

作答时间与反应依赖关系建模： 基于双因子模型视角^{*}

郭小军¹ 柏小云¹ 罗照盛²

(¹ 赣南师范大学教育科学学院/教育经济研究中心, 江西 赣州 341000)

(² 江西师范大学心理学院, 南昌 330022)

摘要 在心理与教育测验中, 测验的计算机化越来越普遍, 使得被试作答的过程性数据的搜集也越来越便利。分层模型的提出为作答时间与反应的联合分析提供了一个基本的建模框架, 且逐渐成为当前最流行的方法。虽然分层模型被广泛使用, 但仅仅通过参数间的关系还不能很好地解释作答时间和反应之间的关系。因此, 一些研究者提出了一系列改进模型, 但仍然存在一些不足。基于双因子模型的新视角, 文中将测验的作答时间与反应分别视为测量被试速度和能力的两个局部因子, 而作答时间与反应又视为综合测量了被试的速度与准确率权衡的一般能力或全局因子。基于此, 文中提出双因子分层模型, 以探讨作答时间与反应的依赖关系。模拟研究发现 Mplus 程序能有效估计双因子分层模型的各参数, 而忽视作答时间与反应依赖关系的分层模型的参数估计结果存在明显的偏差。在实例数据分析中, 相较于分层模型, 双因子分层模型的各模型拟合指数表现更好。此外, 不同被试在不同项目上的作答时间与反应存在不同的依赖关系, 从而对被试的作答准确率与时间产生不同的影响。

关键词 作答时间, 反应, 依赖关系, 分层模型, 双因子模型

分类号 B841

1 引言

随着计算机在测验实施中的普及, 心理与教育测验也通常以计算机化的形式进行。计算机化的测验不仅能非常便捷地采集被试的作答反应数据, 还能获得被试的作答时间数据。在被试的作答数据中, 不仅作答反应能反映被试的重要作答信息, 作答时间也是被试的重要信息来源(Luce, 1986; van der Linden, 2009)。为了联合分析被试的作答时间与反应, van der Linden (2007)提出了分层模型(Hierarchical Model, HM), 它是目前最流行的建模方法之一。HM 被广泛地应用于传统测验中(Klein Entink, Kuhn, et al., 2009; Lu et al., 2021; Scherer et al., 2015), 并逐渐扩展到认知诊断测验(Liang et al., 2023; Zhan et al., 2018)、人格测验(Ranger, 2013)以

及认知实验(Loeys et al., 2011)的数据分析领域。

在 HM 建模框架中, 作答时间与反应模型构成第一层, 被试参数间的相关与项目参数间的相关构成第二层(van der Linden, 2007)。一些研究者发现作答时间与反应间存在依赖关系(Bolsinova & Tilmstra, 2016; van der Linden & Glas, 2010)。分层建模通过第二层参数间的相关来解释作答时间和反应之间的相关性, 但是仅仅通过参数间的相关只能解释作答时间与反应的部分相关。因此, 传统的 HM 无法有效地解释作答时间和反应之间的相关性影响。为了更好地反映作答时间与反应之间的依赖关系, 而不仅仅通过参数之间的相关性来获得, 一些研究者将作答时间与反应的依赖关系融入到分层模型中, 以进一步扩展模型。在作答时间与反应的残差相关上, Ranger 和 Ortner (2012)通过项目水平参数进一

收稿日期: 2023-06-27

^{*} 江西省社会科学“十四五”基金项目(21JY49)、江西省教育厅科学技术研究项目(GJJ2201241)资助。

通信作者: 罗照盛, E-mail: luozs@126.com

步解释两者的残差相关, 但并未考虑被试水平的依赖性; Meng 等(2015)则基于被试与项目参数乘积来解释残差相关, 但是他们将项目参数固定为非负数, 导致在作答时间和反应之间的残差相关性上具有相同的符号。另外, 一些研究者将残差时间纳入作答反应模型中。Bolsinova, De Boeck 和 Tijmstra (2017)在反应模型中引入了残差时间, 并允许项目的残差相关性在符号上不同, 但并未考虑被试水平的差异。为了弥补这一不足, Bolsinova, Tijmstra 和 Molenaar (2017)进一步扩展至被试水平。Bolsinova 和 Molenaar (2018)则考虑了残差时间对作答反应的非线性影响。然而, 将残差时间直接添加到作答反应模型中会导致被试在所有项目上的作答准确率可以始终趋近于 1, 违背了项目反应理论的基本假设。因此, 一些研究者单独构建线性(Guo et al., 2020)和非线性模型(郭小军, 罗照盛, 2019), 以探究残差时间对作答反应的影响, 但未考虑被试水平的差异。从上述模型可以看出, 不同研究者提出了一系列的模型, 然而总是存在不同方面的弊端, 模型也变得越来越复杂, 估计方法也更加困难, 这并不利于 HM 的应用。

在分析测验项目的依赖关系时, 最常用的项目反应理论模型包括双因子模型(bifactor model)和题组模型(testlet model)。双因子模型通过一般或全局因子(general factor)解释所有项目的共同变异, 同时使用局部因子或特定因子(specific factor)解释额外部分项目的共同变异(毛秀珍 等, 2018)。比如在瑞典标准推理测验中, 不同维度的项目作答时间分别测了被试的知觉辨别速度、类同比较速度和比较推理速度等等, 同时整个测验的作答时间也测量了被试的一般速度特质(郭小军 等, 2022)。在题组模型中, 多个项目共用相同刺激的情况构成题组(testlet), 此时被试对这些项目的作答依赖于对该内容整体的理解(詹沛达 等, 2013)。例如, 在篇章阅读测验中, 属于同一篇章的多个项目就构成一个典型的题组(郑蝉金 等, 2011)。关于双因子模型与题组模型的关系, 双因子模型通过特定因子与全局因子的载荷之比来反映题组效应的大小, 而题组模型是双因子模型的一个特例, 双因子模型更为复杂(魏丹 等, 2017), 因此双因子模型更具有一般性。在心理与教育测验中, 不仅可以采集被试在测验项目上的作答时间, 还可以采集被试的作答反应。从双因子模型的视角看, 测验项目的作答时间与反应可以视为测量了不同的局部因子。具体而言, 测验

的作答时间测量了被试的速度特质, 而测验的作答反应测量了被试的能力特质。同时, 被试在作答测验过程中还会受到时间与准确率的综合影响(Bolsinova, Tijmstra, & Molenaar, 2017; Davison et al., 2012), 即一般潜在特质或全局因子, 或速度与准确率权衡能力(Chen et al., 2018)。因此, 这样的测验结构非常符合双因子模型的特点, 为解决测验作答时间与反应依赖关系的联合分析提供了全新视角。

综合上述研究可以发现, 已考虑作答时间与反应依赖关系的 HM 存在不同方面的弊端。此外, 由于贝叶斯估计方法对数学基础要求较高以及适合该模型的统计软件的限制, 阻碍了 HM 的广泛应用。然而, 基于双因子模型视角提出的双因子分层模型(Bifactor hierarchical model, Bi-HM)为作答时间与反应的依赖关系提供了新的思路。而且 Bi-HM 可以使用标准的潜变量建模软件进行拟合分析, 例如 Mplus、LISREL 和 Amos 等, 这将极大地促进 Bi-HM 的应用与推广。本文内容结构安排如下: 首先回顾 van der Linden (2007)的 HM; 其次介绍 Bi-HM 的提出; 然后通过模拟研究评估 Bi-HM 参数估计效果; 接下来利用实例数据评估模型的拟合效果; 最后进行总结与展望。

2 双因子分层模型

2.1 模型构建

在介绍 Bi-HM 模型前, 先介绍 van der Linden (2007)提出的 HM。在 HM 中, 第一层是作答反应与作答时间模型。作答反应模型为两参数 logistic 模型, 可以表示为

$$P(X_{ij}=1|\theta_i)=\frac{\exp(a_j\theta_i-d_j)}{1+\exp(a_j\theta_i-d_j)} \quad (1)$$

其中 X_{ij} 为被试 i 在项目 j 上的作答反应, 1 表示作答正确, 0 表示作答错误; a_j 为项目的区分度参数, d_j 为项目难度相关参数, θ_i 为被试的能力特质。

被试 i 在项目 j 上的对数作答时间 $\log RT_{ij}$ 服从正态分布, 作答时间模型可以表示为

$$\log RT_{ij} \sim N(\beta_j - \tau_i, \alpha_j^{-2}) \quad (2)$$

其中 β_j 是项目作答时间强度参数, α_j 是项目作答时间的区分度参数, τ_i 为被试作答速度特质。

在 HM 的第二层为被试参数间的相关与项目参数间的相关, 通常假设被试能力特质与速度特质服从二元正态分布, 项目难度相关参数与作答时间强度参数服从二元正态分布, 如式(3)和式(4)所示。

$$\begin{bmatrix} \theta_i \\ \tau_i \end{bmatrix} \sim MVN \left(\mu_P = [\mu_\theta \quad \mu_\tau], \Sigma_P = \begin{bmatrix} \sigma_\theta^2 & \sigma_{\theta\tau} \\ \sigma_{\theta\tau} & \sigma_\tau^2 \end{bmatrix} \right) \quad (3)$$

$$\begin{bmatrix} d_j \\ \beta_j \end{bmatrix} \sim MVN \left(\mu_I = [\mu_d \quad \mu_\beta], \Sigma_I = \begin{bmatrix} \sigma_d^2 & \sigma_{d\beta} \\ \sigma_{d\beta} & \sigma_\beta^2 \end{bmatrix} \right) \quad (4)$$

在对测验的作答反应与时间进行分析时, HM 仅仅通过被试参数间的关系(式 3)和项目参数间的关系(式 4)来量化作答时间与反应的关系是不够的(如图 1a 所示)。Bi-HM 将测验的作答时间与反应视为测量被试能力和速度的两个局部因子, 同时测验的作答时间与反应又综合测量了被试的一般潜在特质或全局因子, 以及速度与准确率权衡能力(如图 1b 所示)。因此, 在借鉴双因子模型思路基础上提出 Bi-HM。

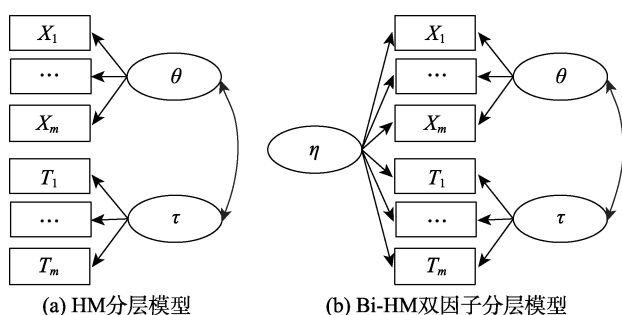


图 1 分层模型与双因子分层模型示意图

在 Bi-HM 的第一层模型中, 作答反应模型与作答时间模型分别可以表示为

$$P(X_{ij} = 1 | \theta_i) = \frac{\exp(a_j \theta_i + \lambda_{1j} \eta_i - d_j)}{1 + \exp(a_j \theta_i + \lambda_{1j} \eta_i - d_j)} \quad (5)$$

$$\log RT_{ij} \sim N(\beta_j - \tau_i + \lambda_{2j} \eta_i, \alpha_j^{-2}) \quad (6)$$

其中 η_i 反映了被试的速度与准确性之间的依赖关系或速度与准确率之间的权衡能力; λ_{1j} 反映了速度与准确率权衡对项目的准确率的影响, λ_{2j} 反映速度与准确率权衡对项目的作答时间的影响; 其他参数内涵与 HM 相同。

项目参数 λ 和被试参数 η 可以联合反映被试在五种不同作答时间与反应之间的关系, 具体结果如表 1 所示。从表中可以看到不同类型的表现: 在类

型一中, 作答时间与作答准确率是相互独立的, 它们之间没有影响; 在类型二中, 被试更注重准确率, 会花费更多的作答时间来提高准确率, 这类似速度与准确率权衡实验强调准确率而忽视时间; 类型三的被试则更注重作答速度, 因此会通过牺牲准确率来提高作答速度或降低作答时间, 这和速度与准确率权衡实验强调作答速度而忽视准确率的表现类似(郭小军 等, 2019); 类型四表现为最佳作答模式, 被试能够同时保持快速和准确地作答; 最后一种类型五与类型四表现完全相反, 被试既作答拖拉又准确率低。

在 Bi-HM 的第二层中, 速度与准确率权衡能力 η_i 作为一般潜在特质, 通常假设一般特质与局部特质之间是彼此独立的(Cai et al., 2011), 但是允许局部特质之间存在相关。因此, 被试能力与速度特质假设服从二元正态分布(式 3), 而项目难度参数与作答时间强度参数也服从二元正态分布(式 4)。

2.2 参数估计与模型识别

对 HM 与 Bi-HM 的参数估计, 采用 Mplus (version 8.1) (Muthén & Muthén, 2019) 中的稳健极大似然估计(Maximum Likelihood Robust, MLR)。在模型识别上, HM 将能力特质 θ 与速度特质 τ 的方差分别固定为 1 (van der Linden, 2007), 而 Bi-HM 需要将速度与准确率权衡特质 η 方差固定为 1, 并且将速度与准确率权衡特质 η 分别与能力 θ 以及速度 τ 的相关固定为 0 (Cai et al., 2011)。此外, Mplus 参数估计无法考虑项目参数间的相关, 但是研究发现忽视项目参数间的相关对 HM 的参数估计并没有影响(Molenaar et al., 2015), 甚至有些拓展的 HM 并未考虑项目参数间的关系(Wang et al., 2013)。需要获取相关 Mplus 代码的研究者可以联系通讯作者。

3 模拟研究

对模拟研究, 主要基于两个目的展开探究, 其一是不同条件对 Bi-HM 参数估计的返真性的影响; 其二是忽视作答时间与反应关系对传统分层模型 HM 所带来的影响。

表 1 作答时间与反应的不同关系类型

参数	关系类型							
	类型一	类型二	类型三	类型四	类型五	类型六	类型七	类型八
λ_{1j}	0	< 0	> 0	> 0	< 0	> 0	< 0	> 0
λ_{2j}	0	< 0	> 0	> 0	< 0	< 0	> 0	< 0
η_i	—	< 0	> 0	< 0	> 0	> 0	< 0	< 0
关系内涵	独立	注重准确率, 忽视速度。	注重作答速度, 忽视准确率。	既注重准确率, 又注重速度。	作答速度慢且准确率低。			

3.1 评价指标

对于模拟结果采用 MSE (Mean squared error) 和平均 Bias 进行评价, MSE 和 Bias 值越小意味参数的估计值与模拟值越接近, 结果越佳。

$$MSE(\hat{\xi}) = \frac{\sum_{r=1}^R \sum_{k=1}^K (\hat{\xi} - \xi)^2}{R * K} \quad (7)$$

$$Bias(\hat{\xi}) = \frac{\sum_{r=1}^R \sum_{k=1}^K (\hat{\xi} - \xi)}{R * K} \quad (8)$$

其中 $\hat{\xi}$ 和 ξ 分别为估计值和模拟值, R 为重复次数, K 为被试量或测验项目量。

3.2 数据生成与分析

在模拟研究中, 根据三个条件展开模拟, 分别为被试量 N (500 和 1000)、测验项目量 m (30 和 60) 以及项目难度相关参数 d_j 与作答时间强度参数 β_j 间的相关(0、0.4 和 0.8)。总共有 12 个不同的组合水平, 每个组合水平重复模拟 50 次, 并分别计算每个参数的 Bias 和 MSE, 以评估参数估计的准确性。

在数据生成过程中, 首先生成被试参数, 根据均值向量为 $\mu_p = (0, 0)$, 方差矩阵中 $\sigma_\theta^2 = \sigma_\tau^2 = 1$, 协方差 $\sigma_{\theta\tau}$ 根据均匀分布 $U(0.2, 0.6)$ 随机生成, 再根据式(3)二元正态分布生成被试能力 θ_i 和速度 τ_i 值(van der Linden, 2007); 被试的速度与准确率权衡能力服从标准正态分布 $\eta_i \sim N(0, 1)$ (Bolsinova,

Tijmstra, & Molenaar, 2017)。其次生成项目参数, 项目难度相关参数 d_j 与作答时间强度参数 β_j 的均值向量为 $\mu_l = (0, 4)$, 方差矩阵中方差 $\sigma_d^2 = \sigma_\beta^2 = 1$, 协方差 $\sigma_{d\beta}$ 为相关系数, 分别设置为 0, 0.4 和 0.8, 再根据式(4)生成每个项目的难度相关参数 d_j 与时间强度参数 β_j 值(Meng et al., 2015); 项目区分度参数 a_j 和项目作答时间区分度参数 α_j 服从左截尾的正态分布 $N(0, 1)I(0,)$ (Meng et al., 2015); 项目水平的速度与准确率权衡参数 λ_{1j} 和 λ_{2j} 服从正态分布 $N(0.5, 1)$ (Guo et al., 2020)。最后, 根据 Bi-HM 的作答反应模型(式 5)与作答时间模型(式 6)生成每个被试在每个项目上的作答反应 X_{ij} 和对数时间 $\log RT_{ij}$ 。

3.3 结果

对 Bi-HM 和 HM 的各参数的估计结果, 通过矩阵图和表格分别进行了呈现, 以更好的反映其趋势和特点。在矩阵图中, 每一行的图形对应 HM 和 Bi-HM 的 MSE 和 Bias 值, 每一列的图形对应不同被试量和测验项目量的组合; 在每一个行列交叉的图形中, 横坐标为 HM 和 Bi-HM 对应的项目参数或被试参数, 纵坐标为 MSE 或 Bias 的值, 不同线段对应不同的项目相关 $\rho_{d\beta}$ 水平。

图 2 和表 2 呈现了 HM 和 Bi-HM 的各项参数的估计结果。从图表中可以看出, 整体而言, Bi-HM 的各参数估计结果较好, 而 HM 的参数估计

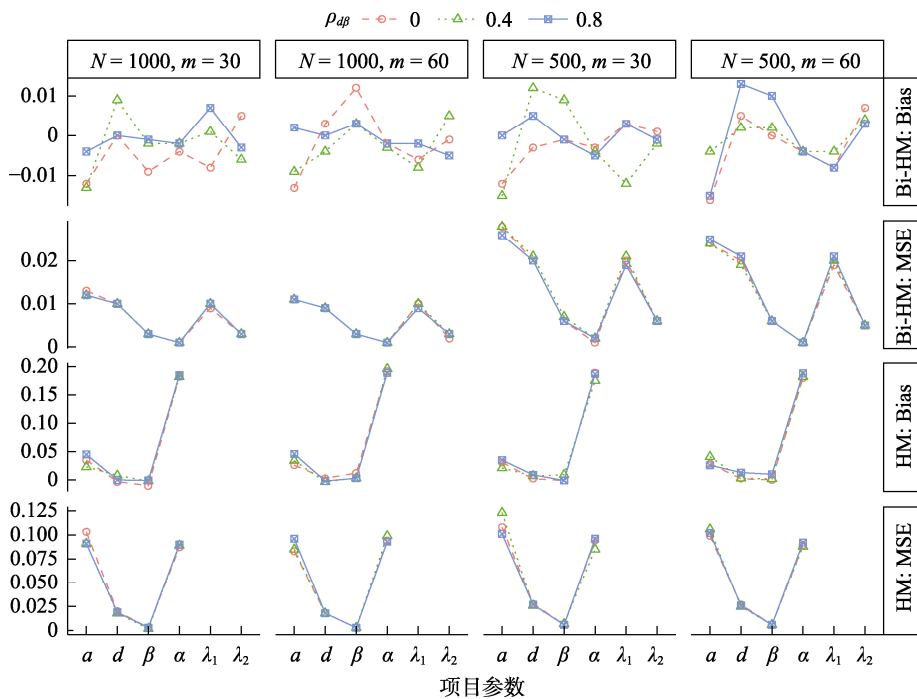


图 2 不同条件 HM 和 Bi-HM 项目参数估计效果

表 2 HM 与 Bi-HM 项目参数返真性

项目参数	模型	$\rho_{b\beta}$	$N = 500$				$N=1000$			
			$m = 30$		$m = 60$		$m = 30$		$m = 60$	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
a	HM	0	0.108	0.031	0.099	0.029	0.103	0.035	0.083	0.026
		0.4	0.123	0.021	0.106	0.041	0.091	0.023	0.085	0.034
		0.8	0.101	0.035	0.102	0.026	0.091	0.045	0.096	0.046
	Bi-HM	0	0.028	-0.012	0.024	-0.016	0.013	-0.012	0.011	-0.013
		0.4	0.028	-0.015	0.024	-0.004	0.012	-0.013	0.011	-0.009
		0.8	0.026	0	0.025	-0.015	0.012	-0.004	0.011	0.002
d	HM	0	0.028	0.002	0.027	0.002	0.02	-0.003	0.018	0.003
		0.4	0.026	0.007	0.025	0.003	0.018	0.008	0.018	-0.003
		0.8	0.027	0.009	0.026	0.013	0.019	0	0.018	-0.002
	Bi-HM	0	0.02	-0.003	0.02	0.005	0.01	0	0.009	0.003
		0.4	0.021	0.012	0.019	0.002	0.01	0.009	0.009	-0.004
		0.8	0.02	0.005	0.021	0.013	0.01	0	0.009	0
β	HM	0	0.006	-0.001	0.006	0	0.003	-0.01	0.003	0.012
		0.4	0.007	0.009	0.006	0.002	0.002	-0.002	0.003	0.003
		0.8	0.006	-0.001	0.006	0.01	0.003	-0.001	0.003	0.003
	Bi-HM	0	0.006	-0.001	0.006	0	0.003	-0.009	0.003	0.012
		0.4	0.007	0.009	0.006	0.002	0.003	-0.002	0.003	0.003
		0.8	0.006	-0.001	0.006	0.01	0.003	-0.001	0.003	0.003
α	HM	0	0.094	0.189	0.089	0.18	0.087	0.182	0.094	0.192
		0.4	0.085	0.175	0.088	0.182	0.089	0.182	0.099	0.196
		0.8	0.096	0.187	0.092	0.188	0.09	0.185	0.093	0.189
	Bi-HM	0	0.001	-0.003	0.001	-0.004	0.001	-0.004	0.001	-0.002
		0.4	0.002	-0.004	0.001	-0.004	0.001	-0.002	0.001	-0.003
		0.8	0.002	-0.005	0.001	-0.004	0.001	-0.002	0.001	-0.002
λ_1	Bi-HM	0	0.02	0.003	0.019	-0.008	0.009	-0.008	0.01	-0.006
		0.4	0.021	-0.012	0.02	-0.004	0.01	0.001	0.01	-0.008
		0.8	0.019	0.003	0.021	-0.008	0.01	0.007	0.009	-0.002
λ_2	Bi-HM	0	0.006	0.001	0.005	0.007	0.003	0.005	0.002	-0.001
		0.4	0.006	-0.002	0.005	0.004	0.003	-0.006	0.003	0.005
		0.8	0.006	-0.001	0.005	0.003	0.003	-0.003	0.003	-0.005

注：HM = van der Linden 的分层模型; Bi-HM = 双因子分层模型。

结果相对较差。在相同被试量($N = 500$ 或 1000)上, Bi-HM 和 HM 的各自项目参数的 MSE 值趋向水平相似, Bias 波动较小。随着被试量从 500 增加到 1000, Bi-HM 的各项参数的 MSE 值都呈现下降趋势, 特别是反应模型的区分度参数 a 的 MSE 值从 0.028 降低到 0.011, 难度相关参数 d 的 MSE 从 0.02 降低到 0.009, λ_1 的 MSE 从 0.021 降低到 0.009, 作答时间模型的各项参数有轻微降低, 且各参数的最大 MSE 值为 0.007; 所有项目参数的 Bias 值在接近 0 的范围内波动。HM 的各项参数的 MSE 值也随着被试量的增加而略微下降, 但是其反应模型的 a

参数的 MSE 从 0.123 降低到 0.083, MSE 值仍然较高, 估计结果较差, 作答时间模型的 α 参数的 MSE 和 Bias 值也都明显偏高, MSE 值始终在 0.096 附近, Bias 值则在 0.19 附近, 其他项目参数的 MSE 值和 Bias 值都相对偏小, 估计较好。在不同项目相关 $\rho_{d\beta}$ 水平上, HM 的各项参数的 MSE 和 Bias 值保持稳定, Bi-HM 的各项参数的 MSE 值也保持不变, 虽然 Bi-HM 的 Bias 值有所波动, 但差异非常微小。

HM 和 Bi-HM 的被试参数的估计结果如图 3 和表 3 所示。从图表中可以看出, 在相同测验项目量($m = 30$ 或 60)上, Bi-HM 和 HM 间的各被试参数

的MSE值随着被试量变化趋势接近, Bias值虽然趋势不同, 但是差异较小。随着测验项目量从30增加到60, Bi-HM的各被试参数的MSE值都明显降低, 特别是能力参数 θ 的MSE值从0.157降低到0.086, 速度与准确率权衡参数 η 的MSE值从0.057降低到0.024; HM被试参数的MSE值随着测验项目量增加

而有所降低, 虽然被试参数能力参数 θ 的MSE从0.283降低到0.194, 但值仍然较高, 估计偏差较大。相对Bi-HM和HM被试参数的MSE值, 两个模型的被试参数的Bias值则都在0附近波动。此外, Bi-HM和HM的各被试参数的MSE值在不同项目相关 $\rho_{d\beta}$ 水平上趋于稳定, 特别是Bi-HM的被试参

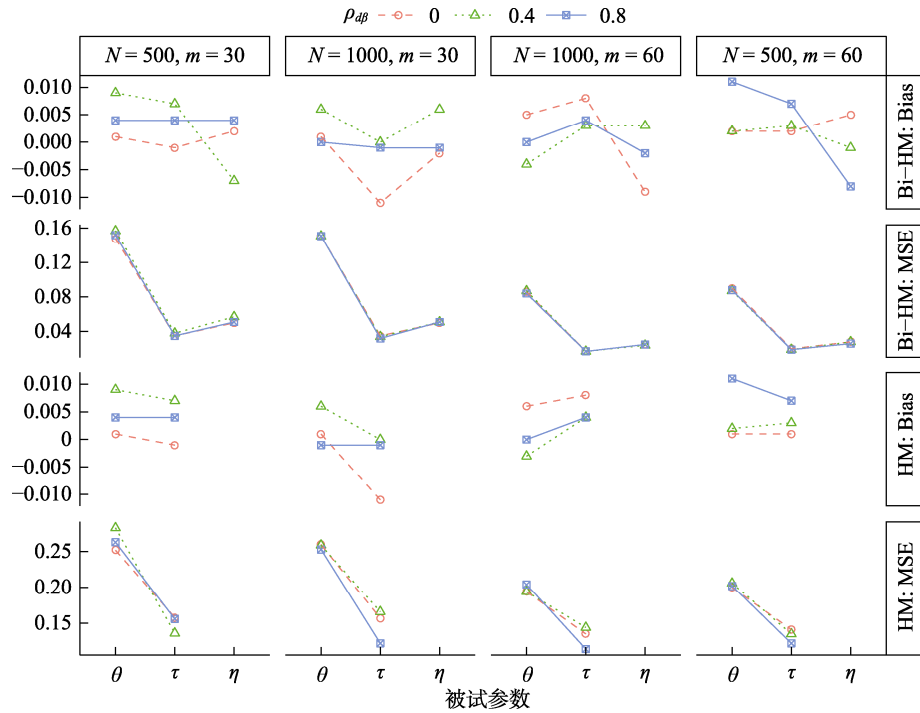


图3 不同条件HM和Bi-HM被试参数估计效果

表3 HM与Bi-HM被试参数返真性

被试参数	模型	$\rho_{d\beta}$	N = 500				N = 1000			
			m = 30		m = 60		m = 30		m = 60	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
θ	HM	0	0.252	0.001	0.2	0.001	0.26	0.001	0.194	0.006
		0.4	0.283	0.009	0.206	0.002	0.259	0.006	0.194	-0.003
		0.8	0.263	0.004	0.202	0.011	0.253	-0.001	0.204	0
	Bi-HM	0	0.148	0.001	0.09	0.002	0.151	0.001	0.086	0.005
		0.4	0.157	0.009	0.087	0.002	0.151	0.006	0.087	-0.004
		0.8	0.152	0.004	0.088	0.011	0.151	0	0.084	0
τ	HM	0	0.158	-0.001	0.141	0.001	0.157	-0.011	0.135	0.008
		0.4	0.136	0.007	0.135	0.003	0.166	0	0.144	0.004
		0.8	0.156	0.004	0.122	0.007	0.122	-0.001	0.114	0.004
	Bi-HM	0	0.035	-0.001	0.02	0.002	0.035	-0.011	0.017	0.008
		0.4	0.038	0.007	0.019	0.003	0.034	0	0.017	0.003
		0.8	0.035	0.004	0.019	0.007	0.032	-0.001	0.017	0.004
η	Bi-HM	0	0.05	0.002	0.028	0.005	0.05	-0.002	0.025	-0.009
		0.4	0.057	-0.007	0.028	-0.001	0.051	0.006	0.024	0.003
		0.8	0.051	0.004	0.026	-0.008	0.051	-0.001	0.025	-0.002

注: HM = van der Linden 的分层模型; Bi-HM = 双因子分层模型。

数的波动完全重叠, Bias 值则呈现随机波动趋势,但是差异较小。

综合而言, Bi-HM 的各参数能够通过 Mplus 程序得到较好的估计结果,而忽视作答反应和时间依赖关系的 HM 的参数估计结果存在明显的偏差。同时,在估计过程中,忽视 $\rho_{d\beta}$ 相关对参数估计结果的影响可以忽略不计。

4 实例数据分析

4.1 数据描述

通过 E-prime 2.0 搜集《瑞文标准推理测验》作答数据,每个项目作答结束后会进入下一个项目,不能返回修改,共 60 个项目,要求被试在 45 分钟内完成。该测验共施测 427 名大学生,原始作答时间数据均事先进行对数转换。

4.2 分析

为了对《瑞文标准推理测验》的作答反应与时间数据进行深入分析,采用 van der Linden 提出的分层模型(HM)和文中提出的双因子分层模型(Bi-HM)来拟合数据。同时,已有研究从项目水平和被试水平分别探讨了作答时间与反应的依赖关系(Bolsinova, Tijmstra, & Molenaar, 2017; Meng et al., 2015; Ranger & Ortner, 2012)。因此,为了进一步探讨 Bi-HM 模型的性能,将增加固定项目参数 λ_{1j} 和 λ_{2j} 为 1 的双因子分层模型(Fix-Bi-HM)来拟合数据。不同模型优劣采用 LL (Loglikelihood)、AIC (Akaike's information criterion)、BIC (Bayesian information criterion) 和 SABIC(Sample-Size Adjusted BIC)四个模型拟合指数进行评价。在不同模型比较中,LL 值越大,模型拟合越好,而其他三个模型拟合指数值越小,该模型拟合效果越好,也越倾向选择该模型为最优模型。

4.3 结果

《瑞文标准推理测验》数据的模型拟合结果如表 4 所示。从表中可以看出, HM 的各模型拟合指数 AIC、BIC 和 SABIC 要小于 Fix-Bi-HM 的各拟合指数,但是明显大于 Bi-HM 的拟合指数,同时 Bi-HM 的 LL 值最大。这表明固定作答时间与反应的项目水平依赖关系后, Bi-HM 拟合急剧变差。同时, Bi-HM 更能符合实际数据的特征,具有更好的拟合效果。

为了更进一步比较 HM 和 Bi-HM 的拟合效果,基于参数估计结果计算了两个模型作答时间的残差,并通过取绝对值后计算项目的平均绝对残差,

表 4 瑞文标准推理测验数据分析中模型-数据拟合指标

模型	NP	LL	AIC	BIC	SABIC
HM	241	-28652.849	57787.698	58765.383	58000.598
Fix-Bi-HM	241	-28740.198	57962.396	58940.081	58175.296
Bi-HM	361	-27862.329	56446.657	57911.156	56765.566

注: NP = 自由参数量; LL = Loglikelihood; AIC = Akaike's information criterion; BIC = Bayesian information criterion; SABIC = Sample-Size Adjusted BIC。

同时结合残差方差 α_j^{-2} 进行比较。分别以 HM 的值为纵轴,以 Bi-HM 的值为横轴,每一个项目值构成一个坐标点,绘制散点图,结果如图 4 所示,其中虚线为对角线。从图中可以看出, logRT 残差的数据点明显位于虚线 HM 的上方,这表明 HM 的各项水平的平均 logRT 残差要明显大于 Bi-HM。同时,残差方差也明显位于虚线 HM 的上方,说明 Bi-HM 的 logRT 残差的波动要明显小于 HM。因此,这些结果进一步表明 Bi-HM 能够更拟合《瑞文标准推理测验》的数据。

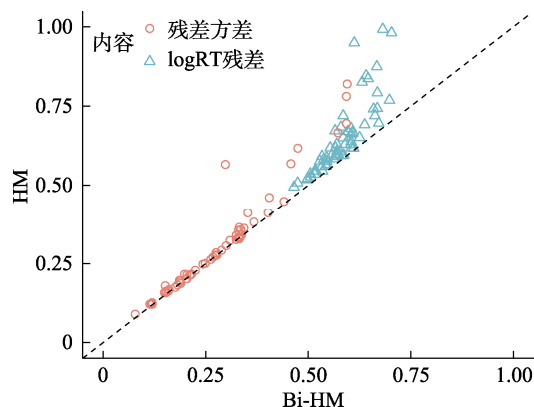


图 4 HM 和 Bi-HM 残差比较

对 Bi-HM 各参数的关系与特点,绘制项目参数的相关矩阵图,结果如图 5 所示。在相关矩阵图中,对角线的直方图为各参数的直方图分布,下三角为各参数的两两散点关系图,上三角为各参数的两两相关系数。从图中可以看出, λ_1 和 λ_2 之间的相关为 0.47,这表明在《瑞文标准推理测验》中速度与准确率权衡对项目水平的准确率和作答时间的一致性影响仅呈现中等水平。此外, λ_1 与项目难度相关参数 d 和时间强度参数 β 相关分别为 0.49 和 0.60,而 λ_2 与项目难度相关参数 d 和时间强度参数 β 相关高达 0.86 和 0.88,表明项目水平的速度与准确率权衡与项目难度相关参数 d 和时间强度参数 β 存在较高相关。最后,难度相关参数 d 和时间强度参数 β 的相关为 0.85,表明项目越难,时间强度也

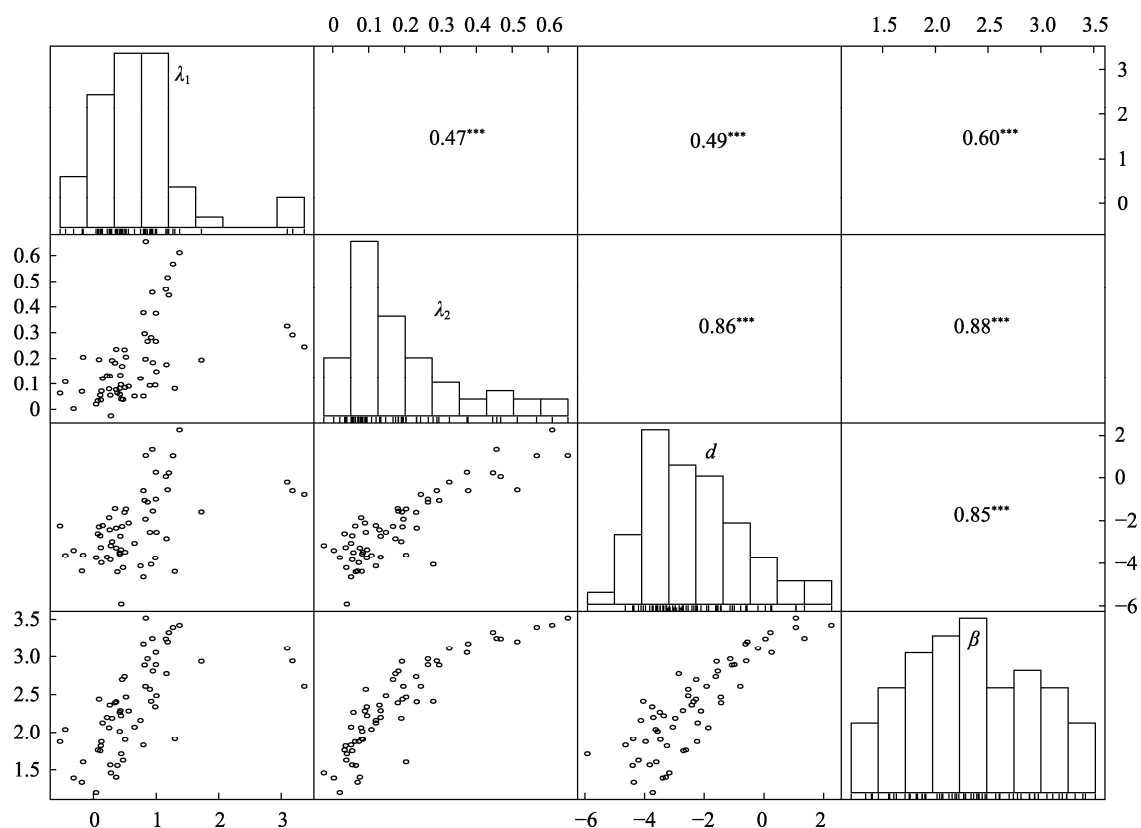


图5 Bi-HM 项目参数的相关矩阵

会越高, 需要时间越多。

在心理和教育测验中, 有许多因素会影响被试者的作答过程和结果。其中, 被试的能力水平会影响其作答的准确性, 作答速度则会影响其完成测验项目所需的时间。同时, 作答速度和准确性之间还存在一种权衡关系。这些影响因素在不同类型的测验中普遍存在, 例如能力测验、人格测验等。为了更好地考虑这些影响, 文中提出了 Bi-HM。下面以《瑞文标准推理测验》为例, 具体探讨 Bi-HM 在这一测验中的心理学分析, 以凸显该模型的特点。从

被试作答过程的总体心理来看, Bi-HM 的大部分项目的 λ_1 值为正值(除难度偏低项目 1、3、15、25 和 37 为负值), 而 λ_2 在所有项目上均为正值(图 5), 结合被试的速度与准确率权衡能力 η 的正负, 可以得出被试在作答过程中主要包括两种作答心理状态: 一种是被试通过花费更多的时间以提高准确率的作答心理(表 1 类型二), 另一种是牺牲准确率以降低作答时间的作答心理(表 1 类型三)。针对具体的项目作答心理, 根据 λ_1 和 λ_2 参数分布选择了两个典型项目(项目 55 和 37), 结果如图 6 所示。图中不

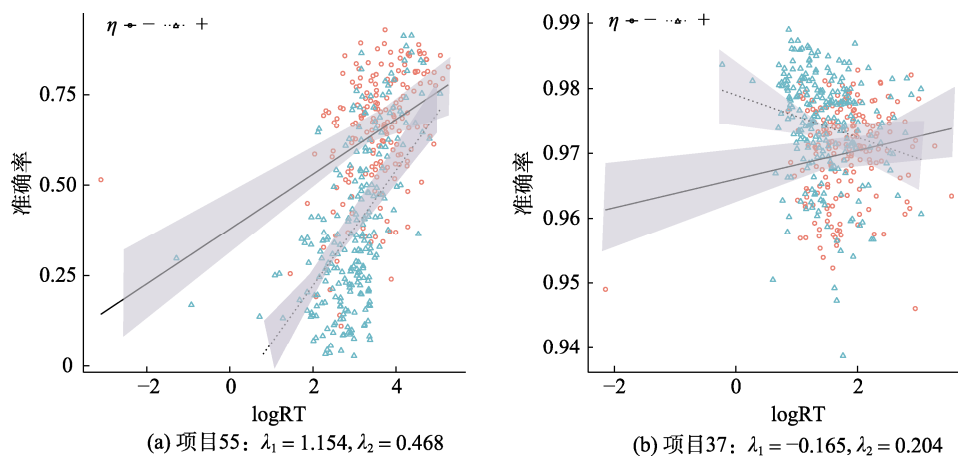


图6 不同类型的作答时间与准确率关系趋势

同直线代表速度与准确率权衡能力 η 为正或负的被试在作答过程中的作答趋势。从图 6(a)可以看出,速度与准确率权衡能力 η 为正(+)的被试在项目 55 上更倾向于花费更多时间以提高准确率(表 1 类型二),而速度与准确率权衡能力 η 为负(-)的被试在该项目上更倾向于牺牲准确率以降低作答时间(表 1 类型三)。从图 6(b)可以看出,速度与准确率权衡能力 η 为负(-)的被试在项目 37 上既注重准确率,又注重速度(表 1 类型四),而速度与准确率权衡能力 η 为正(+)的被试在该项目上的作答速度慢且准确率低(表 1 类型五)。因此,不同项目上的作答时间与反应之间存在不同类型的关系,进而对被试的作答准确率与时间产生不同影响。

5 总结和展望

作答时间与反应都反映了被试重要的作答信息,分层模型(Hierarchical Model, HM)(van der Linden, 2007)的提出为联合分析作答时间与反应提供了基本框架,并逐渐成为当前最流行的建模方法。然而, HM 只通过项目参数间的关系与被试参数间的关系来解释作答时间与反应之间的依赖关系,这在某种程度上是不够的。虽然不同研究者提出了一系列改进模型,但仍存在一些局限性(郭小军等, 2019; Bolsinova et al., 2017, 2018; Range et al., 2012)。为了更好地反映作答时间与反应间的依赖关系,文中从双因子模型视角出发,将测验的作答时间与反应分别视为测量了被试速度和能力两个局部因子,而作答时间与反应又综合测量了被试的速度与准确率权衡的一般能力或全局因子,基于此构建了双因子分层模型(Bifactor hierarchical model, Bi-HM)。Bi-HM 能灵活地反映作答时间与反应间的不同关系,并且可以通过常用的结构方程模型软件对模型参数进行估计,这对 Bi-HM 的应用非常有利。模拟研究发现,基于 MLR 的 Mplus 程序能有效估计 Bi-HM 的各参数,并且参数估计结果不受项目参数间的相关水平的影响。然而, HM 因忽视作答时间与反应间的依赖关系进而导致参数估计结果存在明显偏差。在实例数据中, Bi-HM 的模型拟合指数明显优于 HM, 并且项目的时间残差和残差方差也明显要小于 HM, 这表明考虑作答时间与反应依赖关系的 Bi-HM 更符合实例数据。同时,被试在大部分项目上的作答心理主要分为增加时间以提高准确率或牺牲准确率以降低作答时间,然而部分简单项目也存在增加作答时间后准确

率反而下降的现象(郭小军等, 2019; Bolsinova, Tijmstra, & Molenaar, 2017), 这可能是由于被试在测试期间效率发生变化的结果(Chen et al., 2018)。

值得肯定的是研究取得了较好的结果,但是仍然存在一些局限性需要进一步研究来完善。首先, Bi-HM 只对单维测验的作答时间与反应进行了模拟与实证研究,基于双因子模型的视角探究作答时间与反应的依赖关系可以从多个方面进行拓展。从测验维度上,可以将 Bi-HM 从单维测验向多维测验拓展(郭小军等, 2022; 魏丹等, 2017),构建多维的双因子分层模型,甚至高阶双因子分层模型;从过程数据的类型上,除了作答时间与反应数据外,还可以向注视点数据(詹沛达, 2022)和鼠标单击/拖动轨迹数据(Liang et al., 2023)等扩展,此时不同过程数据类型视为测量多个局部或特定因子进行双因子分层模型的构建,探讨不同过程数据间的相互影响;从作答时间与反应的依赖关系的形式上,文中只探讨了线性的关系,但是也有研究发现有的项目的作答时间与反应之间存在非线性关系(郭小军等, 2019; Bolsinova et al., 2018; Chen et al., 2018),例如将速度与准确率权衡能力定义为二项式形式。其次,虽然模拟研究发现未考虑项目参数间关系的 Bi-HM 的参数可以通过 Mplus 程序进行有效估计,但是未来的研究可以比较贝叶斯估计和 Mplus 的 MLR 估计结果,以进一步探究忽视项目参数间相关的影响(Molenaar et al., 2015)。最后,在作答时间模型的适用性上,对数正态时间模型可能并不适用所有的时间数据,因此可以考虑对模型进行不同形式的变换或者选择其他模型来更好地适应作答时间的分布(孟祥斌, 2016; Klein Entink, van der Linden, & Fox, 2009; Wang et al., 2013)。

参考文献

- Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika*, 82(4), 1126–1148.
- Bolsinova, M., & Molenaar, D. (2018). Modeling nonlinear conditional dependence between response time and accuracy. *Frontiers in Psychology*, 9, 1525.
- Bolsinova, M., & Tijmstra, J. (2016). Posterior predictive checks for conditional independence between response time and accuracy. *Journal of Educational and Behavioral Statistics*, 41(2), 123–145.
- Bolsinova, M., Tijmstra, J., & Molenaar, D. (2017). Response moderation models for conditional dependence between response time and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 70(2), 257–279.
- Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological*

- Methods*, 16(3), 221–248.
- Chen, H., De Boeck, P., Grady, M., Yang, C.-L., & Waldschmidt, D. (2018). Curvilinear dependency of response accuracy on response time in cognitive tests. *Intelligence*, 69, 16–23.
- Davison, M. L., Semmes, R., Huang, L., & Close, C. N. (2012). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement*, 72(2), 245–263.
- Guo, X., & Luo, Z. (2019). The trade-off between speed and accuracy: Evaluation and modeling for participants in the response states. *Studies of Psychology and Behavior*, 17(5), 589–595.
- [郭小军, 罗照盛. (2019). 速度与准确率权衡: 被试反应状态评价与建模. *心理与行为研究*, 17(5), 589–595.]
- Guo, X., Luo, Z., & Yan, J. (2022). Modeling of the effect on multidimensional latent speeds in the between-item multidimensional response time. *Journal of Psychological Science*, 45(5), 1222–1229.
- [郭小军, 罗照盛, 严娟. (2022). 项目间多维测验作答时间数据分析: 潜在特质速度间效应建模. *心理科学*, 45(5), 1222–1229.]
- Guo, X., Luo, Z., & Yu, X. (2020). A speed-accuracy tradeoff hierarchical model based on cognitive experiment. *Frontiers in Psychology*, 10, 2910.
- Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, 14(1), 54–75.
- Klein Entink, R. H., van der Linden, W. J., & Fox, J.-P. (2009). A box-cox normal model for response times. *British Journal of Mathematical and Statistical Psychology*, 62(3), 621–640.
- Liang, K., Tu, D., & Cai, Y. (2023). Using process data to improve classification accuracy of cognitive diagnosis model. *Multivariate Behavioral Research*, 58(5), 969–987.
- Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, 76(3), 487–503.
- Lu, J., Wang, C., & Shi, N. (2021). A mixture response time process model for aberrant behaviors and item nonresponses. *Multivariate Behavioral Research*, 58(1), 71–89.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Mao, X., Xia, M., & Xin, T. (2018). Full-information item bifactor analysis: Model parameter estimation and application. *Advances in Psychological Science*, 26(2), 358–367.
- [毛秀珍, 夏梦连, 辛涛. (2018). 全信息项目双因子分析: 模型、参数估计及其应用. *心理科学进展*, 26(2), 358–367.]
- Meng, X.-B. (2016). A log-skew-normal model for item response times. *Journal of Psychological Science*, 39(3), 727–734.
- [孟祥斌. (2016). 项目反应时间的对数偏正态模型. *心理科学*, 39(3), 727–734.]
- Meng, X.-B., Tao, J., & Chang, H.-H. (2015). A conditional joint modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement*, 52(1), 1–27.
- Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. J. (2015). A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology*, 68(2), 197–219.
- Muthén, L. K., & Muthén, B. O. (2019). *Mplus: The comprehensive modeling program for applied researchers: user's guide*. Muthén & Muthén.
- Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, 48, 37–50.
- van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72(3), 287–308.
- van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, 46(3), 247–272.
- van der Linden, W. J., & Glas, C. A. W. (2010). Statistical tests for conditional independence in a hierarchical model for speed and accuracy on test items. *Psychometrika*, 75(1), 120–139.
- Wang, C., Chang, H.-H., & Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, 66(1), 144–168.
- Wei, D., Liu, H., & Zhang, D. (2017). Multidimensional rasch testlet model: An extension and generalization of MRCMLM. *Acta Psychologica Sinica*, 49(12), 1604–1614.
- [魏丹, 刘红云, 张丹慧. (2017). 多维题组反应模型: 多维随机系数多项 Logistic 模型的应用拓展. *心理学报*, 49(12), 1604–1614.]
- Ranger, J. (2013). Modeling responses and response times in personality tests with rating scales. *Psychological Test and Assessment Modeling*, 55(4), 361–382.
- Ranger, J., & Ortner, T. (2012). The case of dependency of responses and response times: A modeling approach based on standard latent trait models. *Psychological Test and Assessment Modeling*, 54(2), 128–148.
- Zhan, P. (2022). Joint-cross-loading multimodal cognitive diagnostic modeling incorporating visual fixation counts. *Acta Psychologica Sinica*, 54(11), 1416–1423.
- [詹沛达. (2022). 引入眼动注视点的联合-交叉负载多模态认知诊断建模. *心理学报*, 54(11), 1416–1423.]
- Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, 71(2), 262–286.
- Zhan, P., Wang, W.-C., & Wang, L. (2013). Testlet response theory: An introduction and new developments. *Advances in Psychological Science*, 21(12), 2265–2280.
- [詹沛达, 王文中, 王立君. (2013). 项目反应理论新进展之题组反应理论. *心理科学进展*, 21(12), 2265–2280.]
- Zheng, C.-J., Guo, C.-Y., & Bian, Y.-F. (2011). Using testlet DIF procedures to detect testlet DIF in chinese passage-based reading testing. *Acta Psychologica Sinica*, 43(7), 830–835.
- [郑蝉金, 郭聪颖, 边玉芳. (2011). 变通的题组项目功能差异检验方法在篇章阅读测验中的应用. *心理学报*, 43(7), 830–835.]

Modeling the dependence between response and response time: A bifactor model approach

GUO Xiaojun¹, BAI Xiaoyun¹, LUO Zhaosheng²

(¹ School of Education Science / Center for Education and Economic Research, Gannan Normal University, Ganzhou 341000, China)

(² School of psychology, Jiangxi Normal University, Nanchang 330022, China)

Abstract

In the realms of psychological and educational testing, the computerization of tests is becoming more prevalent, facilitating the acquisition of process data from test-takers. In the domain of process data, response time and response represent the two most commonly utilized variables. Responses provide critical insights into the answers provided by test-takers, while response time, as an essential source of information, is increasingly garnering attention from researchers. The proposal of hierarchical model (HM) has provided a fundamental modeling framework for the joint analysis of response time and response, and it is becoming increasingly popular in current research practices. However, relying solely on the association between item and subject parameters is insufficient to adequately explain the correlation between response time and response. Consequently, researchers have proposed various enhanced models to address these limitations, although some challenges persist.

The bifactor model explains common variance through a general or global factor, while a local or specific factor explains the common variance of additional partial items. In psychological and educational testing, it is possible to capture not only the test-takers' response times on test items but also their responses. From the perspective of the bifactor model, response times and responses to test items measure different local factors. Specifically, a test's response time measures the test-taker's speed trait, while the response to the test measures their ability trait. Test-takers are also influenced by a combination of time and accuracy when responding to the test, known as general latent traits or global factors, or speed-accuracy trade-off ability. This test structure aligns well with the bifactor model and provides a new perspective on analyzing the relationship between test-taking response time and response dependence. Based on this, this study proposes a bifactor hierarchical model (Bi-HM) to explore the dependency between response time and response.

In the simulation study, it was found that the MPLUS program utilizing MLR (Maximum Likelihood Robust), could accurately estimate the parameters of the Bi-HM and was not influenced by the level of item parameter correlation. Conversely, when disregarding the relationship between response time and response in the HM, notable bias in the parameter estimates occurred. In the empirical data analysis, the Bi-HM demonstrated significantly superior model fit indices compared to the HM. Moreover, the Bi-HM effectively captured the dependency between response and response time at both the participant and item levels. This dependency is closely associated with item difficulty and time intensity factors.

Based on the findings mentioned above, it is evident that the Bi-HM, which adopts a bifactor model perspective, excels in parameter estimation and data fitting, demonstrating excellent scalability.

Keywords response time, response, dependency relationship, hierarchical model, bifactor model