

基于成员间高阶认知交互的团队认知诊断建模*

詹沛达¹ 王志谋¹ 褚高红¹ 郝宁^{2,3}

(¹浙江师范大学心理学院, 金华 321004) (²华东师范大学心理与认知科学学院, 上海市心理健康与危机干预重点实验室, 上海 200062) (³合肥师范学院, 青少年心理健康与危机智能干预安徽省哲学社会科学重点实验室, 合肥 230601)

摘要 团队合作作为复杂社会系统中实现共同目标的关键行为模式, 其表现高度依赖于由成员间认知交互形成的团队认知。对团队认知的诊断测量不仅有助于探索团队成员间的认知交互规律, 还有助于识别影响团队行为表现的具体原因, 进而为实施有针对性干预、提升团队表现提供重要量化依据。然而, 目前尚缺乏具有团队认知诊断功能的心理测量模型。对此, 本研究构建了一个适用于分析团队合作作答数据的团队认知诊断模型——Team-DINA 模型。实证研究和模拟研究结果表明新模型适用于分析团队合作作答数据, 可实现同时评估团队整体和成员个体的认知属性掌握情况和认知能力水平; 不仅填补了团队认知诊断领域的方法学空白, 还进一步拓展了认知诊断测量范式的适用范围。

关键词 团队认知诊断, 团队合作, 团队认知, 认知诊断模型, DINA 模型

分类号 B841

1 引言

在以提升新质生产力为导向的发展阶段, 单纯的知识储备型人才已难以满足社会需求, 团队合作能力、问题解决能力和创造力等核心素养已成为新时代人才培养的关键(辛涛等, 2013; 杨向东, 2022)。随着社会分工日益精细化, 团队合作已逐渐取代个体工作, 成为现代社会生产和生活的基本组织形式。《中共中央国务院关于深化教育体制机制改革的意见》、《教育部关于全面深化课程落实立德树人根本任务的意见》和《义务教育课程方案(2022年版)》等政策文件均明确指出, 培养学生的合作能力是实现终身发展和社会需求的重要途径。

团队合作(teamwork)是指两个或两个以上个体通过相互协作实现共同目标的方式; 对其研究有助于探讨人际互动中的认知与行为规律, 理解合作过程中个体之间的交互影响(Salas et al., 2017; Zhang et al., 2024)。作为团队合作的核心构念, 团队认知(team cognition)是由成员个体间认知交互形成的团

队水平心理构念(如, 共同理解和知识共享), 是影响团队行为表现的主要因素(Cannon-Bowers et al., 1993; Salas et al., 2017)。团队认知包含了合作问题解决能力(Li et al., 2023)、团队创造力(Lu & Hao, 2024)和团队智力(Woolley et al., 2010)等一系列团队水平的认知因素构念。已有研究表明即便是由高能力个体组成的团队, 也不必然有优异表现, 凸显探索团队认知的重要性(Xue et al., 2018; Woolley et al., 2015)。对团队认知的诊断测量不仅有助于探索团队成员间的认知交互规律, 还有助于识别影响团队行为表现的具体原因, 进而为实施有针对性干预、提升团队表现提供重要量化依据(Lacerenza et al., 2018; Salas et al., 2007)。

近些年, 遵循“改进结果评价”导向, 客观量化个体认知现状并提供诊断反馈的测评模式日渐普及。其中, 认知诊断是一种基于诊断测评数据识别个体认知属性(如, 知识或技能)掌握优劣势并提供相应诊断反馈的测评范式。认知诊断有助于发挥测评的诊断和改进功能, 有助于识别影响个体绩效的

收稿日期: 2025-03-04

* 浙江省哲学社会科学规划领军人才培养专项课题(25QNYC010ZD)资助。

通信作者: 詹沛达, E-mail: pdzhan@gmail.com, zhan@zjnu.edu.cn; 郝宁, E-mail: nhao@psy.ecnu.edu.cn

具体原因,为促进个体认知发展提供了方法学支持(Tang & Zhan, 2021)。作为认知诊断的核心技术环节,认知诊断模型(cognitive diagnosis model, CDM)或诊断分类模型刻画了潜在认知属性与外显行为之间的概率关系,其建构的合理性(如,模型与测验情境的匹配度)直接影响诊断结果的准确性和可解释性。目前,针对不同的测验情境和诊断需求,已经有许多 CDM 被提出(von Davier & Lee, 2019),比如,确定性输入噪音“与”门(deterministic input, noisy "and" gate; DINA)模型(Junker & Sijtsma, 2001)。然而,现有 CDM 均局限于诊断单一个体的认知属性,不满足团队合作情境中对团队认知的诊断测量需求,限制了其在以团队为基本组织形式的实践场景中的应用。

随着计算机(网络)化测评的普及,对团队合作表现的测评已经逐步从基于自评或他评问卷的形式(Andersson et al., 2017)转变为基于计算机化合作任务的形式(Hesse et al., 2015; OECD, 2017)。这种转变促使研究者们构建新的心理测量模型来分析团队作答数据(如, Andrews et al., 2017; Li et al., 2023; Wilson et al., 2017; Yuan et al., 2019)。比如, Wilson 等人(2017)通过对单维 Rasch 模型引入团队水平随机效应参数来表征成员作答概率之间的互依性,假设成员个体能力和团队能力共同影响其在团队合作任务中的表现; Li 等人(2023)使用两维项目反应理论模型对个体合作问题解决认知能力和社交能力进行了测量,发现两者呈中等程度正相关。然而,已有模型多局限于成员个体构念的测量(即,将合作问题解决能力界定为个体参与团队合作以解决问题的能力),未关注由成员个体构念交互形成的团队构念(即,团队认知)对团队合作表现的影响(Cooke, 2015; Salas et al., 2017)。同时,这些模型均基于 Rasch 模型或项目反应理论模型构建,多数仅能提供一个笼统的个体合作问题解决能力估计值。由于缺乏细颗粒度诊断结果,这些模型难以揭示影响合作有效性的具体原因(如,合作无效是源于缺乏哪个特定的团队认知属性?)。此外,还有一些研究聚焦于团队成员之间的合作关系和神经同步性,而非团队认知的诊断测量;比如,基于网络分析的团队成员合作关系/模式研究(Zhu & Bergner, 2017)和基于相关分析获取脑间神经同步性的人际互动质量研究(Lu et al., 2023; Zhang et al., 2024)。总之,现有模型和方法不满足团队认知诊断需求,难以为团队合作提供有针对性的补救或培养

建议。

对此,本项目拟探究一种适用于团队合作作答数据分析的团队认知诊断建模框架,该框架聚焦于构建由成员个体间认知交互形成的团队认知对团队合作作答的影响机理模型,以期满足对团队认知的诊断测量需求,填补团队合作情境中缺乏诊断分析模型的空白。下文首先简单回顾适用于个体认知属性诊断的 DINA 模型;然后,以 DINA 模型为基础构建适用于分析团队合作作答数据的团队合作 DINA (teamwork-DINA, Team-DINA)模型;再然后,基于实证研究展现新模型的现实可应用性及相对优势,并通过两则模拟研究探究新模型的心理测量学性能;最后,总结研究结果和局限,并探讨未来研究方向。

2 团队认知诊断建模

2.1 个体认知诊断 DINA 模型

本研究以 DINA 模型为基础模型构建团队认知诊断模型:

$$P(Y_{ni} = 1 | \mathbf{a}_n) = g_i + (1 - s_i - g_i) \prod_{k=1}^K \alpha_{nk}^{q_{ik}}, \quad (1)$$

式中, $P(Y_{ni} = 1 | \mathbf{a}_n)$ 表示具有属性掌握模式 $\mathbf{a}_n = (\alpha_{n1}, \alpha_{n2}, \dots, \alpha_{nK})'$ 的参与者 n 正确作答题目 i 的概率; $\alpha_{nk} \in \{0, 1\}$ 为参与者 n 对属性 k 的掌握情况, $\alpha_{nk} = 1$ 表示掌握, $\alpha_{nk} = 0$ 反之; $q_{ik} \in \{0, 1\}$ 为题目 i 对属性 k 的考察情况, $q_{ik} = 1$ 表示考察, $q_{ik} = 0$ 反之;进而 $\prod_{k=1}^K \alpha_{nk}^{q_{ik}}$ 表示参与者 n 对题目 i 的理想作答

结果; g_i 和 s_i 分别为题目 i 的猜测和失误概率,分别表示未掌握全部考察属性的参与者(即,理想作答为错误的参与者)呈现正确作答的概率和掌握全部考察属性的参与者(即,理想作答为正确的参与者)呈现错误作答的概率。

考虑到认知属性之间存在相关并减少待估计参数数量,可进一步假设参与者对认知属性的掌握受一般认知能力的影响(de la Torre & Douglas, 2004):

$$\text{logit}(P(\alpha_{nk} = 1 | \theta_n)) = \gamma_{0k} + \gamma_{1k} \theta_n, \quad (2)$$

式中, θ_n 表示参与者 n 的一般认知能力,通常被约束为 $\theta_n \sim N(0, 1)$ 以保证模型可识别; γ_{0k} 和 γ_{1k} 分别表示属性 k 的掌握容易度和区分度。

DINA 模型假设当且仅当参与者掌握题目考察的所有认知属性时,其才有较高的概率正确作答该

题目。本研究选用 DINA 模型为基础模型处于以下两个原因：(1) DINA 模型作为最具代表性的 CDM，已经得到了广泛的研究和应用，研究者和实践者对其特性具有充分了解；(2) 相比于一般化饱和模型而言，DINA 模型对数据量的要求较低 (Chiu et al., 2018)，符合小规模团队合作项目的需求；(3) 本研究重点在于拓展传统 CDM 的团队合作数据分析功能，而非界定多个认知属性之间的结合逻辑 (即，缩合规则 [condensation rule])。因此，除了 DINA 模型外，亦可以使用其他 CDM 作为基础模型，但这并非本研究的核心内容。

2.2 团队认知诊断模型

2.2.1 模型构建

在构建模型前，我们先强调 4 个关键点。第一，本研究聚焦于团队合作中最基础的标准双人 (dyadic) 模式，即假设任一团队 t ($t = 1, 2, \dots, T$) 包含两名成员 A 和 B。第二，本研究中“团队合作作答 (teamwork response)”是指两团队成员合作后统一提交的反映团队整体决策的答案 (如，Woolley et al., 2010; Zhang et al., 2024)。第三，本研究针对的团队合作测验同时包含各成员独立作答题目和两成员合作作答题目；这种组合设计为评估团队合作的收益提供了一个基准 (Hao et al., 2017)。基于此，整个测验包含 3 个题目作答矩阵，分别是 2 个各成员的独立作答矩阵 (Y_A 和 Y_B) 和 1 个两成员的合作作答矩阵 (Y_{AB})。第四，基于团队认知的概念 (如，Cannon-Bowers et al., 1993; Salas et al., 2017)，本研究将“团队认知能力 (team cognitive ability)”界定为由成员间一般认知能力共同形成的团队一般认知能力，是一种团队水平连续变量；并将“团队认知属性 (team cognitive attribute)”界定为由成员间个体认知共同形成的对特定认知属性的共同理解或共享认知，是一种与个体认知属性相对应的团队水平分类变量。同时，假设团队认知能力直接影响团队认知属性的掌握。

设定 Y_{Ai} 和 Y_{Bi} 分别表示团队 t 中成员 A 和 B 对独立作答题目 i ($i = 1, 2, \dots, I_1$) 的作答， $Y_{(AB)i} \equiv Y_{ti}$ 表示包含成员 A 和 B 的双人团队 t 对合作作答题目 i ($i = 1, 2, \dots, I_2$) 的作答。对于独立作答而言，可以直接迁移应用传统 DINA 模型，有：

对于成员 A：

$$P(Y_{Ai} = 1 | \alpha_A) = g_i + (1 - s_i - g_i) \prod_{k=1}^K \alpha_{Ak}^{q_{ik}}, \quad (3a)$$

$$\text{logit}(P(\alpha_{Ak} = 1 | \theta_A)) = \gamma_{0k} + \gamma_{1k} \theta_A, \quad (4a)$$

对于成员 B：

$$P(Y_{Bi} = 1 | \alpha_B) = g_i + (1 - s_i - g_i) \prod_{k=1}^K \alpha_{Bk}^{q_{ik}}, \quad (3b)$$

$$\text{logit}(P(\alpha_{Bk} = 1 | \theta_B)) = \gamma_{0k} + \gamma_{1k} \theta_B, \quad (4b)$$

进一步，

$$\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & r_{AB} \sigma_A \sigma_B \\ r_{AB} \sigma_A \sigma_B & \sigma_B^2 \end{pmatrix} \right), \quad (5)$$

式中， r_{AB} 为两成员间一般认知能力的相关系数； μ_A 和 μ_B 为两成员一般认知能力的均值； σ_A^2 和 σ_B^2 为两成员的一般认知能力的方差，其余参数除区分了两成员身份 (A 或 B) 外含义同公式 (1) 和 (2) 中的；为保证模型可识别，可约束 $\mu_A = \mu_B = 0$ ， $\sigma_A^2 = \sigma_B^2 = 1$ 。该迁移应用相当于把 N 名参与者分为两组，每组包含 $T = N/2$ 名参与者，并进行分组估计。

进一步，对合作作答而言，本研究假设 (1) 团队认知能力影响团队认知属性的掌握概率，(2) 团队认知属性的掌握状况影响合作作答的正确概率，进而有：

$$P(Y_{ti} = 1 | \alpha_t) = g_i + (1 - s_i - g_i) \prod_{k=1}^K \alpha_{tk}^{q_{ik}}, \quad (6)$$

$$\text{logit}(P(\alpha_{tk} = 1 | \theta_A, \theta_B)) = \gamma_{0k} + \gamma_{1k} \theta_t, \quad (7)$$

$$\theta_t = \lambda_1 \theta_A + \lambda_2 \theta_B + \varepsilon_t, \quad (8)$$

式中， $\alpha_{AB} \equiv \alpha_t = (\alpha_{t1}, \alpha_{t2}, \dots, \alpha_{tk})'$ 为与个体认知属性相对应的团队认知属性模式，其中 α_{tk} 表示团队成员 A 和 B 对认知属性 k 的共同理解或共享认知； $\theta_{AB} \equiv \theta_t$ 为团队认知能力，反映团队合作水平或合作质量，受成员 A 和 B 的一般认知能力的共同影响， λ_1 和 λ_2 为相应的载荷； θ_A 和 θ_B 同样满足二元正态分布 (公式 (5))； $\varepsilon_{AB} \equiv \varepsilon_t$ 为残差项，表示团队认知能力中无法被两成员的一般认知能力共同解释的部分，可将其约束为 $\varepsilon_t \sim N(0, 1 - \lambda_1^2 - \lambda_2^2 - 2\lambda_1\lambda_2r_{AB})$ 以保证模型可识别 (即，约束 $\theta_t \sim N(0, 1)$)。根据共享心智模型 (shared mental model) 等团队认知理论 (Klimoski & Mohammed, 1994)，团队认知可能同时涵盖认知维度和合作维度 (李美娟 等, 2022; Hesse et al., 2015; Li et al., 2023)；因此，该残差项也可被笼统地理解为其他影响团队认知能力的合作维度因素的综合体 (如，信息交流、进度调控)，用于反映团队成员间的合作/社交质量。公式 (8) 可以有效解释团队认知超越成员个体认知的一些典型现象。例如，当残差项取绝对值较大的负值时 (如，低质量合作)，即使两名成员的能力水平较高，也可能导

致团队认知为负值,从而使团队认知属性的掌握概率较低;反之,当残差项取绝对值较大的正值时(如,高质量合作),即使两名成员能力水平较低,也可能导致团队认知为正值,从而使团队认知属性的掌握概率较高。图 1 呈现了 Team-DINA 的建模示意图。整体而言,Team-DINA 模型包括独立作答模型和合作作答模型两部分。

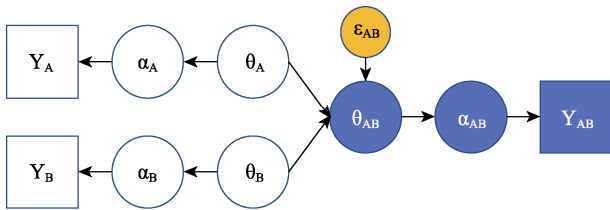


图 1 基于高阶潜在结构的团队认知诊断建模示意图
注: θ_A 和 θ_B : 成员 A 和 B 的一般认知能力; α_A 和 α_B : 成员 A 和 B 的个体认知属性向量; Y_A 和 Y_B : 成员 A 和 B 的独立作答向量; θ_{AB} : 团队认知能力; α_{AB} : 团队认知属性向量; Y_{AB} : 团队合作作答向量; ϵ_{AB} : 残差项(其他合作维度因素)。彩图见电子版,下同。

2.2.2 参数估计

本研究使用全贝叶斯马尔可夫链蒙特卡洛算法对 Team-DINA 进行参数估计,并基于 JAGS 软件实现。网络版附录 S1 章节中呈现了各参数的先验分布设定,关于如何使用 JAGS 进行贝叶斯参数估计可参见 Zhan, Jiao, Man 等人(2019)。

2.2.3 与已有合作问题解决能力测量模型的区别概述

如引言中所述,目前已经有一些针对合作问题解决任务的心理测量模型,这些模型多基于 21 世纪技能的评价与教学(ATC21S)框架(Hesse et al., 2015)或 2015 年国际学生评估项目(PISA 2015)框架(OECD, 2017)构建。两框架均认为合作问题解决同时涉及认知维度和社交维度,这与已有的团队认知理论(如,共享心智模型)相关研究一致。基于此,Team-DINA 和已有模型均认为有效团队合作受到认知和社交两维度的共同影响,只不过前者聚焦于测量已经形成的团队认知,而不是可能影响团队认知形成的社交维度因素;因此,公式 8 中的残差项只能笼统地反映团队成员间的合作/社交质量,并非测量目标。具体而言,Team-DINA 和已有模型的主要区别有 3 点:

(1)就测量目标而言,Team-DINA 兼顾个体构念(如,个体认知属性)与团队构念(如,团队认知属性)的测量,而已有模型多局限于对个体构念(如,个体合作问题解决能力)的测量;

(2)就测量用途而言,Team-DINA 适用于诊断测验,旨在诊断个体或团队的认知属性掌握情况,以揭示影响合作有效性的具体认知原因,并辅以获取个体或团队一般认知能力的相对(排序)位置;而已有模型多适用于描述测验,主要目的在于获取个体合作问题解决能力的相对位置。当然,这一差异与模型中潜变量的类型有关:Team-DINA 假设个体/团队的认知属性为类别变量、一般认知能力为连续变量,而已有模型多假设个体合作问题解决能力为连续变量;

(3)就分析数据而言,Team-DINA 聚焦于作答精度数据(即,结果数据),着重构建团队认知对团队合作作答数据的影响机制,不涉及对社交过程数据(如,语音或文本交流)的分析与利用;而部分已有模型(如, Li et al., 2023)则尝试挖掘并利用社交过程数据,以实现成员个体社交维度构念的测量。

3 实证数据分析

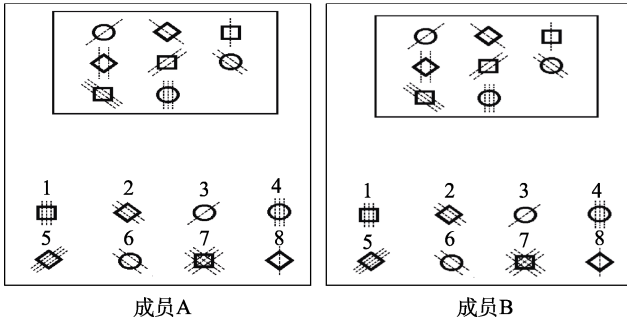
3.1 数据描述和分析

本研究使用一则双人矩阵推理测验数据来展现 Team-DINA 模型的实践可应用性。该数据包含 67 个双人团队(即, 134 名参与者)对 18 道来自短版瑞文高级推理测验(Zhan, Chen, Man et al., 2024)的题目的作答数据;网络版附录表 A1 呈现了 67 个团队的成员组成信息。该测验中前 9 题为独立作答题目,后 9 题为合作作答题目;分别考察 5 个个体认知属性(Carpenter et al., 1990; Wei et al., 2024): (α_1)行恒定规则(constant in a row)、(α_2)配对渐变规则(quantitative pairwise progression)、(α_3)加减规则(addition/subtraction)、(α_4)三值分布规则(distribution of three)和 (α_5)二值分布规则(distribution of two);和 5 个相应的团队认知属性: (α_{AB1})对行恒定规则的共同理解、(α_{AB2})对配对渐变规则的共同理解、(α_{AB3})对加减规则的共同理解、(α_{AB4})对三值分布规则的共同理解和(α_{AB5})对二值分布规则的共同理解。测验 Q 矩阵如图 2a 所示,其中题目序号为原版瑞文高级推理测验中的题目序号。需要说明的是,该测验题目数量相对于认知属性数量而言偏少,且由于缺乏单位矩阵,该 Q 矩阵并不具有完备性(丁树良 等, 2010; 宋丽红 等, 2024; Köhn & Chiu, 2017),这在一定程度上会影响诊断结果的准确性。

a. 测验Q矩阵

独立作答题目										合作作答题目								
	1	3	4	9	13	19	22	35	36	5	7	10	14	15	21	23	31	34
α_1																		
α_2																		
α_3																		
α_4																		
α_5																		
α_{AB1}																		
α_{AB2}																		
α_{AB3}																		
α_{AB4}																		
α_{AB5}																		

b. 独立作答题目示例



c. 合作作答题目示例

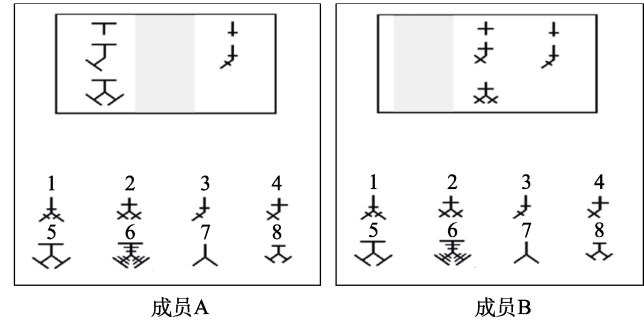


图 2 实证研究中的测验 Q 矩阵、独立作答题目示例和合作作答题目示例

注：图 2a 中题目序号为高级瑞文推理测验的原始题目序号。

对于独立作答题目(图 2b), 两成员在各自屏幕上查看相同的题目界面并独立作答, 作答过程中禁止交流。对于合作作答题目(图 2c), 可采用非对称设计(asymmetric design) (如, Hesse et al., 2015; Yuan et al., 2019), 通过遮挡矩阵区中部分图形的方式引入资源互依(resource interdependence), 创建信息不对称题目, 促使成员间通过沟通进行信息交流, 以实现合作。两成员在各自屏幕上查看相同题目的不同界面; 作答过程中允许言语交流, 为避免肢体动作等非语言交流, 两成员相背而坐。需要强调的是, 对于合作作答题目, 尽管两成员各自看到的界面不同, 但对于题目本身而言, 其考察的认知属性没有发生变化。

根据任务设定, 合作作答题目中, 允许成员各自提交自己的答案(仅当两成员均提交正确答案时, 团队合作作答才记为正确), 且当一名成员提交答案后需等待另一名成员提交答案后再同步进入下一题; 该设定便于采集成员间作答一致性(即, 两成员提交答案是否一致, 无论对错)和成员间相互等待时间(即, 两成员提交答案的时间差)等可反映团队合作质量的变量。最终, 形成 3 个反映团队合作质量的合作行为变量: 成员间作答一致性(z_1 ; $M = 8.611$, $SD = 1.044$)、成员间交流频次(z_2 ; $M = 172.642$, $SD = 101.720$)和成员间相互等待总时长(z_3 ; $M = 24.843$, $SD = 37.958$)。

3.2 数据分析

使用 Team-DINA 模型分析数据。参数估计时, 使用 5 条马尔科夫链(随机起点), 每条链包含

20,000 次迭代(预热 10,000 次), 稀疏值为 1; 最终剩余 50,000 次迭代用于参数估计。对于认知属性, 使用后验众数表示参数估计结果, 对于其他连续变量则使用后验均值表示参数估计结果。使用潜在量尺缩减因子(potential scale reduction factor, PSRF)作为参数估计收敛检验指标, 通常 $PSRF < 1.1$ 表示参数估计收敛(Brooks & Gelman, 1998)。使用后验预测模型检验(posterior predictive model checking, PPMC; Gelman et al., 2013)来评估模型-数据绝对拟合, 其中后验预测概率(posterior predictive probability, ppp) < 0.05 或 > 0.95 表示模型不拟合数据; 本研究中, PPMC 关注真实数据和预测数据之间的差异(Zhan, Chen, Wang et al., 2024)。

3.3 结果

首先, 所有模型参数估计值满足 $PSRF < 1.1$, 且绝大多数满足 $PSRF < 1.01$, 表明参数估计已经收敛。其次, Team-DINA 模型对三个数据集(Y_A 、 Y_B 和 Y_{AB})的 ppp 值分别为 0.340、0.429 和 0.274, 均处于 0.05 ~ 0.95 范围内, 表明新模型能够拟合数据。

图 3 呈现了实证研究的数据分析结果。首先, 图 3a 呈现了题目参数估计值, 结果表明该测验中题目的猜测度参数整体偏高, 这可能与该测验的整体难度偏低有关(cf., Liu et al., 2023)。图 3b 呈现了 67 个团队中两成员的个体认知属性和团队认知属性的混合比例分布。发现具有认知属性模式(11111)的人数占比最高, 且该认知属性模式中团队整体的掌握数量高于成员各自的掌握数量; 一方面表明该参与者群体的推理能力或流体智力整体偏高, 另一

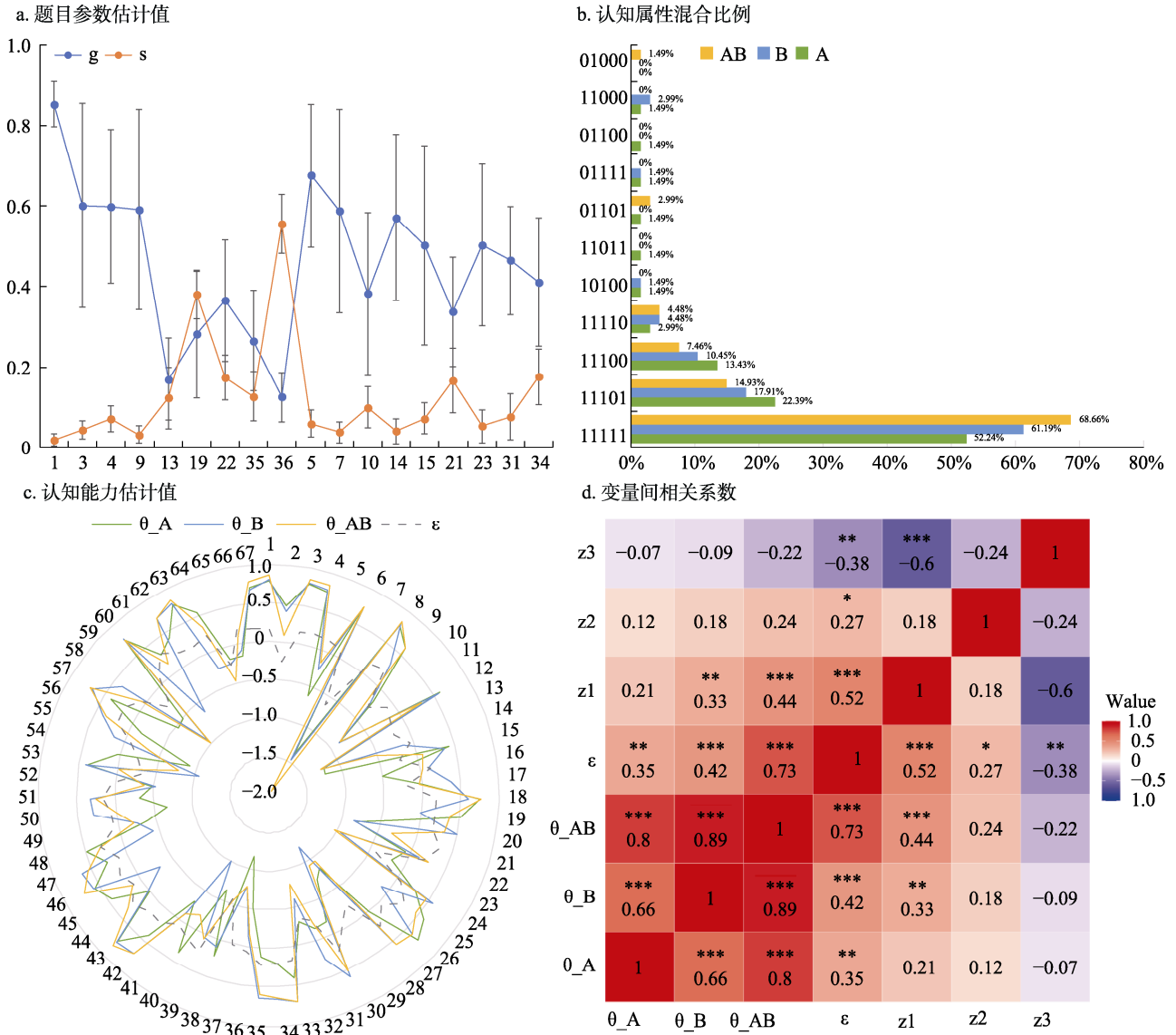


图 3 实证研究分析结果

注: g: 题目猜测参数; s: 题目失误参数; A: 成员 A, B: 成员 B, AB: 团队; θ : 一般能力; ϵ : 残差(其他合作维度因素); z_1 : 成员间作答一致性; z_2 : 成员间交流频次; z_3 : 成员间相互等待总时长. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

方面也表明团队合作整体带来了正收益, 成员之间存在一定程度的互补促进作用。

图 3c 呈现了 67 个团队中两成员的个体一般认知能力估计值、团队认知能力估计值和残差项。首先, 发现团队认知能力与两成员的个体一般认知能力之间均存在高水平显著正相关($r(\hat{\theta}_{AB}, \hat{\theta}_A) = 0.80$ ($p < 0.001$), $r(\hat{\theta}_{AB}, \hat{\theta}_B) = 0.89$ ($p < 0.001$), $r(\hat{\theta}_A, \hat{\theta}_B) = 0.66$ ($p < 0.001$))。其次, 各团队的残差项均在 0 附近波动, 表明两成员的个体一般认知能力的确能够解释部分团队认知能力, 但仍有少部分是无法单纯地被成员个体的认知能力所解释, 即团队认知并非成员个体认知的简单组合。进一步, 根据残差项和三个合作行为变量之间的相关系数(图

3d)可发现, 残差项水平越高, 则成员间作答一致性更高($r(\hat{\epsilon}_{AB}, z_1) = 0.517$ ($p < 0.001$))、成员间交流频次越高($r(\hat{\epsilon}_{AB}, z_2) = 0.268$ ($p = 0.028$))且成员间相互等待的时长越短($r(\hat{\epsilon}_{AB}, z_3) = -0.379$ ($p = 0.002$)), 一定程度上表明残差项与合作维度因素有关。

为进一步展现新模型的实践价值, 表 1 中呈现了 7 个团队的数据分析结果。团队 1 是一个典型的“双优成员”团队, 无论是成员个体还是团队整体均有较高的认知能力水平和认知属性掌握模式, 两成员在独立作答题目和合作作答题目上均获得满分, 且适当频次的交流和较短的相互等待时长也反映出两成员具有较好的合作积极性和合作同步性。与之相对, 尽管团队 2 中两成员的个体认知水平尚可,

表 1 7 个示例团队的潜在变量与外显变量

团队	θ_A	θ_B	θ_{AB}	ε_{AB}	α_A	α_B	α_{AB}	Raw _A	Raw _B	Raw _{AB}	z_1	z_2	z_3
1	0.79	0.81	0.86	0.17	11111	11111	11111	9	9	9	9	118	17.01
2	0.48	0.40	0.09	-0.25	11111	11111	11101	8	7	6	9	141	17.11
7	-1.38	-1.48	-1.95	-0.63	11100	10100	01000	5	5	0	3	66	100.45
15	0.43	0.37	0.07	-0.23	11111	11111	11101	8	7	7	7	9	232.37
48	0.48	-0.43	0.29	0.29	11111	01100	11111	8	4	9	9	419	5.64
56	-0.17	0.40	0.35	0.18	01101	11111	11111	6	6	8	9	405	8.71
66	-0.17	-0.09	0.11	0.18	11101	11111	11111	6	5	8	9	468	48.12

注: θ_A 和 θ_B : 成员 A 和 B 的一般认知能力; α_A 和 α_B : 成员 A 和 B 的个体认知属性向量; θ_{AB} : 团队认知能力; α_{AB} : 团队认知属性向量; ε_{AB} : 残差项(其他影响因素); Raw_A 和 Raw_B: 成员 A 和 B 在独立作答题目上的总分; Raw_{AB}: 团队在合作作答题目上的总分; z_1 : 成员间作答一致性; z_2 : 成员间交流频次; z_3 : 成员间相互等待总时长(秒).

且他们的交流频次和相互等待时长也反映出他们合作积极且同步性较高,但他们在合作作答题目上的得分却低于在独立作答题目上的,单凭观测变量难以解释;进一步,根据团队认知属性的估计结果发现,是两人对第四个认知属性的共同理解出现了偏差(即,个体均掌握该认知属性但缺失相应的团队认知属性)。团队 7 是一个典型的“双差成员”团队,即两成员个体认知水平均较低,且较低的作答一致性、交流频次和较长的相互等待时长也反映出他们未进行有效的合作交流,团队认知水平和团队认知属性掌握均较差,进而导致合作作答得分为 0。团队 15 与团队 2 类似,都是两名能力尚可的成员组成的团队,但不同的是,导致他们没有形成较高水平团队认知的主要原因可能来自合作维度因素,这点从他们的三个合作行为变量上能够反映出。团队 48 和 56 都是一名相对水平较高的成员与另一名相对水平较低的成员的组合,频繁的交流 and 极短的相互等待时长反映出两成员之间有较高的合作积极性和同步性(即,高水平成员带领低水平成员解决问题),这点从两团队的团队认知能力水平和团队认知属性掌握情况上也能反映出来。团队 66 是一个较为典型的两名中等偏低水平成员通过有效合作(频繁的沟通和相对较短的相互等待时长)获得正收益的团队,表现为合作作答题目得分高于独立作答题目得分。

整体而言,实证研究结果表明 Team-DINA 模型具有实践可应用性,且有助于深入理解团队合作表现并识别影响团队合作表现的具体认知因素。然而,由于实证研究中的测验情境比较单一且不够理想(如,题目数量偏少且 Q 矩阵不具备完备性),无法充分反映新模型的心理测量学性能,需要做进一步的模拟研究。

4 模拟研究

本研究包含 2 个模拟研究,其中,模拟研究 1 旨在探究 Team-DINA 模型在多种测验条件下的性能表现,模拟研究 2 旨在探究团队合作测验中独立作答题目和合作作答题目的不同占比对 Team-DINA 模型表现的影响。

4.1 模拟研究 1

4.1.1 数据生成与分析

模拟研究 1 包括 3 个操纵变量:(1)团队数量(T): 50、100 和 200;(2)分半题目数量(I): 10、15 和 30;(3)两成员一般认知能力相关($r_{AB} \equiv \rho$): -0.6、0 和 0.6。此外,属性数量(K)固定为 5 个,对应 Q 矩阵见图 4a。其中, $I = 15$ 和 30 时,测验的题目数量是认知属性数量的 3~5 倍经验范围内,且独立作答题目和合作作答题目的 Q 矩阵均包含两个可达矩阵以保证其完备性和模型可识别性;而 $I = 10$ 时,测验的题目数量相对认知属性数量偏少,且仅包含 1 个可达矩阵以保证完备性,但不保证模型可识别性(丁树良等,2010;宋丽红等,2024;Xu & Zhang, 2016)。

参考 Zhan, Jiao, Liao 等人(2019),题目参数依据二元正态分布生成:

$$\begin{pmatrix} \text{logit}(g_i) \\ \text{logit}(s_i) \end{pmatrix} \sim MVN \left(\begin{pmatrix} -2.197 \\ -2.197 \end{pmatrix}, \begin{pmatrix} 1 & -0.6 \\ -0.6 & 1 \end{pmatrix} \right). \quad (9)$$

此时,猜测参数和失误参数均呈现正偏态分布(均值约为 0.1,最小值约为 0.01,最大值约为 0.6)且两者之间呈中低程度负相关,更符合实际情况。属性容易度参数固定为 $\gamma_0 = (-1.5, -1, 0, 1, 1.5)'$,属性区分度均固定为 $\gamma_{1k} = 1.5$ 。两成员一般认知能力依据二元正态分布生成:

$$\begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (10)$$

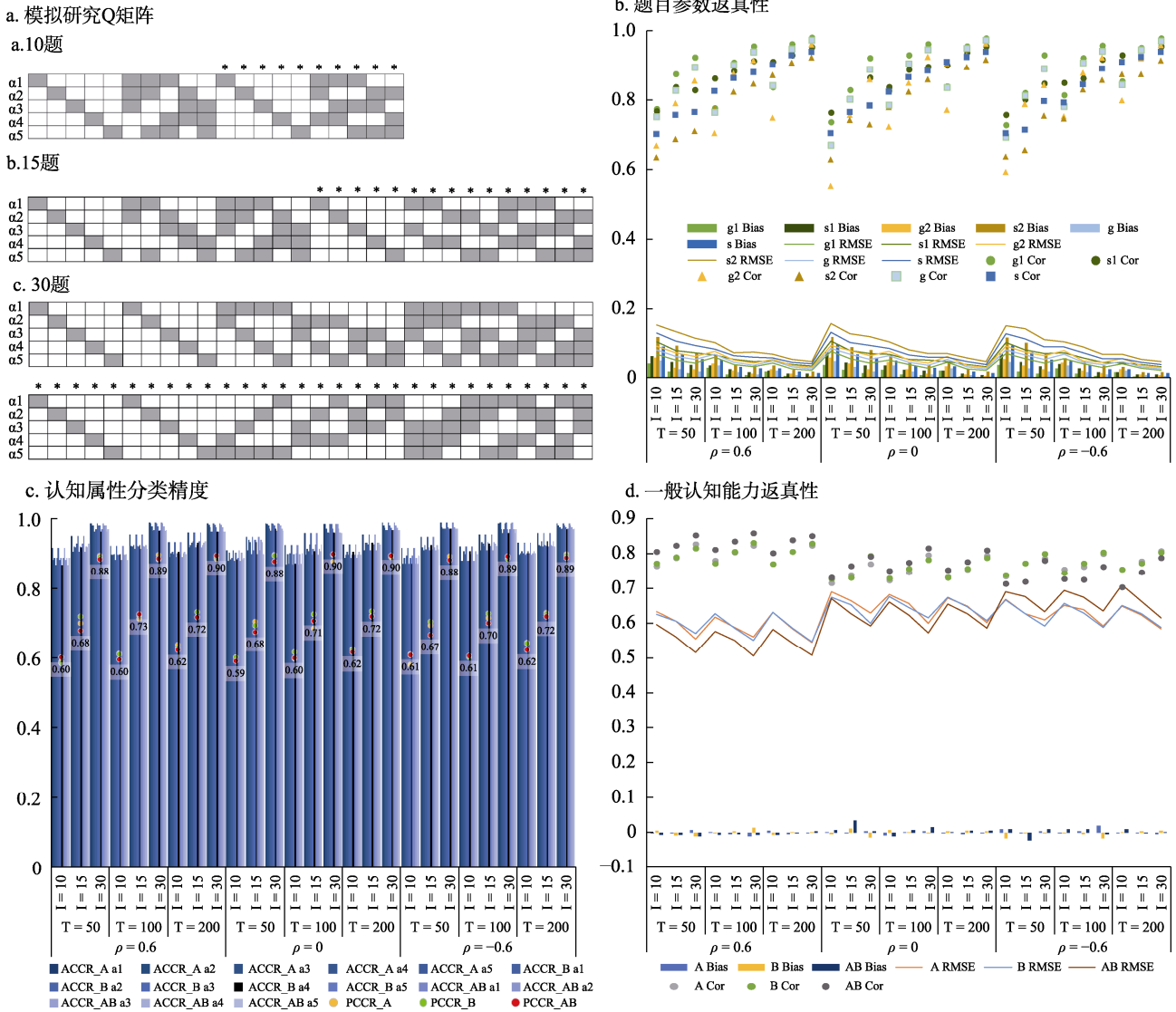


图 4 模拟研究 1 Q 矩阵及参数估计返真性

注: T: 团队数量; ρ : 成员间认知能力的相关; I: 分半测验的长度; Bias: 偏差; RMSE: 均方根误差; Cor: 真值与估计值的相关系数; A: 成员 A; B: 成员 B; AB: 团队。(a)模拟研究 Q 矩阵, 灰色为“1”、空白为“0”; 标记*为合作作答题目。(b)题目参数返真性, g1: 独立作答题目猜测参数; s1: 独立作答题目失误参数; g2: 合作作答题目猜测参数; s2: 合作作答题目失误参数; g: 全部题目猜测参数; s: 全部题目失误参数。(c)认知属性分类精度: ACCR: 属性判准率; PCCR: 模式判准率; 数值为 PCCR_AB。(d)一般认知能力返真性。

两成员一般认知能力对团队认知能力的载荷均固定为 $\lambda_1 = \lambda_2 = 0.5$ 。基于 $\alpha_{Ak} \sim Bernoulli(P(\alpha_{Ak} = 1|\theta_A))$ 和 $\alpha_{Bk} \sim Bernoulli(P(\alpha_{Bk} = 1|\theta_B))$ 生成个体认知属性, 基于 $\alpha_{tk} \sim Bernoulli(P(\alpha_{tk} = 1|\theta_t))$ 生成团队认知属性。最后, 独立作答数据依据 $Y_{Ai} \sim Bernoulli(P(Y_{Ai} = 1|\alpha_A))$ 和 $Y_{Bi} \sim Bernoulli(P(Y_{Bi} = 1|\alpha_B))$ 生成, 合作作答数据依据 $Y_{ti} \sim Bernoulli(P(Y_{ti} = 1|\alpha_t))$ 生成。为减少随机误差, 每种模拟测验条件下各生成 $R = 30$ 组数据。

使用 Team-DINA 模型分析数据。采用两条马尔可夫链(初始值随机), 每条链包含 10,000 次迭代(预热 5,000 次)。使用 PSRF 检验各参数是否收敛。

对于题目参数和一般认知能力参数, 使用偏差(Bias)、均方根误差(root mean square error, RMSE)和真值与估计值之间的相关系数(Cor)作为参数估计返真性指标。对于认知属性, 分别计算两组成员的个体认知属性(模式)判准率和团队整体的团队认知属性(模式)判准率, 具体而言: 个体认知属性判准率 $ACCR_{\#k} = \sum_{r=1}^{R=30} \sum_{n=1}^N I(\hat{\alpha}_{\#nk} = \alpha_{\#nk}) / NR$, 个体认知属性模式判准率 $PCCR_{\#} = \sum_{r=1}^{R=30} \sum_{n=1}^N I(\hat{\alpha}_{\#n} = \alpha_{\#n}) / NR$, 其中, #表示成员 A 或 B, $\hat{\alpha}_{\#n} = (\hat{\alpha}_{\#n1}, \hat{\alpha}_{\#n2}, \dots, \hat{\alpha}_{\#nK})'$; 而团队认知属性判准率 $ACCR_{tk} =$

$\sum_{r=1}^{R=30} \sum_{t=1}^T I(\hat{\alpha}_{tk} = \alpha_{tk}) / TR$, 团队认知属性模式判准率

$PCCR_t = \sum_{r=1}^{R=30} \sum_{t=1}^T I(\hat{\mathbf{a}}_t = \mathbf{a}_t) / TR$, 其中, $\hat{\mathbf{a}}_t = (\hat{\alpha}_{t1}, \hat{\alpha}_{t2}, \dots, \hat{\alpha}_{tK})'$, 其余参数含义同上。

4.1.2 结果

图 4b~4d 呈现了模拟研究 1 结果, 具体结果详见网络版附录 S2 章节。图 4b 呈现了题目参数估计返真性。首先, 在整体趋势上可发现(1)题目参数估计返真性随团队数量的增加和题目数量的增加而增加; (2)成员间认知能力的相关对题目参数估计返真性影响较小, 但在中等程度正相关($\rho = 0.6$)时的返真性略优于其他两种相关情况。其次, 独立作答题目的参数返真性优于合作作答题目的参数返真性, 这很可能是独立作答数据量(即, 2 个独立作答数据矩阵)多于合作作答数据量(即, 1 个合作作答数据矩阵)导致的。

图 4c 呈现了认知属性分类精度。整体而言, 无论是成员个体认知属性还是团队认知属性的判准率均较高, 且三者之间的差异较小; 与已有包含 5 个认知属性的研究相比(如, Zhan et al., 2022; Zhan, 2023), 分类精度符合预期。另外, 认知属性分类精度随题目数量的增加而增加, 而团队数量和成员间认知能力的相关对认知属性分类精度的影响很小。

图 4d 呈现了一般认知能力估计返真性。整体趋势上可发现(1)一般认知能力返真性随题目数量的增加而增加, (2)成员间认知能力具有正相关时, 一般认知能力的返真性优于另外两种相关情况, (3)团队数量对一般认知能力返真性影响很小。另外, 一个有趣的现象是, 当成员间认知能力具有正相关时, 团队认知能力的返真性优于两成员个体的认知能力返真性; 当成员间认知能力不相关时, 团队认知能力的返真性与两成员个体的认知能力返真性接近; 而当成员间认知能力具有负相关时, 团队认知能力的返真性劣于两成员个体的认知能力返真性; 该结果意味着团队成员的认知能力同质性有助于提高团队认知能力的估计精度。

整体而言, 模拟研究 1 结果表明 Team-DINA 模型在不同测验情境下具有良好的心理测量学性能。

4.2 模拟研究 2

4.2.1 数据生成与分析

为简化研究, 模拟研究 2 固定团队数量为 100, 分半题目数量为 15, 两成员间一般认知能力相关

系数为 0.6, 仅操纵独立作答题目和合作作答题目的数量分配这一个自变量(S-C), 包括 5-25、10-20、15-15、20-10 和 25-5 共 5 个水平(前者为独立作答题目数量, 设定独立作答题目或合作作答题目数量最小值为 5 以保证其至少包含 1 个可达矩阵)。题目参数、一般认知能力参数和认知属性的生成方式, 以及测验 Q 矩阵均与模拟研究 1 保持一致。数据分析过程与评价指标均与模拟研究 1 保持一致。

4.2.2 结果

图 5 呈现了模拟研究 2 结果, 具体结果详见网络版附录 S2 章节。图 5a 呈现了题目参数返真性, 可发现, (1)无论是独立题目的, 合作题目的, 还是所有题目的猜测参数的返真性均随独立题目数量的增加而增加; (2)随独立题目数量的增加, 独立题目的失误参数返真性逐渐增加, 而合作题目的失误参数返真性逐渐下降; (3)所有题目的失误参数返真性在独立题目和合作题目数量相等时最优。图 5b 呈现了认知属性分类精度, 可发现, 随着独立题目数量的增加, 无论是个体认知属性还是团队认知属性的 ACCR 均有所提高, 两成员个体认知属性模式的 PCCR 也有所提高, 但团队认知属性模式的 PCCR 有更大幅度的下降。图 5c 呈现了一般认知能力参数返真性, 可发现, (1)随着独立题目数量的增加, 两成员个体认知能力的返真性稍有提高; (2)团队认知能力的返真性在独立题目和合作题目数量相等时最优。整体而言, 在独立题目和合作题目数量相等时, Team-DINA 模型中各参数的返真性比较适中, 这也反映出团队合作任务中需包含独立题目和合作题目的必要性(Hao et al., 2017)。同时, 研究者可根据测量目的的不同略微调整独立题目和合作题目的占比; 比如, 如果测量目的更关注团队认知属性的模式判准率, 则可适当增加合作题目的数量, 而如果测量目的更关注成员个体认知属性的模型判准率, 则可适当增加独立题目的数量。

5 讨论与总结

团队合作作为复杂社会系统中实现共同目标的关键行为模式, 其效能高度依赖于团队认知这一核心构念(即, 通过成员间认知交互形成的共享心理表征(如, 共同理解))。对团队认知的诊断测量不仅有助于探索团队成员间的认知交互规律, 还有助于识别影响团队行为表现的具体原因, 进而为实施有针对性干预、提升团队表现提供重要量化依据。然而, 现有 CDM 均局限于诊断单一个体的认知属

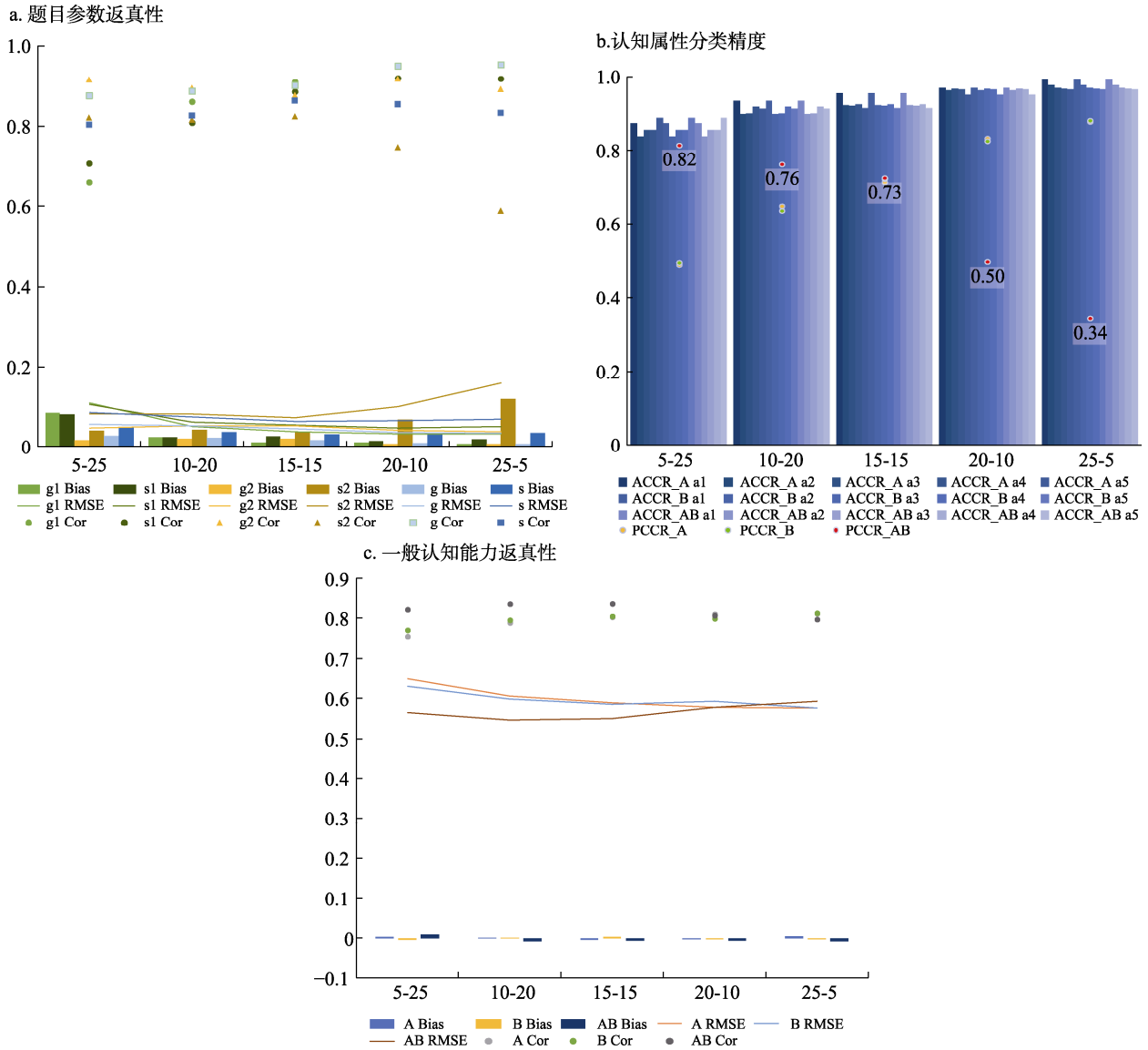


图 5 模拟研究 2 参数估计返真性

注: T: 团队数量; ρ : 成员间认知能力的相关; I: 分半测验的长度; Bias: 偏差; RMSE: 均方根误差; Cor: 真值与估计值的相关系数; A: 成员 A; B: 成员 B; AB: 团队。(a)题目参数返真性, g1: 独立作答题目猜测参数; s1: 独立作答题目失误参数; g2: 合作作答题目猜测参数; s2: 合作作答题目失误参数; g: 全部题目猜测参数; s: 全部题目失误参数。(b)认知属性分类精度: ACCR: 属性判准率; PCCR: 模式判准率; 数值为 PCCR_AB。(c)一般认知能力返真性。

性, 不满足团队合作情境中对团队认知的诊断测量需求; 同时, 现有基于项目反应理论构建的合作问题解决能力测量模型虽可实现对成员个体的认知能力(如, 参与合作问题解决的能力)的笼统测量, 但缺乏认知诊断功能, 难以揭示影响团队合作表现的具体原因。对此, 本研究以 DINA 模型为基础, 构建了适用于分析团队合作作答数据的团队认知诊断模型——Team-DINA 模型。首先, 通过一则双人合作矩阵推理测验展现了新模型的实践可应用性, 发现 Team-DINA 模型适用于分析团队合作作答数据, 能够识别影响团队合作表现的具体认知因素。

然后, 通过两个模拟研究探究了新模型的心理测量学性能, 研究发现(1)Team-DINA 模型在不同测验情境下具有良好的心理测量学性能, 适当增加题目数量和团队数量有助于进一步提高模型参数估计精度; (2)当团队合作测验包含相等数量的独立题目和合作题目时, Team-DINA 模型中各参数的返真性比较均衡且适宜, 且可根据测量目的适当调整独立题目和合作题目的占比。

相较于局限于个体认知的传统 CDM, 新模型采用潜在高阶认知交互的方式构建成员个体认知与团队认知之间的联结, 以刻画团队认知的形成机

理。具体而言,新模型通过高阶潜在结构联结两成员间个体一般认知能力,并引入残差项以表征无法由个体认知能力解释的团队层面差异(如,社交维度因素),以形成团队认知能力,进而影响团队认知属性的掌握。相较于简单取成员个体认知属性的并集或交集的方式(即,机械式合并),新模型有如下两点优势:(1)可避免对团队认知的有偏评价。比如,取并集的隐含假设是只要有一名成员掌握某认知属性,则团队整体必然掌握该认知属性。与之相对,取交集的隐含假设是只有两成员均掌握某认知属性,则团队整体才掌握该认知属性。显然,两种方式均过于机械和极端,无法灵活反映团队内部认知共享程度或社交/合作质量的实际差异;进而导致前者很可能会高估团队认知水平,后者很可能会低估团队认知水平。而新模型可根据成员个体认知能力和交互残差项等综合因素推断团队认知能力和团队认知属性的掌握概率,相对更灵活、普适性更强;(2)可捕捉成员间认知整合的促进或阻碍效应。新模型专门设置残差项,用以吸纳成员认知交互中的促进或阻碍效应(如,高质量合作带来的协同促进,或沟通障碍带来的认知损耗),从而可以解释为什么某些高能力个体组成的团队仍表现不佳,或低能力个体组成的团队反而表现不错,这些典型团队合作现象是简单的并集/交集方式无法揭示的。综上,本研究所提出的团队认知诊断建模更符合团队认知的形成与演化机理,为团队合作情景下的认知诊断测量提供了新的分析范式。

当然,本研究也有一定的局限性有待未来做进一步的拓展研究。第一,本研究是基于相对简单的 DINA 模型构建的,尽管替换不同的 CDM 并不影响本研究的核心研究内容,未来仍可尝试基于其他 CDM(如,多级评分模型、多分(概率态)属性模型或一般化广义模型)构建新的团队认知诊断模型,以满足不同的诊断需求。第二,本研究仅针对双人团队开展研究,尽管双人合作模式是所有团队合作模式的基础,但面对日益复杂的团队合作形式,有必要进一步探究适用于多人团队合作模式(如,“无领导小组”)的团队认知诊断模型。第三,本研究假设团队认知能力是对各成员一般认知能力的线性回归,尽管研究结果表明新模型具有良好的心理测量学性能,但该假设显然简化了成员间认知交互的复杂性,有必要在今后的研究中进一步探究团队认知的形成机理并构建更适宜的认知交互机理模型。第四,尽管在团队认知能力的建模中,我们指出残差

项可能代表了合作维度因素的综合体,但尚没有做深入探究,有待进一步完善。第五,本研究聚焦于传统作答精度数据建模,没有涉及合作过程数据(如,信息交流和操作流程),一定程度上忽略了问题解决过程中蕴含的丰富信息(韩雨婷等,2022;刘耀辉等,2022),未来可尝试探究如何引入合作过程数据或对其建模,以利用其蕴含的丰富信息提高诊断测量精度。第六,在本研究的实证研究和模拟研究中,均设定独立作答题目优先于合作作答题目呈现给参与者。尽管理论上题目的呈现顺序并不影响模型的使用,但如果合作作答题目出现在独立作答题目之前,成员间的合作交流可能会影响后续的独立作答(概率),从而导致独立作答(概率)之间不再满足局部独立性。因此,独立作答和合作作答题目的呈现顺序对结果产生的具体影响程度,仍有待在后续研究中进一步检验。第七,为聚焦于核心研究问题,本研究也简化或忽视了一些可能影响研究结果的问题,比如,模拟研究中操纵的自变量类型还不够丰富,实证研究中 Q 矩阵不具完备性,以及并未探讨 Q 矩阵的验证与修订(当然,团队合作测验的 Q 矩阵验证与修订本就是一个未来研究可关注的问题),未来有必要进一步开展相应研究。整体而言,本研究只是团队认知诊断领域的一次尝试,必然存在一些不足和局限,但我们仍希望该研究能起到“抛砖引玉”作用,吸引更多研究者关注该议题并开展研究。

综上所述,本研究构建的 Team-DINA 模型适用于分析团队合作作答数据,可实现同时评估团队整体和成员个体的认知属性掌握情况和认知能力水平;本研究不仅填补了团队认知诊断领域的方法学空白,还进一步拓展了认知诊断测量范式的适用范围。

参 考 文 献

- Andersson, D., Rankin, A., & Dipree, 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., Kerr, D., Mislevy, R., von Davier, 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>
- 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>
- 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>
- 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., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353. <https://doi.org/10.1007/BF02295640>
- Ding, S. L., Yang, S. Q., & Wang, W. Y. (2010). The importance of reachability matrix in constructing cognitively diagnostic testing. *Journal of Jiangxi Normal University (Natural Sciences Edition)*, *34*, 490–494.
- [丁树良, 杨淑群, 汪文义. (2010). 可达矩阵在认知诊断测验编制中的重要作用. *江西师范大学学报(自然科学版)*, *34*, 490–494.]
- Gelman, A., Carlin, J. B., Stern, H.S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>
- Han, Y., Xiao, Y., & Liu, H. (2022). Feature extraction and ability estimation of process data in the problem-solving test. *Advances in Psychological Science*, *30*(6), 1393–1409. <https://doi.org/10.3724/SP.J.1042.2022.01393>
- [韩雨婷, 肖悦, 刘红云. (2022). 问题解决测验中过程数据的特征抽取与能力评估. *心理科学进展*, *30*(6), 1393–1409.]
- 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. https://doi.org/10.1007/978-3-319-33261-1_9
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_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>
- 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](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*, 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 teamwork. *American Psychologist*, *73*(4), 517–531. <https://doi.org/10.1037/amp0000295>
- 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>
- Li, M., Liu, H., & Zhang, Y. (2022). The application of computational psychometrics in the assessment of key competencies: A case of collaborative problem solving. *Educational Research*, *506*(3), 127–137.
- [李美娟, 刘红云, 张咏梅. (2022). 计算心理测量理论在核心素养测评中的应用——以合作问题解决测评为例. *教育研究*, *506*(3), 127–137.]
- Liu, Y., Xu, H., Chen, Q., & Zhan, P. (2022). The measurement of problem-solving competence using process data. *Advances in Psychological Science*, *30*(3), 522–535. <https://doi.org/10.3724/SP.J.1042.2022.00522>
- [刘耀辉, 徐慧颖, 陈琦鹏, 詹沛达. (2022). 基于过程数据的问题解决能力测量及数据分析方法. *心理科学进展*, *30*(3), 522–535.]
- 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>
- Lu, K., & Hao, N. (2024). The interpersonal neural coupling in group creative ideation. *Psychological Review*. Advance online publication. <https://doi.org/10.1037/rev0000524>
- OECD. (2017). PISA 2015 collaborative problem-solving framework. In *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving* (pp. 131–188). PISA, OECD Publishing, Paris. <https://doi.org/10.1787/9789264281820-en>
- 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). Switzerland: Springer. https://doi.org/10.1007/978-3-319-33261-1_2
- 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.
- Song, L., Wang, W., & Ding, S. (2024). Q-matrix theory and its application in cognitive diagnostic assessment. *Advances in Psychological Science*, *32*(6), 1010–1030. <https://doi.org/10.3724/SP.J.1042.2024.01010>
- [宋丽红, 汪文义, 丁树良. (2024). 认知诊断评估中 Q 矩阵理论及应用. *心理科学进展*, *32*(6), 1010–1030.]
- 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>
- 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>
- Wei, J., Luo, L., Cai, Y., & Tu, D. (2024). A multistrategy cognitive diagnosis model incorporating item response times based on strategy selection theories. *Journal of*

- Educational and Behavioral Statistics*, 49(4), 658–686. <https://doi.org/10.3102/10769986231200469>
- 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, 686–688. <https://doi.org/10.1126/science.1193147>
- Xin, T., Jiang, Y., & Liu, X. (2013). Model construction of key competencies of students in compulsory education period. *Journal of Beijing Normal University (Social Science)*, (1), 5–11.
- [辛涛, 姜宇, 刘霞. (2013). 我国义务教育阶段学生核心素养模型的构建. *北京师范大学学报(社会科学版)*, (1), 5–11.]
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81(3), 625–649. <https://doi.org/10.1007/s11336-015-9471-z>
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *NeuroImage*, 172, 527–537. <https://doi.org/10.1016/j.neuroimage.2018.02.007>
- Yang, X. (2022). Competence and its model building: A theoretical analysis. *Journal of East China Normal University (Educational Science)*, 40(11), 41–57. <https://doi.org/10.16382/j.cnki.1000-5560.2022.11.004>
- [杨向东. (2022). 作为理论构念的素养及其模型构建. *华东师范大学学报(教育科学版)*, 40(11), 41–57.]
- 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. (2023). Deterministic input, noisy mixed modeling for identifying coexisting condensation rules in cognitive diagnostic assessment. *Journal of Intelligence*, 11, 55. <https://doi.org/10.3390/jintelligence11030055>
- 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*, <http://dx.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/10.1177/0146621618781594>
- 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/10.3102/1076998619826040>
- Zhan, P., Man, K., Wind, S. A., & 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, Y., Hu, Y., Ma, F., Cui, H., Cheng, X., & Pan, Y. (2024). Interpersonal educational neuroscience: A scoping review of the literature. *Educational Research Review*, 42, 100593. <https://doi.org/10.1016/j.edurev.2024.100593>
- Zhu, M., & Bergner, Y. (2017). Network models for teams with overlapping membership. In A. von Davier, M. Zhu, & P. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 303–314). Switzerland: Springer. https://doi.org/10.1007/978-3-319-33261-1_19

Teamwork cognitive diagnostic modeling with higher-order cognitive interaction between team members

ZHAN Peida¹, WANG Zhimou¹, CHU Gaohong¹, HAO Ning^{2,3}

⁽¹⁾ School of Psychology, Zhejiang Normal University, Jinhua 321004, China)

⁽²⁾ School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China)

⁽³⁾ Key Laboratory of Philosophy and Social Science of Anhui Province on Adolescent Mental Health and Crisis Intelligence Intervention, Hefei Normal University, Hefei 230601, China)

Abstract

Teamwork involves individuals collaborating to achieve shared goals that exceed individual capabilities. As a team-level construct, team cognition plays a critical role in effective teamwork by integrating individual expertise and fostering shared understanding. Identifying the causes of inefficiencies or poor performance in teams is essential for implementing targeted interventions and promoting the development of team cognition. However, traditional cognitive diagnosis models (CDMs), which are limited to individual problem-solving contexts, are inadequate for diagnosing team cognition in collaborative settings.

To address this, this study introduces a teamwork DINA (Team-DINA) model, an extension of the traditional DINA model for individual cognitive diagnosis, specifically designed to analyze teamwork response data. The proposed model constructs the formation mechanism of team members' individual cognitions into team cognition based on a higher-order latent structure. Further, it is hypothesized that individual cognition affects the mastery of individual cognitive attributes, while team cognition influences the mastery of team-level cognitive attributes. This allows for the simultaneous assessment of cognitive attribute mastery and the collective cognitive ability of both the team and its individual members. To demonstrate the practical applicability of the model, a dyadic teamwork matrix reasoning test was conducted. Additionally, two simulation studies were carried out to evaluate the psychometric performance of the model.

The results of the empirical study revealed that the Team-DINA model provides a deeper understanding of teamwork performance and identifies specific cognitive factors affecting team outcomes. The simulation studies highlighted two main findings: (1) The model demonstrates strong psychometric performance across various test scenarios, with improvements in model accuracy achieved by increasing the number of items and teams; (2) When the test includes an equal number of independent and collaborative response items, the model parameters are recovered in a balanced and appropriate manner, allowing for flexible adjustments of item proportions to enhance measurement accuracy.

Overall, the proposed Team-DINA model fills a critical methodological gap in the field of team cognition diagnosis and expands the application of cognitive diagnostic measurement paradigms from individual cognitive diagnosis to teamwork cognitive diagnosis.

Keywords teamwork cognitive diagnosis, teamwork, team cognition, cognitive diagnosis model, DINA model

附录:

S1. Team-DINA 模型的贝叶斯 MCMC 参数估计先验分布

参考 Zhan et al. (2019)中设定, 首先, 根据局部独立性假设, 有

$$\begin{aligned}
 Y_{Ai} &\sim \text{Bernoulli}(P(Y_{Ai} = 1|\alpha_A)) , \\
 Y_{Bi} &\sim \text{Bernoulli}(P(Y_{Bi} = 1|\alpha_B)) , \\
 Y_{it} &\sim \text{Bernoulli}(P(Y_{it} = 1|\alpha_t)) , \\
 \alpha_{Ak} &\sim \text{Bernoulli}(P(\alpha_{Ak} = 1|\theta_A)) , \\
 \alpha_{Bk} &\sim \text{Bernoulli}(P(\alpha_{Bk} = 1|\theta_B)) , \\
 \alpha_{tk} &\sim \text{Bernoulli}(P(\alpha_{tk} = 1|\theta_t)) .
 \end{aligned}$$

对题目参数而言, 有:

$$\begin{aligned}
 s_i &\sim \text{Beta}(1,1) , \\
 g_i &\sim \text{Beta}(1,1)I(g_i < 1 - s_i) .
 \end{aligned}$$

对两成员认知能力参数而言, 有:

$$\begin{aligned}
 \begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix} &\sim \text{MVN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) , \\
 \rho &\sim \text{Unif}(-1,1) .
 \end{aligned}$$

对团队认知能力而言, 有:

$$\begin{aligned}
 \theta_t &= \lambda_1\theta_A + \lambda_2\theta_B + \varepsilon_t , \\
 \lambda_1 &\sim \text{Norm}(0,1)I(0 < \lambda_1 < 1) , \\
 \lambda_2 &\sim \text{Norm}(0,1)I(0 < \lambda_2 < 1) , \\
 \varepsilon_t &\sim N(0, 1 - \lambda_1^2 - \lambda_2^2 - 2\lambda_1\lambda_2\rho) ,
 \end{aligned}$$

对属性容易度和区分度参数而言, 有:

$$\begin{aligned}
 \gamma_{0k} &\sim N(0,4) , \\
 \gamma_{1k} &\sim N(0,4)I(\gamma_{1k} > 0) .
 \end{aligned}$$

S2. 模拟研究补充结果

表 A1. 实证研究中 67 个团队的成员组成

背景变量	成员组成	组数
性别	男-男	7
	男-女	12
	女-女	48
独生子女	独生-独生	6
	独生-非独生	35
	非独生-非独生	26
学历	本科-本科	48
	本科-研究生	7
	研究生-研究生	12

表 A2. 模拟研究 1 题目参数估计返真性

ρ	T	I	g1			s1			g2			s2			g			s		
			Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor
0.6	50	10	0.042	0.069	0.775	0.064	0.104	0.768	0.060	0.092	0.668	0.118	0.153	0.634	0.051	0.081	0.751	0.091	0.128	0.702
		15	0.019	0.052	0.875	0.047	0.079	0.838	0.030	0.072	0.789	0.092	0.134	0.686	0.025	0.062	0.828	0.069	0.107	0.757
		30	0.015	0.043	0.925	0.037	0.073	0.830	0.025	0.062	0.854	0.072	0.115	0.709	0.020	0.053	0.896	0.054	0.094	0.765
	100	10	0.029	0.059	0.776	0.036	0.066	0.863	0.042	0.077	0.703	0.068	0.101	0.772	0.036	0.068	0.765	0.052	0.084	0.825
		15	0.011	0.038	0.910	0.026	0.054	0.885	0.021	0.052	0.876	0.037	0.073	0.824	0.016	0.045	0.901	0.031	0.064	0.863
		30	0.007	0.032	0.957	0.018	0.046	0.914	0.013	0.043	0.915	0.037	0.074	0.847	0.010	0.038	0.938	0.027	0.060	0.880
	200	10	0.018	0.044	0.839	0.020	0.048	0.911	0.025	0.054	0.748	0.035	0.068	0.872	0.022	0.049	0.842	0.028	0.058	0.901
		15	0.005	0.027	0.963	0.013	0.039	0.930	0.012	0.036	0.929	0.023	0.053	0.907	0.009	0.032	0.948	0.018	0.046	0.929
		30	0.002	0.021	0.982	0.011	0.033	0.953	0.007	0.029	0.959	0.018	0.048	0.922	0.005	0.025	0.972	0.015	0.041	0.940
0	50	10	0.038	0.076	0.736	0.074	0.107	0.763	0.058	0.093	0.551	0.118	0.157	0.627	0.048	0.085	0.669	0.096	0.132	0.703
		15	0.022	0.056	0.830	0.044	0.078	0.803	0.042	0.080	0.757	0.089	0.127	0.740	0.032	0.068	0.801	0.066	0.103	0.765
		30	0.014	0.043	0.922	0.036	0.069	0.865	0.024	0.063	0.860	0.080	0.120	0.729	0.019	0.053	0.889	0.058	0.094	0.783
	100	10	0.024	0.051	0.828	0.035	0.068	0.837	0.045	0.078	0.722	0.066	0.103	0.778	0.035	0.064	0.785	0.051	0.086	0.823
		15	0.011	0.038	0.930	0.022	0.053	0.888	0.025	0.052	0.848	0.046	0.081	0.824	0.018	0.045	0.906	0.034	0.067	0.866
		30	0.008	0.029	0.962	0.020	0.051	0.897	0.012	0.043	0.925	0.038	0.071	0.859	0.010	0.036	0.946	0.029	0.061	0.885
	200	10	0.020	0.045	0.841	0.021	0.046	0.900	0.034	0.061	0.771	0.047	0.071	0.903	0.027	0.053	0.836	0.034	0.059	0.909
		15	0.004	0.027	0.957	0.013	0.038	0.939	0.011	0.037	0.929	0.025	0.057	0.894	0.007	0.032	0.949	0.019	0.048	0.923
		30	0.003	0.022	0.980	0.009	0.032	0.956	0.007	0.030	0.963	0.018	0.047	0.916	0.005	0.026	0.972	0.014	0.040	0.939
-0.6	50	10	0.037	0.070	0.729	0.070	0.102	0.759	0.054	0.092	0.592	0.115	0.151	0.635	0.046	0.081	0.692	0.092	0.126	0.703
		15	0.019	0.056	0.822	0.046	0.081	0.802	0.035	0.075	0.788	0.101	0.143	0.654	0.027	0.065	0.812	0.074	0.112	0.715
		30	0.012	0.044	0.931	0.034	0.070	0.848	0.025	0.063	0.843	0.072	0.111	0.755	0.019	0.054	0.890	0.053	0.090	0.796
	100	10	0.028	0.056	0.814	0.039	0.073	0.850	0.045	0.078	0.751	0.057	0.104	0.745	0.036	0.067	0.782	0.048	0.088	0.791
		15	0.012	0.040	0.923	0.028	0.057	0.862	0.022	0.050	0.878	0.048	0.089	0.829	0.017	0.045	0.907	0.038	0.073	0.844
		30	0.007	0.031	0.959	0.017	0.045	0.919	0.013	0.044	0.923	0.035	0.068	0.858	0.010	0.038	0.940	0.026	0.057	0.890
	200	10	0.017	0.041	0.854	0.017	0.044	0.931	0.026	0.055	0.799	0.032	0.069	0.874	0.021	0.048	0.845	0.025	0.056	0.910
		15	0.004	0.028	0.952	0.010	0.038	0.948	0.011	0.037	0.919	0.017	0.053	0.874	0.008	0.032	0.945	0.014	0.045	0.925
		30	0.002	0.021	0.979	0.010	0.033	0.957	0.008	0.031	0.959	0.017	0.046	0.913	0.005	0.026	0.970	0.014	0.040	0.939

表 A3. 模拟研究 1 认知属性判准率

ρ	T	I	ACCR_A					ACCR_B					ACCR_AB					PCCR_A	PCCR_B	PCCR_AB
			a1	a2	a3	a4	a5	a1	a2	a3	a4	a5	a1	a2	a3	a4	a5			
0.6	50	10	0.917	0.881	0.888	0.866	0.888	0.917	0.881	0.888	0.866	0.888	0.917	0.881	0.888	0.866	0.888	0.589	0.585	0.603
		15	0.951	0.905	0.921	0.918	0.928	0.951	0.905	0.921	0.918	0.928	0.951	0.905	0.921	0.918	0.928	0.701	0.721	0.679
		30	0.986	0.981	0.981	0.963	0.969	0.986	0.981	0.981	0.963	0.969	0.986	0.981	0.981	0.963	0.969	0.895	0.892	0.885
	100	10	0.921	0.898	0.896	0.899	0.881	0.921	0.898	0.896	0.899	0.881	0.921	0.898	0.896	0.899	0.881	0.614	0.611	0.596
		15	0.955	0.922	0.920	0.924	0.914	0.955	0.922	0.920	0.924	0.914	0.955	0.922	0.920	0.924	0.914	0.714	0.723	0.726
		30	0.989	0.986	0.978	0.965	0.967	0.989	0.986	0.978	0.965	0.967	0.989	0.986	0.978	0.965	0.967	0.897	0.886	0.887
	200	10	0.934	0.891	0.901	0.905	0.900	0.934	0.891	0.901	0.905	0.900	0.934	0.891	0.901	0.905	0.900	0.638	0.630	0.624
		15	0.961	0.917	0.925	0.930	0.936	0.961	0.917	0.925	0.930	0.936	0.961	0.917	0.925	0.930	0.936	0.732	0.733	0.718
		30	0.987	0.983	0.976	0.963	0.963	0.987	0.983	0.976	0.963	0.963	0.987	0.983	0.976	0.963	0.963	0.888	0.891	0.895
0	50	10	0.910	0.879	0.903	0.885	0.887	0.910	0.879	0.903	0.885	0.887	0.910	0.879	0.903	0.885	0.887	0.602	0.603	0.591
		15	0.948	0.910	0.939	0.933	0.899	0.948	0.910	0.939	0.933	0.899	0.948	0.910	0.939	0.933	0.899	0.705	0.695	0.675
		30	0.986	0.984	0.979	0.972	0.965	0.986	0.984	0.979	0.972	0.965	0.986	0.984	0.979	0.972	0.965	0.898	0.896	0.878
100	10	0.925	0.869	0.898	0.872	0.900	0.925	0.869	0.898	0.872	0.900	0.925	0.869	0.898	0.872	0.900	0.598	0.617	0.600	
	15	0.958	0.897	0.913	0.908	0.927	0.958	0.897	0.913	0.908	0.927	0.958	0.897	0.913	0.908	0.927	0.686	0.726	0.708	

续表

ρ	T	I	ACCR_A					ACCR_B					ACCR_AB					PCCR_A	PCCR_B	PCCR_AB
			a1	a2	a3	a4	a5	a1	a2	a3	a4	a5	a1	a2	a3	a4	a5			
		30	0.985	0.984	0.971	0.961	0.958	0.985	0.984	0.971	0.961	0.958	0.985	0.984	0.971	0.961	0.958	0.878	0.897	0.900
	200	10	0.928	0.888	0.896	0.905	0.896	0.928	0.888	0.896	0.905	0.896	0.928	0.888	0.896	0.905	0.896	0.627	0.625	0.618
		15	0.955	0.909	0.930	0.932	0.927	0.955	0.909	0.930	0.932	0.927	0.955	0.909	0.930	0.932	0.927	0.723	0.735	0.720
		30	0.988	0.980	0.974	0.967	0.968	0.988	0.980	0.974	0.967	0.968	0.988	0.980	0.974	0.967	0.968	0.892	0.896	0.896
-0.6	50	10	0.915	0.870	0.888	0.870	0.895	0.915	0.870	0.888	0.870	0.895	0.915	0.870	0.888	0.870	0.895	0.582	0.611	0.610
		15	0.947	0.917	0.909	0.924	0.912	0.947	0.917	0.909	0.924	0.912	0.947	0.917	0.909	0.924	0.912	0.695	0.705	0.667
		30	0.989	0.975	0.971	0.971	0.971	0.989	0.975	0.971	0.971	0.971	0.989	0.975	0.971	0.971	0.971	0.893	0.883	0.882
	100	10	0.920	0.879	0.880	0.908	0.900	0.920	0.879	0.880	0.908	0.900	0.920	0.879	0.880	0.908	0.900	0.607	0.603	0.607
		15	0.954	0.910	0.934	0.928	0.923	0.954	0.910	0.934	0.928	0.923	0.954	0.910	0.934	0.928	0.923	0.716	0.728	0.701
		30	0.987	0.988	0.970	0.958	0.969	0.987	0.988	0.970	0.958	0.969	0.987	0.988	0.970	0.958	0.969	0.887	0.887	0.893
	200	10	0.932	0.896	0.902	0.907	0.902	0.932	0.896	0.902	0.907	0.902	0.932	0.896	0.902	0.907	0.902	0.645	0.643	0.625
		15	0.959	0.924	0.922	0.935	0.921	0.959	0.924	0.922	0.935	0.921	0.959	0.924	0.922	0.935	0.921	0.731	0.726	0.721
		30	0.986	0.981	0.976	0.970	0.972	0.986	0.981	0.976	0.970	0.972	0.986	0.981	0.976	0.970	0.972	0.900	0.895	0.890

表 A4. 模拟研究 1 认知能力参数估计返真性

ρ	T	I	A			B			AB		
			Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor
0.6	50	10	-0.001	0.633	0.764	0.005	0.626	0.772	-0.006	0.595	0.806
		15	-0.003	0.605	0.788	-0.008	0.605	0.790	-0.006	0.561	0.823
		30	0.008	0.555	0.827	-0.010	0.570	0.815	-0.009	0.516	0.853
	100	10	0.000	0.618	0.780	-0.001	0.626	0.773	-0.006	0.576	0.812
		15	-0.004	0.589	0.803	0.004	0.586	0.805	-0.005	0.550	0.836
		30	-0.010	0.561	0.824	0.014	0.550	0.831	-0.006	0.506	0.858
	200	10	0.005	0.631	0.770	-0.006	0.632	0.770	-0.005	0.582	0.802
		15	-0.004	0.584	0.807	0.002	0.586	0.805	-0.001	0.544	0.839
		30	0.000	0.545	0.823	0.001	0.547	0.830	0.004	0.508	0.851
0	50	10	0.001	0.690	0.715	-0.005	0.674	0.727	0.007	0.670	0.732
		15	-0.003	0.665	0.739	0.011	0.653	0.732	0.035	0.627	0.765
		30	0.003	0.629	0.772	-0.014	0.601	0.795	0.004	0.592	0.792
	100	10	-0.008	0.683	0.723	0.007	0.676	0.729	-0.010	0.660	0.752
		15	0.002	0.657	0.746	0.000	0.644	0.758	0.007	0.623	0.774
		30	0.003	0.599	0.796	-0.001	0.615	0.783	0.015	0.572	0.816
	200	10	-0.003	0.673	0.733	0.003	0.674	0.732	0.001	0.655	0.752
		15	-0.004	0.649	0.755	0.006	0.647	0.757	0.005	0.628	0.777
		30	-0.002	0.601	0.795	0.004	0.608	0.789	0.005	0.586	0.809
-0.6	50	10	0.010	0.668	0.736	-0.015	0.666	0.739	0.010	0.690	0.713
		15	-0.002	0.627	0.773	-0.001	0.628	0.773	-0.023	0.676	0.720
		30	0.004	0.609	0.787	-0.002	0.592	0.801	0.009	0.633	0.780
	100	10	-0.002	0.650	0.754	0.000	0.657	0.746	0.010	0.694	0.728
		15	0.003	0.639	0.763	-0.004	0.629	0.772	0.009	0.675	0.726
		30	0.019	0.593	0.800	-0.015	0.588	0.805	-0.004	0.635	0.763
	200	10	0.000	0.649	0.754	0.001	0.650	0.755	0.009	0.710	0.705
		15	-0.003	0.622	0.778	0.003	0.628	0.773	-0.001	0.662	0.748
		30	-0.004	0.584	0.808	0.005	0.589	0.803	0.002	0.616	0.788

表 A5. 模拟研究 2 题目参数估计返真性

S-C	g1			s1			g2			s2			g			s		
	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor
5-25	0.085	0.109	0.660	0.080	0.106	0.707	0.017	0.046	0.916	0.041	0.081	0.821	0.028	0.057	0.876	0.048	0.086	0.803
10-20	0.023	0.051	0.861	0.024	0.061	0.809	0.021	0.053	0.894	0.042	0.082	0.816	0.022	0.052	0.887	0.036	0.075	0.826
15-15	0.011	0.038	0.910	0.026	0.054	0.885	0.021	0.052	0.876	0.037	0.073	0.824	0.016	0.045	0.901	0.031	0.064	0.863
20-10	0.010	0.033	0.950	0.015	0.047	0.919	0.008	0.041	0.918	0.067	0.101	0.748	0.009	0.035	0.950	0.032	0.065	0.854
25-5	0.008	0.033	0.954	0.018	0.050	0.918	0.006	0.038	0.892	0.120	0.160	0.589	0.007	0.034	0.953	0.035	0.068	0.833

表 A6. 模拟研究 2 认知属性判准率

S-C	ACCR_A					ACCR_B					ACCR_AB					PCCR_A	PCCR_B	PCCR_AB
	a1	a2	a3	a4	a5	a1	a2	a3	a4	a5	a1	a2	a3	a4	a5	A	B	AB
5-25	0.875	0.838	0.855	0.857	0.887	0.875	0.838	0.855	0.857	0.887	0.875	0.838	0.855	0.857	0.887	0.490	0.495	0.816
10-20	0.934	0.899	0.900	0.919	0.912	0.934	0.899	0.900	0.919	0.912	0.934	0.899	0.900	0.919	0.912	0.650	0.637	0.763
15-15	0.955	0.922	0.920	0.924	0.914	0.955	0.922	0.920	0.924	0.914	0.955	0.922	0.920	0.924	0.914	0.714	0.723	0.726
20-10	0.969	0.963	0.967	0.964	0.951	0.969	0.963	0.967	0.964	0.951	0.969	0.963	0.967	0.964	0.951	0.834	0.827	0.498
25-5	0.991	0.976	0.968	0.967	0.964	0.991	0.976	0.968	0.967	0.964	0.991	0.976	0.968	0.967	0.964	0.879	0.882	0.344

表 A7. 模拟研究 2 认知能力估计返真性

S-C	A			B			AB		
	Bias	RMSE	Cor	Bias	RMSE	Cor	Bias	RMSE	Cor
5-25	0.003	0.648	0.755	-0.004	0.629	0.771	0.010	0.564	0.822
10-20	0.001	0.606	0.789	0.001	0.599	0.796	-0.007	0.547	0.836
15-15	-0.004	0.589	0.803	0.004	0.586	0.805	-0.005	0.550	0.836
20-10	-0.002	0.578	0.811	0.000	0.593	0.800	-0.005	0.578	0.807
25-5	0.005	0.576	0.812	-0.001	0.575	0.813	-0.008	0.592	0.798