork. However, the waiting time between the two members did not play a significant role in predicting teamwork quality when considered alongside the former two factors. In other words, while waiting time alone was significantly negatively correlated with teamwork quality, its effect diminished when the other two variables were taken into account.

To showcase the diagnostic capabilities of the proposed models, Table 4 presents seven illustrative teams deliberately chosen to highlight distinct and characteristic outcomes, thereby facilitating a clearer understanding of the model outputs. When the total score of unified responses does not exceed the sum of individual responses from the two members, the quality of teamwork is typically poor. Additionally, greater consistency in the responses submitted by team members, more frequent communication, and shorter waiting times between them are generally associated with higher teamwork quality, supporting

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

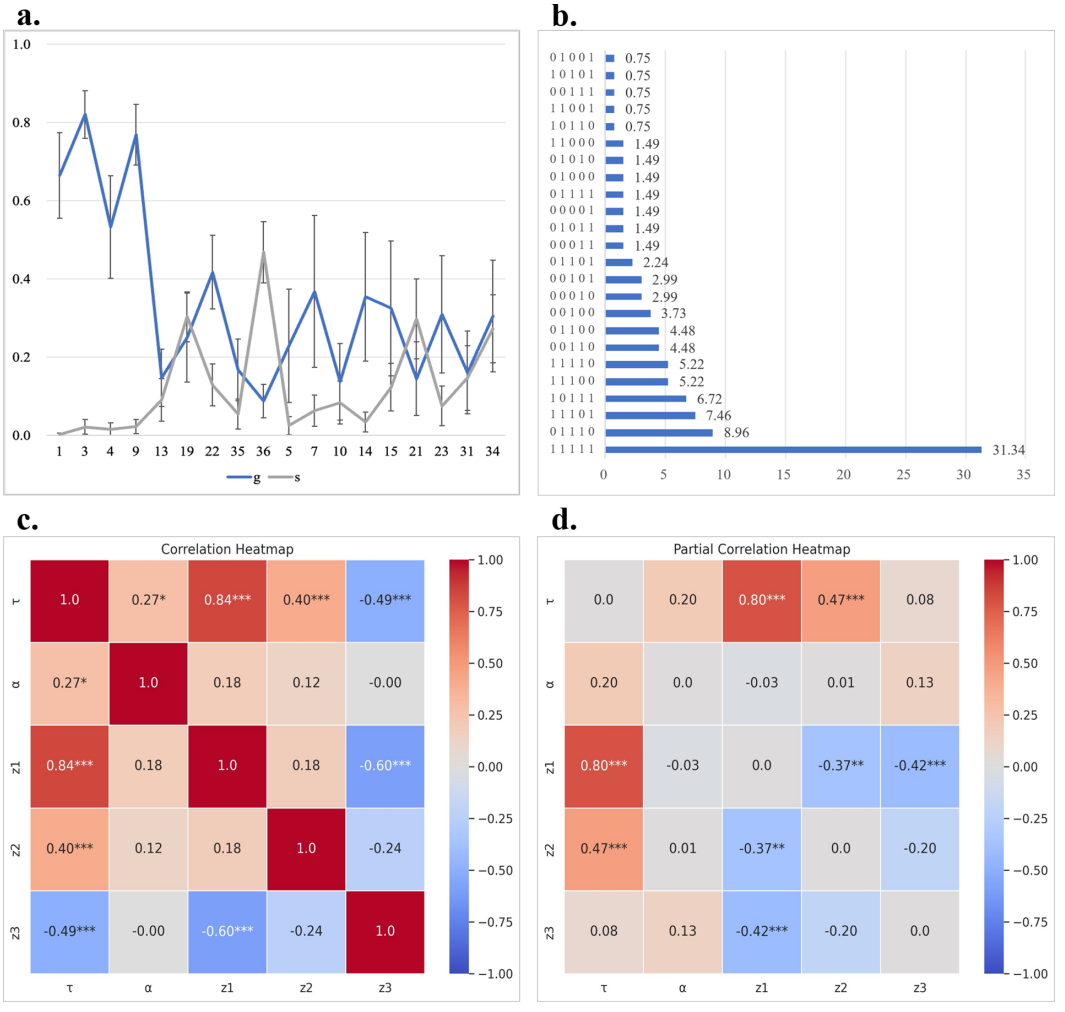


Figure 9. Summary of parameter estimates in the empirical study.
Note: (a) Item parameter estimates, g: guessing parameter, s: slipping parameter. (b) Mixing proportions of individual attribute patterns. (c) Correlation heatmap, τ: teamwork-effect estimates; α: total number of attributes mastered by the team; z_1: the number of teamwork items answered consistently by two team members; z_2: the frequency of communication between team members; z_3: the time difference between two team members submitting their responses to teamwork items. (d) Partial correlation heatmap; *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

the earlier findings. Further analysis of the teams reveals additional insights. Team 1, composed of two high-level members, achieved perfect scores on both individual and teamwork items, supported by efficient communication with below-average frequency and waiting times. Similarly, Team 2 also comprised high-level members, but their unified response score was lower than their individual scores, despite identical answers. This suggests effective communication but potential biases in their shared understanding of the items, hindering their collaborative performance. In contrast, Team 7, consisting of two low-level members, lacked the shared mastery of attributes necessary for team cognition, making it difficult to perform on teamwork items despite communication efforts. Team 15 provides an example of high-level members demonstrating poor teamwork, marked by minimal communication and significantly above-average waiting times, indicating near-independent work on teamwork items. Teams 23 and 48, featuring one high-level and one low-level member, showed divergent outcomes. Team 48 achieved ultra-short waiting times and perfect collaborative scores through frequent communication,

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

Table 4. Information about seven example teams

								$\text{Score}_{\text{Individual}}$		
Team	$\boldsymbol{\alpha}_A$	$\boldsymbol{\alpha}_B$	τ	α	z_1	z_2	z_3	A	B	$\text{Score}_{\text{Unified}}$
1	11111	11111	0.443	10	9	118	17.009	9	9	9
2	11111	11111	−0.694	10	9	141	17.113	8	7	6
7	01100	00001	−4.003	3	3	66	100.454	5	5	0
15	11111	11101	−1.025	9	7	9	232.367	8	7	7
23	11110	00100	−1.010	5	8	90	17.540	7	5	5
48	11111	01000	1.027	6	9	419	5.641	8	4	9
50	01110	11110	0.919	7	9	405	8.709	5	7	8

Note: $\boldsymbol{\alpha}_A$ and $\boldsymbol{\alpha}_B$: attribute patterns for members A and B; τ: teamwork-effect estimates; α: total number of attributes mastered by the team; z_1: the number of teamwork items answered consistently by two team members; z_2: the frequency of communication between team members; z_3: the time difference between two team members submitting their responses to teamwork items; $\text{Score}_{\text{Individual}}$: total score of individuals on individual response items; $\text{Score}_{\text{Unified}}$: total score of a team on collaborative response items.

while the high-level member in Team 23 failed to lead effectively, resulting in no substantial improvement in overall performance. Finally, Team 50 exemplifies how frequent communication can enable a team to exceed the performance levels of its individual members, demonstrating the potential for teamwork to enhance collective outcomes. These examples illustrate that team performance depends not only on the attributes mastered by members but also on the quality of their teamwork. Furthermore, they underscore the proposed models' ability to identify specific factors influencing team performance.

Overall, the results of the empirical study validate the model setup, confirm the interpretability of the model parameters, and underscore the benefits of incorporating teamwork-effect parameters.

5. Discussion

This study proposed 12 Team-CDMs to diagnose team members' cognitive attributes and assess teamwork quality during collaborative tasks. These models effectively identify whether poor team performance arises from individual knowledge or skill deficiencies or low-quality collaboration, such as ineffective communication. By highlighting specific areas for improvement, the proposed approach offers valuable insights for optimizing team-building strategies and implementing targeted interventions to enhance team cognition and overall performance.

Two simulation studies were conducted. Simulation Study 1 demonstrated the robustness and effectiveness of the models in recovering item and teamwork-related parameters. Classification accuracy for cognitive attributes and attribute patterns was consistently high across all models, confirming their suitability for teamwork-based diagnostic assessments. Explanatory models incorporating covariates like communication frequency and response consistency outperformed standard models, while models using the teamwork-separate mode generally achieved better parameter recovery than those employing the teamwork-unified mode. Increasing the number of items significantly improved parameter recovery and classification accuracy, underscoring the importance of sufficient data volume. Additive interaction mechanisms also showed a slight edge in recovery performance. Simulation Study 2 highlighted the necessity of integrating both teamwork-effect and interaction mechanisms for accurate modeling of team cognition. Unconstrained models achieved the highest classification accuracy and parameter recovery, while models that excluded teamwork quality or team cognitions showed diminished performance, particularly for collaborative response items. Models ignoring whole team cognitions entirely performed better than those neglecting teamwork quality, emphasizing the latter's greater impact. Furthermore, analyzing individual or teamwork response items in isolation resulted in suboptimal performance, reinforcing the need to incorporate both individual contributions and collaborative processes in teamwork tasks. Finally, a teamwork reasoning task illustrated the application of the models.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

The results indicated that team cognition significantly impacts teamwork quality, with the consistency of responses and frequency of communication positively predicting performance. Waiting time between members, however, did not play a significant role when considered alongside the other two factors. The models successfully identified specific reasons for poor team performance, such as insufficient knowledge or ineffective collaboration, demonstrating their practical utility in real-world assessments.

Despite promising results, further research is needed to address several limitations. First, this study used a constrained CDM—the LLM—as the base model for illustration. While extending it to a generalized CDM without specific condensation rules poses no theoretical challenges, the psychometric performance of such extensions requires further investigation.

Second, the study focused on dyadic teams, simplifying the modeling process but limiting the applicability of the proposed models to broader teamwork scenarios. Future research should explore diagnostic data from teams with three or more members to expand the utility of these models.

Third, while this study provides insights into measuring social/collaborative abilities through the teamwork-effect parameter (i.e., team-level random-effect parameter or teamwork covariates), the data analyzed were outcome-based (i.e., item response accuracy). Observed collaborative behaviors, documented through speech, were not modeled. Future research could integrate such social process data into the Team-CDM framework for a more direct assessment of social abilities.

Fourth, team cognition in this study was constructed using three interaction mechanisms—disjunctive, additive, and conjunctive. While the results demonstrate that the proposed models exhibit good psychometric properties, these mechanisms inevitably simplify the complexities inherent in cognitive interactions among team members. Therefore, future research should further investigate the underlying mechanisms of team cognition formation and develop more nuanced models of cognitive interaction to more accurately capture and interpret the interdependence among team members.

Fifth, the proposed models assume that all cognitive attributes contribute to team cognition through a uniform interaction mechanism. However, in practice, different cognitive attributes may follow distinct interaction mechanisms in forming team cognition. For example, one attribute may follow a disjunctive mechanism, while another attribute follows an additive one. In future work, it would be valuable to develop a generalized interaction framework that allows each cognitive attribute to follow a potentially different interaction mechanism.

Sixth, the proposed models under the teamwork-unified response mode assume equal contributions from both team members. However, in real-world scenarios, one member may dominate the decision-making process. Therefore, it would be valuable to collect data that capture the relative influence of individual members and to extend the model to account for unequal contributions to team decision-making.

Seventh, to avoid overloading the results with too much information, certain issues that may have affected the study's outcomes were simplified. For example, the types and levels of independent variables (e.g., item parameters) manipulated were insufficiently diverse. Hence, future research directions could prioritize exploring the performance of new models based on richer manipulated variables in order to provide more theoretical guidance for practical applications.

Eighth, with the rise of technology-enhanced assessments, the use of multimodal data—such as response time (Sinharay & Johnson, 2020; Zhan, Chen, Wang, et al., 2024), eye-tracking (Liu et al., 2023; Man & Harring, 2021; Zhan et al., 2022), brain activations (Jeon et al., 2021), and action sequences (Fu et al., 2024; Han et al., 2022)—is becoming increasingly feasible. Future efforts should integrate multimodal data into teamwork assessments to provide more accurate diagnoses and comprehensive feedback.

Nineth, with the rise of AI agents based on large language models (e.g., ChatGPT), researchers are investigating their impact on teamwork, particularly in human–machine or human–AI collaboration (Burton et al., 2024). Whether AI agents should be treated as independent team members or merely tools to enhance efficiency will shape future teamwork assessment research.

Lastly, Bayesian estimation depends heavily on prior distributions, which reflect analysts' assumptions. This study adopted priors from existing research without comparing their effects on parameter

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

estimation. Given that prior distributions can strongly influence results in small-scale projects, future studies should carefully select priors based on context-specific considerations, rather than simply replicating those used in this study.

In summary, this study represents a preliminary attempt in the domain of teamwork cognitive diagnosis. Despite its limitations, we hope that it serves as a catalyst for further exploration and encourages more researchers to engage with this emerging area of inquiry.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/psy.2025.10036.

Data availability statement. The datasets generated and/or analyzed during this study are not publicly available due to privacy restrictions, but they are available from the corresponding author upon reasonable request.

Author contributions. Conceptualization: P.Z.; Data curation: Z.W., G.C., H.Q.; Formal analysis: P.Z.; Funding acquisition: P.Z.; Investigation: G.C., H.Q.; Project administration: P.Z.; Software: Z.W.; Supervision: P.Z.; Writing—original draft: P.Z.; Writing—review and editing: P.Z.

Funding statement. This study was supported by the Ministry of Education (MOE) in China Project of Humanities and Social Sciences (Grant No. 24YJA190019) and the Zhejiang Provincial Philosophy and Social Sciences Planning Leading Talents Training Project (Grant No. 25QNYC010ZD).

Competing interests. The authors declared no potential competing interests with respect to the research, authorship, and/or publication of this article.

Ethical standards. The ethical approval was acquired from the Ethics Committee of Zhejiang Normal University (ZSRT2023020).

Informed consent. Informed consent was obtained from all individual participants included in the study.

References

Andersson, D., Rankin, A., & Diptee, D. (2017). Approaches to team performance assessment: A comparison of self-assessment reports and behavioral observer scales. *Cognition, Technology & Work*, *19*(4), 517–528. https://doi.org/10.1007/s10111-017-0428-0

Andrews, J. J., Kerr, D., Mislevy, R. J., von Davier, A. A., Hao, J., & Liu, L. (2017). Modeling collaborative interaction patterns in a simulation-based task. *Journal of Educational Measurement*, *54*(1), 54–69. https://doi.org/10.1111/jedm.12132

Austin, J. R. (2003). Transactive memory in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance. *Journal of Applied Psychology*, *88*(5), 866–878. https://doi.org/10.1037/0021-9010.88.5.866

Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.2307/1390675

Burton, J. W., Lopez-Lopez, E., Hechtlinger, S., Rahwan, Z., Aeschbach, S., & Lorenz-Spreen, P. (2024). How large language models can reshape collective intelligence. Nature Human Behaviour. Advance online publication. https://doi.org/10.1038/s41562-024-01959-9

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Lawrence Erlbaum Associates.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: The modified maximum global discrimination index method. *Educational and Psychological Measurement*, *70*(6), 902–913. https://doi.org/10.1177/0013164410366693

Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika*, *83*(2), 355–375. https://doi.org/10.1007/s11336-017-9595-4

Cooke, N. (2015). Team cognition as interaction. *Current Directions in Psychological Science*, *24*(6), 415–419. https://doi.org/10.1177/0963721415602474

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. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353. https://doi.org/10.1007/BF02295640

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, *95*(1), 32–53. https://doi.org/10.1037/a0017328

Devine, D. J., & Philips, J. L. (2001). Do smarter teams do better? A meta-analysis of cognitive ability and team performance. *Small Group Research*, *32*(5), 507–532. https://doi.org/10.1177/104649640103200501

Fu, Y., Zhan, P., Chen, Q., & Jiao, H. (2024). Joint modeling of action sequences and action time in computer-based interactive tasks. *Behavior Research Methods*, *56*, 4293–4310. https://doi.org/10.3758/s13428-023-02178-2

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC Press. https://doi.org/10.1201/b16018

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

Han, Y., Liu, H., & Ji, F. (2022). A sequential response model for analyzing process data on technology-based problem-solving tasks. *Multivariate Behavioral Research*, *57*(6), 960–977. https://doi.org/10.1080/00273171.2021.1932403

Hao, J., Liu, L., Kyllonen, P., Flor, M., & von Davier, A. A. (2019). Psychometric considerations and a general scoring strategy for assessments of collaborative problem solving. *ETS Research Report Series*, *2019*(1), 1–17. https://doi.org/10.1002/ets2.12276

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. In A. von Davier, M. Zhu, & P. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 135–156). Springer.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Jeon, M., De Boeck, P., Luo, J., Li, X., & Lu, Z.-L. (2021). Modeling within-item dependencies in parallel data on test responses and brain activation. *Psychometrika*, *86*, 239–271. https://doi.org/10.1007/s11336-020-09741-2

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

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*(2), 403–437. https://doi.org/10.1016/0149-2063(94)90021-3

Köhn, H. F., & Chiu, C.-Y. (2017). A procedure for assessing the completeness of the Q-matrices of cognitively diagnostic tests. *Psychometrika*, *82*(1), 112–132. https://doi.org/10.1007/s11336-016-9536-7

Lacerenza, C. N., Marlow, S. L., Tannenbaum, S. I., & Salas, E. (2018). Team development interventions: Evidence-based approaches for improving team performance. *American Psychologist*, *73*(4), 517–531. https://doi.org/10.1037/amp0000295

Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315374604

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, *88*(4), 587–604. https://doi.org/10.1037/0021-9010.88.4.587

Li, M., Liu, H., Cai, M., & Yuan, J. (2023). Estimation of individuals' collaborative problem solving ability in computer-based assessment. *Education and Information Technologies*, *29*, 483–515. https://doi.org/10.1007/s10639-023-12271-w

Liang, Q., de la Torre, J., & Law, N. (2021). Do background characteristics matter in children's digital literacy skills? An investigation with cognitive diagnosis models. *Computers in Human Behavior*, *122*, 106850. https://doi.org/10.1016/j.chb.2021.106850

Liu, S., Cai, Y., & Tu, D. (2018). On-the-fly constraint-controlled assembly methods for multistage adaptive testing for cognitive diagnosis. *Journal of Educational Measurement*, *55*(4), 595–613. https://doi.org/10.1111/jedm.12194

Liu, Y., Zhan, P., Fu, Y., Chen, Q., Man, K., & Luo, Y. (2023). Using a multi-strategy eyetracking psychometric model to measure intelligence and identify cognitive strategy in raven's advanced progressive matrices. *Intelligence*, *100*, 101782. https://doi.org/10.1016/j.intell.2023.101782

Lu, K., Gao, Z., Wang, X., Qiao, X., He, Y., Zhang, Y., & Hao, N. (2023). The hyper-brain neural couplings distinguishing high-creative group dynamics: An fNIRS hyperscanning study. *Cerebral Cortex*, *33*(5), 1630–1642. https://doi.org/10.1093/cercor/bhac161

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187–212. https://doi.org/10.1007/BF02294536

Man, K., & Harring, J. R. (2021). Assessing preknowledge cheating via innovative measures: A multiplegroup analysis of jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement*, *81*(3), 441–465. https://doi.org/10.1177/0013164420968630

OECD. (2017). PISA 2015 collaborative problem solving framework. In *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (revised ed., pp. 131–188). Paris: OECD Publishing. https://doi.org/10.1787/9789264281820-en

Park, Y. S., Xing, K., & Lee, Y.-S. (2018). Explanatory cognitive diagnostic models: Incorporating latent and observed predictors. *Applied Psychological Measurement*, *42*(5), 376–392. https://doi.org/10.1177/0146621617738012

Pan, Y., Cheng, X., & Hu, Y. (2022). Three heads are better than one: Cooperative learning brains wire together when a consensus is reached. Cerebral Cortex. Advance online publication. https://doi.org/10.1093/cercor/bhac161

Plummer, M. (2015). *JAGS: Just another Gibbs sampler*. https://mcmc-jags.sourceforge.io

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Ablex Publishing.

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine*, *78*(5), B77–B85.

Salas, E., Reyes, D. L., & Woods, A. L. (2017). The assessment of team performance: Observations and needs. In A. von Davier, M. Zhu, & P. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 21–36). Springer. https://doi.org/10.1007/978-3-319-33261-1_2

Sinharay, S., & Johnson, M. (2020). The use of item scores and response times to detect examinees who may have benefited from item preknowledge. *British Journal of Mathematical and Statistical, Psychology*, *73*(3), 397–419. https://doi.org/10.1111/bmsp.12187

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

Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal learning diagnostic assessment. *AERA Open*, *7*(1), 1–15. https://doi.org/10.1177/23328584211060804

Templin, J., & Henson, R. A. (2006). Measurement of pathological gambling: A cognitive diagnostic modeling approach. *Psychological Assessment*, *18*(1), 27–36. https://doi.org/10.1037/1040-3590.18.1.27

von Davier, A. A., & Halpin, P. F. (2013). Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series, RR-13-41*. https://doi.org/10.1002/j.2333-8504.2013.tb02348.x

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307. https://doi.org/10.1002/j.2333-8504.2005.tb01993.x

von Davier, M., & Lee, Y.-S. (2019). *Handbook of diagnostic classification models: Models and model extensions, applications, software packages*. Springer. https://doi.org/10.1007/978-3-030-05584-4

Wang, S., Zhang, S., Douglas, J., & Culpepper, S. (2018). Using response times to assess learning progress: A joint model for responses and response times. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 45–58. https://doi.org/10.1080/15366367.2018.1435105

Wilson, M., & De Boeck, P. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9

Wilson, M., Gochyyev, P., & Scalise, K. (2017). Modeling data from collaborative assessments: Learning in digital interactive social networks. *Journal of Educational Measurement*, *54*(1), 85–102. https://doi.org/10.1111/jedm.12134

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, *24*(6), 420–424. https://doi.org/10.1177/0963721415599543

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688. https://doi.org/10.1126/science.1193147

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, *81*(3), 625–649. https://doi.org/10.1007/s11336-015-9474-2

Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two lesscreative individuals turn into a highly-creative pair. *NeuroImage*, *172*, 527–537. https://doi.org/10.1016/j.neuroimage.2018.02.007

Yuan, J. L., Xiao, Y., & Liu, H. Y. (2019). Assessment of collaborative problem solving based on process stream data: A new paradigm for extracting indicators and modeling dyad data. *Frontiers in Psychology*, *10*, 369. https://doi.org/10.3389/fpsyg.2019.00369

Zhan, P., Chen, Q., Man, K., & Hao, N. (2024). Implementing cognitive component constraints in the development and validation of short-form Raven's advanced progressive matrices based on item response theory. *PsyArxiv*, https://doi.org/10.31234/osf.io/3cbm4

Zhan, P., Chen, Q., Wang, S., & Zhang, X. (2024). Longitudinal joint modeling for assessing parallel interactive development of latent ability and processing speed using responses and response times. *Behavior Research Methods*, *56*(3), 1656–1677. https://doi.org/10.3758/s13428-023-02113-5

Zhan, P., Jiao, H., Liao, M., & Bian, Y. (2019). Bayesian DINA modeling incorporating within-item characteristics dependency. *Applied Psychological Measurement*, *43*, 143–158. https://doi.org/10.1177/0146621618798660

Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics*, *44*(4), 473–503. https://doi.org/10.3102/1076998619856594

Zhan, P., & Qiao, X. (2022). Diagnostic classification analysis of problem-solving competence using process data: An item expansion method. *Psychometrika*, *87*, 1529–1547. https://doi.org/10.1007/s11336-022-09855-9

Zhan, P., Man, K., Wind, S., & Malone, J. (2022). Cognitive diagnosis modeling incorporating response times and fixation counts: Providing comprehensive feedback and accurate diagnosis. *Journal of Educational and Behavioral Statistics*, *47*(6), 736–776. https://doi.org/10.3102/10769986221111085

Zhang, M., Yin, Z., Zhang, X., Zhang, H., Bao, M., & Xuan, B. (2024). Neural mechanisms distinguishing two types of cooperative problem-solving approaches: An fNIRS hyperscanning study. *NeuroImage*, *291*, 120587. https://doi.org/10.1016/j.neuroimage.2024.120587

Cite this article: Zhan, P., Wang, Z., Chu, G. and Qiao, H., (2025). Teamwork Cognitive Diagnostic Modeling. *Psychometrika*, 1319–1345. https://doi.org/10.1017/psy.2025.10036

https://doi.org/10.1017/psy.2025.10036 Published online by Cambridge University Press