

文章编号: 1000-5862(2016)02-0133-07

# 多水平多维 IRT 模型在学业质量监测中的应用

康春花 孙小坚 顾士伟 曾平飞\*

(浙江师范大学教师教育学院 浙江 金华 321004)

**摘要:** 用多水平多维 IRT( ML-MIRT) 模型对学生学业情况进行分析, 进而拓宽该模型的适用范围. 首先采用不包含任何预测变量的零模型对项目参数和学生参数进行分析, 然后使用包含学生预测变量的 ML-MIRT 模型分析 2 个预测变量对学生学业能力的影响. 研究结果表明: 试题的难度基本上贯穿了整个能力量尺, 试题也具有较好的区分度. 3 个能力维度的跨级相关系数均达到了显著性水平, 父亲的受教育程度对学生的 3 个能力维度均有显著影响, 而母亲的受教育程度无显著影响. 因此, 编制的测验较为合理, ML-MIRT 模型适用于监测学生学业水平, 父亲的受教育程度对学生学业水平有着重要影响.

**关键词:** 多水平多维 IRT 模型; 学业能力; 质量监测; 父母受教育程度

**中图分类号:** B 842.1 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2016.02.05

## 0 引言

学校效能的评估作为当前教育测验中的一个重要课题, 其结果的好坏对学生以及学校均有着重要影响. 为保证评估的公平性及质量, 选取适当的评估方法是至关重要的. 随着多水平分析技术的日臻成熟, 研究者将项目反应理论(IRT)和多水平分析技术相结合, 形成多水平项目反应理论( Multilevel Item Response Theory, MLIRT), 从而为学校效能的评估提供了一种较为可靠的评价方法. MLIRT 将项目反应模型嵌套于多水平模型中, 并对 IRT 模型中的学生能力参数进行回归分析, 以考查能力参数的随机效应和固定效应<sup>[1]</sup>. 相对于传统的效能评估方法, 多水平分析方法具有其自身的优势: (i) 将学生的变异分解成不同层次的变异, 更加科学和合理地分析了学生间存在差异的可能原因, 以及检验各因素对学生能力产生的影响<sup>[2]</sup>; (ii) 其将学生学业能力看作无法直接观测的潜在变量, 并且对该潜在变量的测量误差予以考虑, 从而降低了犯(i)错误的可能性, 增加了测验的精确性, 这对于测验评估的进步有着极大的推进作用<sup>[3-6]</sup>.

目前, MLIRT 在测验中的主要假设是单维性假设, 即测验只测量了一个能力特质. 即使某测验包含多种能力维度, 多数研究者亦将其分割成多个只测量一种能力维度的子测验, 以此来适应单维 IRT 模

型<sup>[7-8]</sup>. 该方法的优点在于对结果的解释较为通俗易懂, 较适用于只为简单地描述学生某一能力水平的测验, 如阅读理解<sup>[9]</sup>; 其不足之处在于无法体现测验中各能力之间的相似性和差异性. 换言之, 单维 IRT 模型认为测验中各能力之间是相互独立的. 而放眼于国内外关于学校效能评估方面的测验, 只单纯测量一个能力特质的情况是极为少见的, 而多数测验是同时测量多种能力. 因此, 发展一种可以同时测量多个能力特质的模型是十分有必要的, 而多维 IRT 模型便可实现该目的, 如此, 便形成了多水平多维项目反应理论( Multilevel Multidimensional Item Response Theory, ML-MIRT) 模型. 目前该模型在跨文化研究中使用较为广泛<sup>[10-11]</sup>. 本研究拟将 ML-MIRT 模型运用于学生学业质量监测中, 并探讨学生预测变量对其学业水平的影响.

## 1 ML-MIRT 简介

ML-MIRT 模型是基于单维 MLIRT 模型发展而来, 主要处理测验中包含多个能力特质的情境. 该模型可以较好地对学生在完成某一测验时所需要的多种能力特质、以及项目和答对概率之间的关系进行公式化, 从而更深入和全面地了解学生, 并对学生提供更加科学和公平的评估<sup>[7]</sup>. 由于分类标准不同, 会有不同类型的多维 IRT 模型, 本研究将对 W. Stout<sup>[12]</sup>提出的补偿性多维 IRT 模型( Compensatory

收稿日期: 2015-11-24

基金项目: 浙江省自然科学基金( LY15C090003) 和教育部基础教育质量监测中心课题( 2014AC001-D) 资助项目.

通信作者: 曾平飞( 1963-), 男, 广西荔浦人, 教授, 博士, 主要从事心理测量与评价方面的研究.

Multidimensional IRT Model, MIRT-C) 进行介绍. 针对 0-1 计分的项目, MIRT-C 的数学表达式为  $P(x_{ij} = 1 | \theta_1, \theta_2, \dots, \theta_K) = \exp(\sum_{k=1}^K a_{jk}\theta_{ik} + d_j) / (1 + \exp(\sum_{k=1}^K a_{jk}\theta_{ik} + d_j))$ , 其中  $P(x_{ij} = 1 | \theta_1, \theta_2, \dots, \theta_K)$  为学生  $i$  在项目  $j$  上的正确作答概率;  $\theta_1, \theta_2, \dots, \theta_K$  为能力的  $K$  个维度;  $a_{jk}$  为项目  $j$  在能力维度  $k$  上的区分度参数;  $d_j$  为项目  $j$  的截距参数, 其与单维 IRT 模型中的难度参数转换方法为<sup>[13]</sup>  $b_j = -d_j / MDISC_j$ ,  $MDISC_j = (\sum_{k=1}^K (a_{jk})^2)^{1/2}$ .

针对 MIRT-C 模型, 采用 logit 函数进行相应转换, 并对其中的能力参数  $\theta_{ik}$  进行回归, 便可得到 ML-MIRT 模型. 根据多水平模型的研究思路<sup>[14]</sup>, ML-MIRT 模型的研究过程主要包括以下步骤:

1) 建立不包含任何预测变量的模型(即零模型), 其函数表达式为

$$\text{logit}(p_{ij}) = \log(p_{ij} / (1 - p_{ij})) = \sum_{k=1}^K a_{jk}\theta_{ik} + d_j, \quad (1)$$

$$\theta_{ik} = \gamma_{0k} + \varepsilon_{ik}, \quad (2)$$

其中(1)式为水平 1 模型(项目水平), (2)式为水平 2 模型(学生水平).  $k(k = 1, 2, \dots, K)$  为能力维度,  $\gamma_{0k}$  为所有学生在第  $k$  维能力上的平均水平,  $\varepsilon_{ik}$  为学生  $i$  在第  $k$  维能力中的随机变异, 其服从均值为 0, 方差为  $\tau_{0k}$  的正态分布, 该值可以看成是学生  $i$  的第  $k$  维能力值.

2) 可将学生的预测变量加入到模型中, 如此便形成含学生层面预测变量的 ML-MIRT 模型, 其水平 1 模型的表达式和零模型中的水平 1 模型相同, 水平 2 模型的公式为

$$\theta_{ik} = \gamma_{0k} + \sum_{h=1}^H \gamma_{hk} X_{ihk} + \varepsilon_{ik}, \quad (3)$$

其中  $H$  为预测变量个数,  $X_{ihk}$  为预测变量,  $\gamma_{hk}$  为回归系数, 表示预测变量对学生能力的重要程度, 其它符号与(2)式相同.

3) 根据步骤 2 的分析结果, 检测(3)式中回归系数的变异程度, 该变异由更高层级的变量(设为  $W$ )而产生的, 表示学生的预测变量  $X$  和其学业能力水平之间的关系在不同的变量  $W$  下会有所不同<sup>[14]</sup>. 如此便形成了 3 个水平的 ML-MIRT 模型(如学生嵌套于学校), 其数学表达式为

(i) 水平 1 模型(项目水平):

$$\text{logit}(p_{isj}) = \log(p_{isj} / (1 - p_{isj})) = \sum_{k=1}^K a_{jk}\theta_{isk} + d_j,$$

(ii) 水平 2 模型(学生水平):

$$\theta_{isk} = \gamma_{0sk} + \sum_{h=1}^H \gamma_{hsk} X_{ihsk} + \varepsilon_{isk},$$

(iii) 水平 3 模型(学校水平):

$$\gamma_{0sk} = \lambda_{00k} + \sum_{v=1}^V \lambda_{0vk} W_{vk} + \zeta_{0sk},$$

$$\gamma_{1sk} = \lambda_{10k} + \sum_{v=1}^V \lambda_{1vk} W_{vk} + \zeta_{1sk},$$

$$\gamma_{2sk} = \lambda_{20k} + \sum_{v=1}^V \lambda_{2vk} W_{vk} + \zeta_{2sk},$$

...

$$\gamma_{Hsk} = \lambda_{H0k} + \sum_{v=1}^V \lambda_{Hvk} W_{vk} + \zeta_{Hsk},$$

其中  $s(s = 1, 2, \dots, S)$  表示学校,  $W_{vk}(v = 1, 2, \dots, V)$  为学校水平的预测变量,  $\lambda_{hvk}(h = 1, 2, \dots, H)$  为回归系数,  $\zeta_{hsk}(h = 1, 2, \dots, H)$  为随机变异, 其服从均值向量为 0, 方差-协方差矩阵为

$$\begin{pmatrix} v_{00} & v_{01} & \cdots & v_{0H} \\ v_{10} & v_{11} & \cdots & v_{1H} \\ \vdots & \vdots & \ddots & \vdots \\ v_{H0} & v_{H1} & \cdots & v_{HH} \end{pmatrix} - \mathbf{v} = (v_{ij})_{H \times H}, i, j = 1, 2, \dots, H$$

的多元正态分布<sup>[15]</sup>.

## 2 实证研究

### 2.1 目的

通过对某市初中生在科学测验中的表现进行分析, 一方面试图推广 ML-MIRT 模型在实践中的应用范围; 另一方面, 为深入了解学生学业能力水平的现状, 以及探讨影响其学业能力的因素, 为国内的教育质量监测机制的发展和完善提供方法上的借鉴.

### 2.2 被试

随机选取某市 791 名初中二年级学生, 其中男生 431 人, 占总数的 54.5%, 女生 360 人, 占总数的 45.5%; 城市学生人数 181(22.9%), 农村学生人数 512(64.7%), 其它情况(包括外来经商、外来务农等)人数为 98(12.4%). 学生将作答一份由 30 道选择题组成的科学测验, 所有学生均独自作答该测验.

### 2.3 材料

研究有 2 个测验材料, 一个为科学测验试卷, 试卷包含了 3 部分内容: 生物(项目 1 ~ 项目 12)、地理(项目 13 ~ 项目 17)和物理(项目 18 ~ 项目 30). 该测验具有较高的信度( $\alpha = 0.835$ ). 另一个材料为学生父母的受教育程度调查问卷, 问卷主要调查了学生父亲和母亲的受教育程度. 其中受教育程度包

含 4 个水平,分别为小学及以下学历(水平 1)、初中学历(水平 2)、高中学历(水平 3)和大专及以上学历(水平 4),父母受教育程度的分布情况如表 1 所示.该材料主要是为验证前人关于父母受教育情况与学生学业成绩之间的关系<sup>[16]</sup>.

表 1 父母受教育水平的分布情况

受教育程度	父亲		母亲	
	频数	百分比 /%	频数	百分比 /%
小学及以下学历	202	25.5	292	36.9
初中学历	391	49.4	348	44.0
高中学历	133	16.8	90	11.4
大专及以上学历	65	8.2	61	7.7

2.4 方法

由于测验只测量了 3 个主要内容,且每个题目只测量了其中一个内容,故本研究将使用项目间 3 维 IRT 模型进行测验分析;同时,考虑到问卷相对简单,故只考虑 2 水平的情况.研究将使用 2 个模型,模型一为零模型,表达式为(1)式和(2)式;模型二为包含学生预测变量的模型,表达式为(1)式和(3)式.模型二主要探讨在加入了学生预测变量后,其对学生学业水平的解释情况.通过对前人关于单维 IRT 模型的路径图进行拓展<sup>[17]</sup>,可得到 ML-MIRT 模型的路径图,如图 1 所示.图 1 中  $\theta_L$   $\theta_E$   $\theta_P$  分别为学生的生物、地理和物理能力; $y_1$   $y_2$   $\cdots$   $y_{30}$  为作答反应; $\xi_1$   $\xi_2$   $\cdots$   $\xi_{30}$  为项目参数(包括区分度和截距参数); $X_L$   $X_E$   $X_P$  则为学生 3 种能力的预测变量.另外,虚方框中的内容表示多维 IRT 模型,直角实方框描述的是项目水平信息,而圆角实方框则描述学生水平的信息.

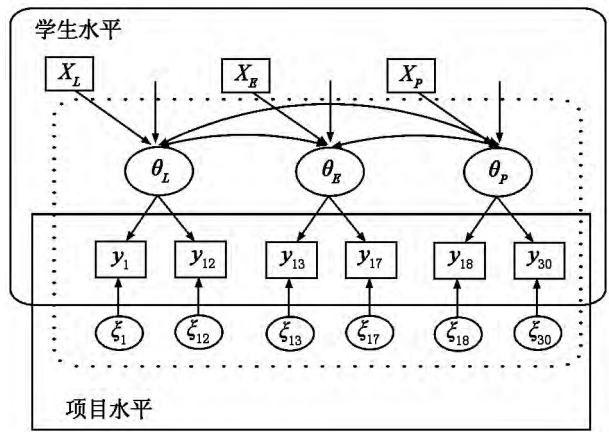


图 1 ML-MIRT 模型的路径图

采用 OpenBUGS 3.23 软件对模型中的各参数进行估计,为保证模型可收敛以及各参数估计值的 MC 误差(MC error)在其标准差的 5% 以下<sup>[18]</sup>,模型一将舍弃前 5 000 迭代次数,模型二将舍弃前

10 000 迭代次数,之后两个模型各自迭代 10 000 次.

2.5 结果

2.5.1 项目参数 表 2 和表 3 分别呈现了 30 个试题的区分度参数和截距参数的相关统计量.表 2 显示 12 个生物题的区分度范围为 0.28 ~ 1.08,其中项目 3 和 7 的区分度相对较小;5 个地理试题的区分度为 0.45 ~ 0.88,有 3 个试题的区分度在 0.5 以下;剩余的 13 个物理试题的区分度范围在 0.47 和 1.75 之间.该测验大部分项目的区分度在 0.50 以上,只有 7 个项目在 0.50 以下,占总测验的 23.3%,其中第 10 和第 17 题非常接近 0.50,说明该测验具有较好的区分能力.同时,项目区分度的 SD 范围为 0.10 ~ 0.40,SD 相对较小,说明区分度估计较为稳定.

表 2 各试题区分度参数估计值

参数	估计值	SD	MC 误差	置信区间	
				2.50%	97.50%
$a_1$	1.002	0.261	0.021	0.573	1.597
$a_2$	0.565	0.138	0.011	0.317	0.859
$a_3$	0.386	0.103	0.007	0.212	0.615
$a_4$	1.016	0.210	0.018	0.644	1.451
$a_5$	0.637	0.145	0.012	0.380	0.946
$a_6$	0.909	0.204	0.017	0.557	1.335
$a_7$	0.279	0.094	0.006	0.123	0.490
$a_8$	1.047	0.247	0.020	0.621	1.576
$a_9$	1.075	0.223	0.019	0.660	1.528
$a_{10}$	0.496	0.123	0.009	0.284	0.763
$a_{11}$	0.679	0.150	0.013	0.415	0.997
$a_{12}$	0.833	0.180	0.015	0.523	1.213
$a_{13}$	0.445	0.085	0.007	0.300	0.629
$a_{14}$	0.468	0.090	0.008	0.311	0.652
$a_{15}$	0.877	0.155	0.014	0.596	1.182
$a_{16}$	0.866	0.177	0.015	0.551	1.220
$a_{17}$	0.496	0.101	0.009	0.323	0.705
$a_{18}$	1.138	0.248	0.023	0.783	1.653
$a_{19}$	0.467	0.119	0.010	0.289	0.732
$a_{20}$	1.436	0.341	0.031	0.968	2.191
$a_{21}$	1.124	0.257	0.024	0.768	1.674
$a_{22}$	0.556	0.133	0.011	0.355	0.846
$a_{23}$	0.953	0.233	0.021	0.621	1.475
$a_{24}$	0.515	0.137	0.011	0.308	0.821
$a_{25}$	1.204	0.275	0.025	0.823	1.803
$a_{26}$	0.985	0.218	0.020	0.667	1.441
$a_{27}$	0.882	0.206	0.018	0.590	1.330
$a_{28}$	1.754	0.407	0.037	1.181	2.628
$a_{29}$	0.788	0.189	0.017	0.519	1.203
$a_{30}$	0.853	0.196	0.018	0.578	1.287

表 3 呈现了各项目的截距参数及其所对应的难度值.生物、地理和物理的难度( $b$ )及其 SD 范围分

别为  $-2.8 \sim 3.16$  ( $0.13 \sim 0.35$ )、 $-0.89 \sim 3.08$  ( $0.26 \sim 0.45$ ) 和  $-2.34 \sim 2.18$  ( $0.17 \sim 0.57$ )。测验中各项的难度基本贯穿了整个能力量尺,其中难度在  $-1$  以下和  $1$  以上的项目各有 5 个,其它 20 个项目难度在  $-1$  到  $1$  之间,通过对这 30 个项目难度进行 Shapiro-Wilk 检验( $W$  检验)发现,这些项目服从均值为  $-0.033$ ,标准差为  $1.35$  的正态分布( $W = 0.96$ ,  $p > 0.05$ )。由此可见,测验难易程度的设定是相对合理的。

表 3 各试题截距参数估计值及相应的难度值

参数	估计值	SD	MC 误差	置信区间		难度 ( $b$ )
				2.50%	97.50%	
$d_1$	1.843	0.340	0.031	1.121	2.484	-1.839
$d_2$	0.108	0.201	0.018	-0.293	0.476	-0.192
$d_3$	-0.387	0.159	0.013	-0.750	-0.094	1.002
$d_4$	0.107	0.327	0.031	-0.528	0.735	-0.105
$d_5$	0.303	0.220	0.020	-0.153	0.709	-0.475
$d_6$	0.835	0.301	0.028	0.230	1.393	-0.918
$d_7$	0.788	0.132	0.009	0.502	1.019	-2.822
$d_8$	1.435	0.340	0.032	0.736	2.068	-1.371
$d_9$	0.146	0.347	0.033	-0.601	0.792	-0.136
$d_{10}$	-0.004	0.179	0.015	-0.379	0.323	0.007
$d_{11}$	-0.572	0.237	0.022	-1.027	-0.138	0.843
$d_{12}$	-2.633	0.304	0.025	-3.305	-2.086	3.160
$d_{13}$	-1.372	0.257	0.023	-1.868	-0.904	3.082
$d_{14}$	-0.375	0.260	0.024	-0.823	0.078	0.802
$d_{15}$	-0.497	0.447	0.044	-1.207	0.302	0.567
$d_{16}$	0.759	0.444	0.043	0.056	1.500	-0.877
$d_{17}$	0.277	0.281	0.026	-0.232	0.778	-0.559
$d_{18}$	-0.385	0.371	0.036	-1.086	0.382	0.338
$d_{19}$	-1.020	0.174	0.015	-1.380	-0.687	2.183
$d_{20}$	0.780	0.459	0.045	-0.137	1.749	-0.543
$d_{21}$	-1.005	0.372	0.036	-1.747	-0.252	0.894
$d_{22}$	-0.137	0.192	0.018	-0.509	0.256	0.246
$d_{23}$	1.191	0.316	0.030	0.565	1.871	-1.250
$d_{24}$	1.204	0.186	0.016	0.842	1.590	-2.338
$d_{25}$	0.474	0.390	0.038	-0.288	1.311	-0.393
$d_{26}$	-1.108	0.320	0.030	-1.703	-0.434	1.125
$d_{27}$	0.132	0.290	0.028	-0.420	0.751	-0.149
$d_{28}$	1.435	0.567	0.055	0.375	2.589	-0.818
$d_{29}$	0.468	0.264	0.025	-0.067	0.990	-0.594
$d_{30}$	-0.124	0.285	0.027	-0.703	0.467	0.146

2.5.2 零模型下的学生参数 表 4 呈现了在零模型条件下得到的学生参数结果。由表 4 可知,学生的地理平均能力( $\gamma_{02}$ )与其它 2 种能力之间存在较大差异,地理平均能力要高于其它 2 种能力;而生物能力( $\gamma_{01}$ )和物理平均能力( $\gamma_{03}$ )间差异相对较小。同时,3 种能力之间的相关性非常高,任意 2 个能力间均达到高相关程度( $r > 0.8$ )。另外,学生在生物能力量尺中的分布范围( $\tau_{01}$ )比另外 2 个能力分布范围更窄,说明学生的生物能力较为集中;学生的物理和生物的平均能力以及分布范围差异不大;而地理

能力则在不同学生之间存在着比较大的差异,学生在此能力上的分布范围( $\tau_{02}$ )明显比生物和物理 2 种能力的分布更为广泛。

表 4 零模型时学生参数描述统计量

参数	估计值	SD	MC 误差	置信区间		显著性 检验
				2.50%	97.5%	
固定效应						<i>t</i> 值
$\gamma_{01}$	0.822	0.198	0.019	0.393	1.135	19.286***
$\gamma_{02}$	1.807	0.335	0.033	1.233	2.377	22.699***
$\gamma_{03}$	0.304	0.286	0.028	-0.402	0.696	6.195***
随机效应						卡方
$\tau_{01}$	1.437	0.586	0.058	0.682	2.985	592.570
$\tau_{02}$	5.013	1.760	0.173	2.725	8.847	620.178
$\tau_{03}$	1.905	0.702	0.069	0.842	3.100	631.880
相关系数						
$r_{12}$	0.935	0.023	0.002	0.884	0.969	
$r_{13}$	0.891	0.025	0.002	0.837	0.935	
$r_{23}$	0.935	0.022	0.002	0.890	0.974	

注:参数中的下标 1、2、3 分别表示生物、地理和物理 3 种能力,\*\*\*表示  $p < 0.001$ ,下同。

为计算学生间的差异对各能力维度总体变异的贡献量,可通过计算学生的随机变异在各能力上的跨级相关系数(intraclass correlation coefficient, ICC)来实现,其公式为

$$ICC = \tau_{0k} / (\tau_{0k} + \sigma^2). \quad (4)$$

值得注意的是,ML-MIRT 中的第 1 水平公式没有误差项,因为有研究指出对于 0-1 计分的多水平 Logistic 模型,其方程式的误差项方差已经被限制,同时根据 Logistic 分布推导出的第 1 水平误差项方差为  $\pi^2/3$ ,约为 3.29<sup>[19]</sup>。所以本研究中的  $\sigma^2$  即为 3.29,将其代入(4)式得到各能力的 ICC 分别为 0.304、0.604、0.367,说明各被试间的变异可分别解释结果变异的 30.4%、60.4% 和 36.7%。与此同时,邱皓政<sup>[20]</sup>基于 Cohen 的标准指出,当  $ICC \geq 0.138$  时,测验结果在不同被试间存在较大的差异,需要使用多水平模型处理这些数据。

2.5.3 含预测变量条件下的学生参数 为考察学生的 2 个预测变量对学生 3 个学业能力的影响,保持层 1 水平不变,在层 2 水平中加入父亲和母亲的受教育程度。经过分析发现,该条件下的项目参数和零模型条件下的项目参数差异不大,故将不再呈现项目参数的估计结果,而该条件下的学生参数结果如表 5 所示。由表 5 可知,加入 2 个预测变量后,学生的生物和物理的平均能力( $\gamma_{01}$  和  $\gamma_{02}$ )较没有加入预测变量时有显著性变化,加入预测变量后,学生在这 2 种能力维度上的平均能力差异变得不显著( $t = 1.486$ ,  $p > 0.05$  和  $t = -1.463$ ,  $p > 0.05$ )。

同时,表中结果显示父亲的受教育程度对学生

的学业能力有显著影响,父亲的受教育程度越高,则学生的生物能力( $\gamma_{11} = 0.309$   $p < 0.001$ )、地理能力( $\gamma_{12} = 0.504$   $p < 0.001$ )和物理能力( $\gamma_{13} = 0.343$   $p < 0.001$ )越好。而母亲的受教育程度对学生的学业能力的作用不显著。另外,从 3 个能力维度的随机部分的估计结果可以看出,在加入父母的受教育程度这 2 个预测变量之后,学生的生物能力的

随机变异由 1.437 下降到 1.372,地理能力的随机变异则由原来的 5.013 下降到 3.724,而学生的物理能力的随机变异由 1.905 下降到 1.795,此二者对 3 个能力维度的解释量分别为 4.52%、25.71% 和 5.77%,即这 2 个预测变量可以解释学生生物能力差异的 4.52%、地理能力差异的 25.71%,以及可以解释物理能力差异的 5.77%。

表 5 含预测变量的模型学生参数统计量

参数	估计值	SD	MC 误差	置信区间		显著性 检验
				2.50%	97.5%	
固定效应						<i>t</i> 值
$\gamma_{01}$	0.209	0.170	0.016	-0.195	0.504	1.486
$\gamma_{02}$	0.867	0.457	0.045	0.170	1.793	3.743***
$\gamma_{03}$	-0.313	0.182	0.018	-0.877	-0.071	-1.463
回归系数						
$\gamma_{11}$	0.309	0.068	0.006	0.184	0.433	6.409***
$\gamma_{21}$	-0.031	0.065	0.006	-0.163	0.097	-0.658
$\gamma_{12}$	0.504	0.161	0.016	0.240	0.798	6.345***
$\gamma_{22}$	-0.103	0.115	0.010	-0.358	0.105	-1.327
$\gamma_{13}$	0.343	0.103	0.010	0.224	0.631	6.220***
$\gamma_{23}$	0.098	0.063	0.006	-0.011	0.243	1.818
随机效应						卡方
$\tau_{01}$	1.372	0.283	0.026	0.942	2.001	602.564
$\tau_{02}$	3.724	1.662	0.164	1.161	7.330	595.651
$\tau_{03}$	1.795	0.705	0.069	0.950	4.010	627.005
相关系数						
$r_{12}$	0.934	0.022	0.002	0.886	0.970	
$r_{13}$	0.887	0.026	0.002	0.831	0.935	
$r_{23}$	0.927	0.022	0.002	0.878	0.962	

3 讨论与结论

本研究将多水平单维 IRT 模型拓展到多水平多维 IRT 模型以吻合测验发展的需要,并通过一个实证研究对 ML-MIRT 模型的适用情况进行了较为深入地探讨。结果表明,该测验项目的设定相对较为合理,只有少数几个项目的区分度在 0.50 以下,同时项目难度基本上涵盖了整个能力量尺,基本上实现了对各能力水平学生的测量。

为探讨可能影响学生学业能力的因素,研究将父亲和母亲的受教育状况等 2 个预测变量加入到模型中。结果显示,父亲的受教育程度可以显著地预测学生的学生能力,父亲受教育程度越高,则学生的学业能力也越强;而母亲的受教育程度则对学生的学业能力没有显著地预测作用。该结果与 J. Ermisch 等<sup>[21]</sup>的结果相同,但与 C. J. Crook 以及庞维国等的研究结果不一致<sup>[22-23]</sup>。笔者认为本研究结果与庞维

国等的结果不一致的原因可能有以下几点:(i)施测被试上的差异。庞维国等的施测对象是省会城市的学生,其母亲受教育程度的异质性较高;而本研究中的学生有 60% 以上属于农村居民,且母亲在 4 个受教育程度上的分布不均衡,由表 1 可知,80% 以上的母亲受教育程度在初中及以下学历,导致从母亲受教育程度来对被试进行分类,样本的异质性程度不高,从而使得研究结果无差异。(ii)在农村地区,一般是父亲的受教育程度较母亲高,对学生学习介入水平也高,并且父亲的受教育程度决定了家庭的整体经济水平,而家庭的整体经济水平会影响着家庭对子女教育的投入水平,从而也会间接影响着学生的学业成绩。(iii)以往研究采用一般线性模型,没有考虑被试在一些情境变量上的异质性,本研究采用考虑情境变量的多水平模型对数据进行分析,也有可能会导致与以往结果不一致的地方。关于父母教育程度对学生学业影响的实质和内部机制,笔者的后续研究拟加大样本在父母受教育程度上的异质性

程度,通过构建中介和调节效应的理论模型,采用多水平多维度 IRT 方法对其进行深入探讨.同时,图 2 结果表明,加入学生预测变量后,模型的固定效应和随机效应的值均要小于零模型条件下的各效应值,进一步说明预测变量对测验结果的解释有较大的帮助.值得注意的是,不管是零模型还是含预测变量的模型,其中关于地理能力的随机变异( $\tau_{02}$ )比较大,并且其估计值的标准差均较大,可能的原因是测验中测量学生地理能力的试题只有 5 个,而较少的试题量使得在对被试的能力的估计时可能受额外因素的影响,从而导致其估计值及标准差均较大.

目前,本研究只探讨了两水平的 MIRT 模型,并

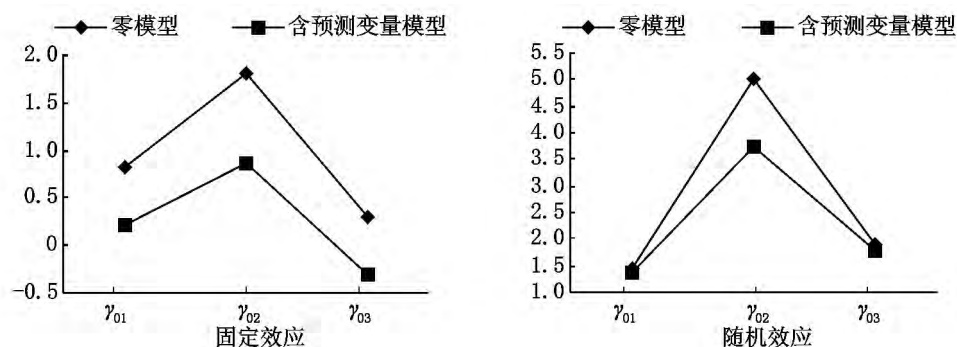


图 2 2 种模型下的固定效应和随机效应值

## 4 参考文献

- [1] 刘慧,简小珠,张敏强,等.多水平 IRT 的发展与应用述评[J].心理科学进展,2012,20(4):627-632.
- [2] 刘红云,骆方.多水平项目反应理论模型在测验发展中的应用[J].心理学报,2008,40(1):92-100.
- [3] Bacci S, Caviezel V. Multilevel IRT models for the university teaching evaluation [J]. Journal of Applied Statistics, 2011, 38(12): 2775-2791.
- [4] Fox J P, Glas C A. Bayesian estimation of a multilevel IRT model using Gibbs sampling [J]. Psychometrika, 2001, 66(2): 271-288.
- [5] Fox J P. Applications of multilevel IRT modeling [J]. School Effectiveness and School Improvement, 2004, 15(3/4): 261-280.
- [6] Hox J. Multilevel analysis: Techniques and applications [M]. New York: Routledge Press, 2010.
- [7] 康春花,辛涛.测验理论的新发展:多维项目反应理论[J].心理科学进展,2010(3):530-536.
- [8] Hühler J, Hartig J, Goldhammer F. Modeling the multidimensional structure of students' foreign language competence within and between classrooms [J]. Psychological Test and Assessment Modeling, 2010, 52(3): 323-340.
- [9] Hartig J, Hühler J. Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality [J]. Zeitschrift für Psychologie/ Journal of Psychology, 2008, 216(2): 89-101.
- [10] De Jong M G, Steenkamp J B E. Finite mixture multilevel multidimensional ordinal IRT models for large scale cross-cultural research [J]. Psychometrika, 2010, 75(1): 3-32.
- [11] Lu Y, Bolt D M. Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style [J]. Large-scale Assessments in Education, 2015, 3(1): 1-18.
- [12] Stout W. Skills diagnosis using IRT-based continuous latent trait models [J]. Journal of Educational Measurement, 2007, 44(4): 313-324.
- [13] 涂冬波,蔡艳,戴海琦,等.多维项目反应理论:参数估计及其在心理测验中的应用[J].心理学报,2011,43(11):1329-1340.
- [14] Lahuis D M, Avis J M. Using multilevel random coefficient modeling to investigate rater effects in performance ratings [J]. Organizational Research Methods, 2007, 10(1): 97-107.
- [15] Wang Wenchung, Qiu Xuelan. A multidimensional and multilevel extension of a random-effect approach to subjective judgment in rating scales [J]. Multivariate Behavioral Research, 2013, 48(3): 398-427.

- [16] Melby J N ,Conger R D ,Fang S A ,et al. Adolescent family experiences and educational attainment during early adulthood [J]. *Developmental Psychology* 2008 ,44( 6) : 1519.
- [17] Fox J P. Bayesian item response modeling: theory and applications [M]. Berlin: Springer Science & Business Media 2010.
- [18] Ntzoufras I. Bayesian modeling using WinBUGS [M]. Manhattan: John Wiley & Sons 2011.
- [19] 温福星. 阶层线性模型的原理与应用 [M]. 北京: 中国轻工业出版社 2009.
- [20] 邱皓政. 量化研究法(二): 统计原理与分析技术 [M]. 台北: 双叶书廊图书公司 2005.
- [21] Ermisch J ,Pronzato C. Causal effects of parents' education on children's education [R]. ISER Working Paper Series , 2010.
- [22] Crook C J. The role of mothers in the educational and status attainment of Australian men and women [J]. *Journal of Sociology* ,1995 ,31( 2) : 45-73.
- [23] 庞维国 徐晓波 林立甲 等. 家庭社会经济地位与中学生学业成绩的关系研究 [J]. *全球教育展望* ,2013 ( 2) : 12-21.
- [24] 康春花 孙小坚 曾平飞. 基于等级反应模型的多水平多侧面评分者模型 [J]. *心理科学* 2016( 1) : 214-223.

## The Application of Multilevel Multidimensional IRT Model in Academic Quality Monitoring Test

KANG Chunhua ,SUN Xiaojian ,GU Shiwei ,ZENG Pingfei\*

( College of Teacher Education ,Zhejiang Normal University ,Jinhua Zhejiang 321004 ,China)

**Abstract:** Analysis the academic achievement of students with multilevel multidimensional IRT ( ML-MIRT) model , and expand the application field of the model. First ,use the zero model ,which does not contain any predictor ,to analysis the item and person parameters; then use the ML-MIRT model that contains two predictors about students to explore the impact of the predictors on students' academic achievement. Difficulty of these items and the ability scale is matching ,and most of items have good discrimination ,the intraclass correlation coefficients ( ICC) of the three abilities are significant ,the education background of father affects the three abilities of students significantly , while the education background of mother doesn't affect them. The test is reasonable ,and the ML-MIRT model is appropriate ,also and the education background of father plays an important role for children's academic achievement.

**Key words:** multilevel multidimensional IRT model; academic achievement; quality monitoring; education background of parents

( 责任编辑: 冉小晓)