

图形推理任务作答规则的多策略诊断

刘铁川¹, 赵玉², 戴海琦³

(1.赣南师范大学教育科学学院, 赣州 341000; 2.赣南医学院
心理学系, 赣州 341000; 3.江西师范大学心理学院, 南昌 330022)

【摘要】 目的:使用多策略测量模型从新的视角诊断图形推理任务的规则掌握情况。**方法:**根据言语分析或视觉空间两种策略对瑞文高级推理测验题目所使用的规则进行编码,从题目参数、模型拟合、内部效度等多个角度比较了基于单策略假设的DINA模型和基于多策略假设的Mix-DINA和MS-DINA模型的诊断结果。**结果:**Mix-DINA模型对瑞文高级推理测验的拟合程度、题目区分度、内部效度最高;DINA模型最差;MS-DINA模型很少考虑多策略过程,高估了部分规则难度。**结论:**使用Mix-DINA模型分析图形推理测验更合理,不仅能够诊断个体的规则掌握状态,还可估计其策略使用倾向。

【关键词】 图形推理; 解题策略; 认知诊断; Mix-DINA模型

中图分类号: R395.1

DOI: 10.16128/j.cnki.1005-3611.2016.03.017

Multiple-Strategy Diagnosis of Rules for Solving Figure Reasoning Tasks

LIU Tie-chuan¹, ZHAO Yu², DAI Hai-qi³

¹School of Educational Science, Gannan Normal University, Ganzhou 341000, China; ²Department of Psychology, Gannan Medical College, Ganzhou 341000, China; ³School of Psychology, Jiangxi Normal University, Nanchang 330022, China

【Abstract】 Objective: To diagnose rules used in solving figure reasoning tasks by utilizing new multiple strategy psychometric models. **Methods:** Items in Raven Advanced Progressive Matrices(RAPM) were coded from the perspective of Verbal-Analytic strategy and Visual-Spatial strategy, and different cognitive process were involved in correspond strategies. Three psychometric models were used to fit RAPM. The first model was DINA model which assumed only one strategy. The other two models were Mix-DINA and MS-DINA, which allowed multiple strategies. Item parameters, goodness of fit and internal validity were compared between these models. **Results:** Mix-DINA model fit RAPM best and provided the highest items' discriminable and internal validity. MS-DINA model rarely diagnose subjects' rules from multiple-strategy perspective and overestimate some rules' difficulty. **Conclusion:** Our data fit RAPM with Mix-DINA model best, which provides both individual's rule mastery state and strategy use tendency.

【Key words】 Figure reasoning; Problem solving strategy; Cognitive diagnosis; Mix-DINA model

智力评估、认知发展障碍研究中广泛使用瑞文推理测验获得个体在一般智力因素上的个体差异^[1]。²¹。Carpenter等人的早期研究认为完成此类图形推理任务必须掌握行不变、数量成对变化、图形加减、三值分布、二值分布等五种规则^[3]。此结果在对图形推理任务解题过程研究中广泛引用^[4]。

然而,研究者们发现完成图形推理任务时个体可能使用不同的解题策略。依据双重编码理论,De-shon等人认为图形推理任务有言语分析(Verbal Analytic)和视觉空间(Visual-Spatial)两种策略,其中前者形成刺激的命题表征,而后者形成刺激的视觉表征,他们进一步通过语词遮蔽效应范式实验证实了其假设^[5]。林崇德等人在总结前人研究基础上,认为

图形推理包括分析、不完全分析、知觉分析、知觉匹配、自主想象、格式塔等六种策略^[6]。

不同策略使用者的内部心理过程存在区别,使图形推理认知过程研究更为复杂。使用言语分析策略涉及的是行不变、数量成对变化、三值分布、二值分布四种规则,而使用视觉空间策略涉及的是重叠、抵消重叠、图形加减、移动、旋转、心理变换六种规则^[5]。然而,目前图形推理任务的测量研究均基于Carpenter等人的规则分类^[7-9],忽略了图形推理任务存在多种解题途径的可能性。DINA模型(Deterministic Inputs, Noisy "And" gate model)是目前研究中使用最多的单策略认知诊断模型之一^[10],模型中被试*i*解答题目*j*的正确概率为 $P(X_{ij}=1)=(1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}$ 。其中, g_j 和 s_j 为题目*j*的猜测参数和失误参数, η_{ij} 为被试是否掌握解决题目*j*所需的全部规则。MS-DINA模型(Multiple-Strategy DINA model)^[11]将DINA模型拓展至多策略领域,其做法是取*M*个策略对应的潜在反

【基金项目】 本研究得到了江西省教育科学规划重点课题(13ZD21.005)和江西省高校人文社会科学研究青年基金项目(XL1302)资助

通讯作者:戴海琦, E-mail: daihaiqi@aliyun.com

应变量最大值作为被试*i*对题目*j*最终潜在潜在反应,即 $\eta_{ij}=\max(\eta_{ij1},\eta_{ij2},\dots,\eta_{ijM})$ 。MS-DINA模型的局限是无法同时报告个体的策略使用倾向和规则掌握状态。

近年来新提出的Mix-DINA模型^[12]可解决上述问题。此模型中被试*i*解答题目*j*的正确概率为 $P(X_{ij}=1)=\sum_{m=1}^M \pi_m g_{jm}^{1-\eta_{ijm}} (1-s_{jm})^{\eta_{ijm}}$ 。其中, π_m 为使用第*m*种策略被试的比例,且 $\sum_{m=1}^M \pi_m=1$; g_{jm} 和 s_{jm} 为题目*j*对使用第*m*种策略者的猜测参数和失误参数; η_{ijm} 表示被试*i*是否掌握使用策略*m*解决题目*j*所需的全部规则。

针对以往单策略作答假设的不足,研究认为图形推理任务存在口语分析和视觉空间两种策略,使用较新的Mix-DINA模型同时分析被试完成图形推理任务时的策略使用倾向和规则掌握状态,并比较其与MS-DINA模型用于多策略测验分析的优势。

1 对象与方法

1.1 作答数据

在南昌市的几所高校使用瑞文高级推理测验测试了628名被试,作答过程无时间限制,以保证被试有充裕时间完成整个测验。剔除无效答卷后,共回收613份完整作答数据。评分后,最高31分,最低2分,平均23.1分。瑞文高级推理测验alpha系数为0.859,由自助法(Bootstrap)得到1000个样本的95%置信区间为(0.834, 0.880),显示研究数据中测验内部一致性良好。

1.2 瑞文高级推理测验的多策略Q矩阵

根据Carpenter等人^[3]和Deshon等人^[5]的分类,对瑞文高级推理测验涉及的规则进行编码。其中,第15题无法根据现有规则进行编码,只有第18题测的是心理变换规则。此外,Deshon等人将第19、20、25题划分为同时需要两种策略的规则才能解决的题目,但这些题目无法根据现有规则系统编码,所以无法确定这三题使用的分析性规则。因此,研究对瑞文高级推理测验的分析不包括第15、18、19、20、25这五个题目。

此外,Q矩阵的界定过程中对一些规则进行了合并,以增加测量次数。由于抵消重叠实际上是重叠规则的特例,只有第11题测了图形加减规则,因此将抵消重叠、重叠其与图形加减规则合并,合并后称为A/S;第26题的正确解决既需要三值分布,也需要进行旋转,但只有此题测量了旋转规则,因此将旋转合并到移动规则。

最终编码结果见表1。分析策略倾向的C、PP、

D3、D2、M、A/S六种规则测量次数分别为13、7、12、7、4、7次,视觉空间策略倾向的分别为8、1、11、4、11、9次。

表1 瑞文高级推理测验认知诊断多策略Q矩阵

题号	分析策略倾向(QA)	视觉空间策略倾向(QV)	题号	分析策略倾向(QA)	视觉空间策略倾向(QV)
1	C, D3	C, D3	21	C, D3	C, D3
2	C, PP	M	22	D2	A/S
3	PP	M	23	D2	A/S
4	C, PP	C, PP	24	C, M	C, M
5	C, PP	M	26	D3, M	D3, M
6	C, PP	M	27	D3	D3
7	A/S	A/S	28	D3	D3
8	D3	D3	29	D3	D3
9	A/S	A/S	30	D2	D2
10	C, PP	M	31	D3, D2, M	D3, D2, M
11	A/S	A/S	32	D3, D2	M
12	A/S	A/S	33	M, A/S	M, A/S
13	C, D3	C, D3	34	D3	D3
14	C, PP	M	35	C, D2, A/S	C, D2, A/S
16	A/S	A/S	36	C, D2	C, D2
17	C, D3	C, D3			

注:C为行不变:每一行的值不变,但每一列的值却会改变。PP为数量成对变化:在相邻的两个单元之间,大小、位置、数量等发生增减变化。D3为三值分布:每一行、每一列中的图形单元或其属性有三种值。D2为二值分布:每一行、每一列中的图形单元只有两种值,不存在第三种值。M为移动:感觉上某图形单元像是连续朝某特定方向移动。A/S为图形加减:两个图形单元以视觉的方式结合成一个整体,但并不重叠,只是共用某边缘。

1.3 统计分析方法

研究将使用DINA、Mix-DINA、MS-DINA等三种认知诊断模型对瑞文高级推理测验的作答数据进行分析。其中,DINA模型基于单策略作答假设,使用的是表1中QA部分;Mix-DINA和MS-DINA模型基于多策略作答假设,使用的是表1中QA和QV两部分。三个模型的参数估计均使用R语言下的自编程序,通过MCMC算法估计模型参数。采用5条链,各链长度均设为8000,取后4000次的均值作为模型参数估计值。

比较各诊断模型的估计结果时,对于能力参数、题目参数等连续变量,通过积差相关分析差异大小;规则诊断结果为离散变量,则计算Cohen's Kappa统计量来评价单个规则与规则掌握模式的一致性。

各模型的整体拟合程度的比较将使用信息准则与各模型的预测值与原始作答的平均绝对残差。各诊断模型为非嵌套模型,故使用AIC、AICc和BIC和DIC等信息准则比较各诊断模型的整体拟合程度。一般认为,BIC加大了对模型复杂程度的惩罚,更偏好简洁模型^[13]。在样本量较小时,使用AIC有可能

选择过拟合(over-fitting)的模型, AICc 对样本量较小的情形进行了校正, Bumham 和 Anderson 建议在什么时候都应时候 AICc^[14], 模拟研究显示, AIC、AICc 的实际表现要好于 BIC。DIC 可以看作是 AIC 指标在贝叶斯统计中的扩展, 其对模型的比较考虑了先验信息。各种信息准则越小的模型对数据拟合越好。

模型内部效度(Internal Validity)比较使用的是项目掌握统计量(Item Mastery Statistics, IMstats)^[15], Imstats 指掌握了某个题目所测全部规则的被试(masters)与未掌握全部规则的被试(non-masters)中正确解答该题的人数比例上的差异。Imstats 越高, 表示模型内部效度越理想^[16]。

2 结 果

2.1 模型拟合程度比较

研究计算了多种信息准则(表2)。其中, Mix-DINA 模型的 AIC、AICc、DIC 等指标最小, 其次是 MS-DINA 模型, DINA 模型的 AIC 最大; 与其他指标相比, BIC 指标加大了对复杂模型的惩罚, 所以与 DINA 模型估计相同数量题目参数的 MS-DINA 模型的 BIC 最小, 而 Mix-DINA 模型的 BIC 指标最大。结合平均绝对残差(越小表示拟合越理想)来看, DINA 模型拟合最差, MS-DINA 模型居中, Mix-DINA 模型最好。

表2 瑞文高级推理测验各模型的拟合统计量

分析指标	DINA	Mix-DINA	MS-DINA
		信息准则	
AIC	14622.972	14322.145	14532.904
AICc	14637.175	14347.162	14547.107
BIC	16034.020	16165.611	15943.951
DIC	15009.728	14696.573	14915.542
平均绝对残差	0.245	0.239	0.243

2.2 不同模型下题目参数比较

本研究使用了瑞文高级推理测验的31个题目, 表3汇总了三个模型下的题目参数估计结果。可以通过题目参数的相关程度和平均绝对差异分析不同模型估计结果的一致性。DINA 与 MS-DINA 两模型得到的失误和猜测参数相关较高, 分别为 0.979、0.877, 而平均绝对差异较小, 分别为 0.017、0.040。但这种差异存在系统性规律, MS-DINA 模型中 87% 题目的 s 参数、32% 题目的 g 参数小于 DINA 模型所得结果。

Mix-DINA 模型分析策略对应的 s_1 、 g_1 参数与 DINA 和 MS-DINA 模型所得结果较为相似, s 参数的相

关分别为 0.949、0.963, g 参数的分别为 0.908、0.810, 它们的平均绝对差异也较小(0.023-0.052)。但视觉空间策略对应的 s_2 、 g_2 参数与 DINA 和 MS-DINA 模型所得结果差异较大, s 参数的平均绝对差异分别为 0.044、0.061, g 参数的分别为 0.147、0.100。而且视觉空间策略倾向的多数题目的 s_2 、 g_2 参数小于 DINA 和 MS-DINA 模型结果(78%-89%)。

表3 瑞文高级推理测验各模型下的题目参数

题目	DINA		Mix-DINA		MS-DINA	
	1-s	g	1-s ₁	g ₁	1-s ₂	g ₂
Item1	0.990	0.587	0.990	0.586		0.990 0.586
Item2	0.955	0.365	0.923	0.430	0.979 0.215	0.956 0.372
Item3	0.977	0.096	0.965	0.177	0.982 0.118	0.977 0.097
Item4	0.946	0.464	0.956	0.363		0.948 0.473
Item5	0.930	0.290	0.890	0.262	0.966 0.286	0.932 0.298
Item6	0.977	0.383	0.968	0.432	0.979 0.234	0.978 0.396
Item7	0.950	0.368	0.950	0.390		0.952 0.424
Item8	0.952	0.530	0.952	0.532		0.954 0.529
Item9	0.989	0.338	0.989	0.369		0.990 0.420
Item10	0.900	0.381	0.867	0.401	0.932 0.284	0.900 0.399
Item11	0.968	0.320	0.969	0.333		0.969 0.394
Item12	0.978	0.455	0.979	0.464		0.979 0.495
Item13	0.836	0.551	0.837	0.552		0.837 0.551
Item14	0.949	0.298	0.937	0.381	0.955 0.164	0.950 0.305
Item16	0.947	0.174	0.951	0.156		0.957 0.174
Item17	0.879	0.465	0.886	0.452		0.884 0.457
Item21	0.861	0.340	0.862	0.348		0.859 0.351
Item22	0.914	0.536	0.845	0.463	0.955 0.198	0.825 0.324
Item23	0.896	0.570	0.778	0.528	0.973 0.164	0.843 0.299
Item24	0.769	0.409	0.712	0.138		0.780 0.488
Item26	0.551	0.365	0.525	0.350		0.560 0.399
Item27	0.640	0.133	0.641	0.132		0.642 0.130
Item28	0.557	0.073	0.558	0.074		0.557 0.077
Item29	0.492	0.084	0.492	0.083		0.494 0.077
Item30	0.782	0.200	0.871	0.229		0.869 0.253
Item31	0.861	0.395	0.889	0.410		0.932 0.438
Item32	0.765	0.160	0.905	0.100	0.596 0.181	0.785 0.153
Item33	0.697	0.203	0.573	0.188		0.735 0.291
Item34	0.686	0.277	0.690	0.263		0.690 0.269
Item35	0.798	0.282	0.833	0.360		0.847 0.360
Item36	0.485	0.127	0.544	0.134		0.554 0.141

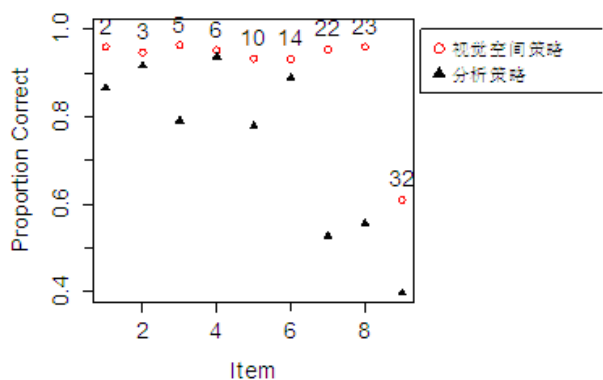
认知诊断模型的 1-s-g 值反映了题目的区分能力(discrimination)。DINA 模型和 MS-DINA 的平均区分能力为 0.505 和 0.507, Mix-DINA 模型下, 全部题目对分析策略倾向被试的平均区分能力为 0.505, 而全部题目对视觉空间策略倾向被试的平均区分能力为 0.555, 若只考察 s_2 和 g_2 参数则平均区分能力为 0.719。综合来看, Mix-DINA 模型参数估计结果较为理想, MS-DINA 模型次之, DINA 模型较差。

分析瑞文高级推理测验的多策略 Q 矩阵中涉及了 6 种规则, 各模型下规则的区分度参数都比较高

(表4)。除规则M外,各模型下其他5个规则的难度顺序与现有研究结果一致^[3]。对于DINA和MS-DINA模型,规则M的难度与D2非常接近,而在Mix-DINA模型下,M的难度远低于D2。

表4 瑞文高级推理测验各模型下的规则参数

规则	DINA		Mix-DINA		MS-DINA	
	lamda0	lamda1	lamda0	lamda1	lamda0	lamda1
C	-2.279	2.027	-2.437	1.448	-2.226	1.894
PP	-1.858	2.615	-2.042	2.087	-1.832	2.524
D3	-0.823	2.469	-0.798	2.481	-0.815	2.258
D2	-0.353	2.366	-0.052	1.716	0.010	2.459
M	-0.438	2.490	-1.426	2.473	-0.061	2.522
A/S	-1.675	2.619	-1.626	2.468	-1.554	2.585



附图 Mix-DINA模型下两种策略在部分题目上的正答比例

2.3 模型内部效度比较

为分析各模型的内部效度,研究计算各模型的Imstats。各题目平均而言,Mix-DINA模型下掌握者与未掌握者正答比例差异最大(M=0.527),而MS-DINA模型(M=0.502)与DINA模型(M=0.501)相差较小。

理论上,对于既可以采用分析策略解决、也可以采用视觉空间策略解决的题目,采用后者解决时其难度更低。表2显示,大多数此类题目的视觉空间策略对应的题目参数都小于分析策略,符合理论预期。而且,倾向于视觉空间策略的个体,其在这一类题目上凭猜测解题的可能性更小,而正确作答比例应更高。由附图可以看出,倾向于使用视觉空间策略的被试在此类题目上的正确作答比例均高于倾向于使用分析策略的被试,平均高出0.172。

由表1可知,第22、23、32题采用分析策略解题时使用的是难度较高的规则D2或D3,而采用分析策略时使用的是难度低很多的规则M或A/S,所以正确作答上的差异也比较大。采用分析策略解决其他几个题目时,所涉及的规则C或PP难度虽然低于视觉空间策略对应的规则M,但使用分析策略时通

常需要同时掌握两个规则,所以不同策略下的正答比例虽存在差异,但相对较小。Mix-DINA模型分析结果与上述理论分析吻合较好。

2.4 基于多策略模型的诊断分析

由前面的分析可知,Mix-DINA与MS-DINA模型均优于基于单策略假设的DINA模型,所以下面只比较分析前两者的诊断结果。整体上,Mix-DINA与MS-DINA模型的诊断一致性并不高(58%),即两模型对42%的被试的规则使用诊断结果存在差异。两模型对各规则诊断的Kappa一致性分别为C(0.450)、PP(0.851)、D3(0.941)、D2(0.915)、M(0.202)、A/S(0.915)。

表5 Mix-DINA模型下的规则掌握模式

序号	规则掌握模式	总频次	总比例	倾向分析策略		倾向视觉空间策略	
				频次	比例	频次	比例
1	000000	5	0.8%			5	1.3%
2	010000	3	0.5%	3	1.3%		
3	010011	3	0.5%			3	0.8%
4	011011	1	0.2%	1	0.4%		
5	100000	13	2.1%	8	3.3%	5	1.3%
6	100001	2	0.3%	2	0.8%		
7	100010	1	0.2%	1	0.4%		
8	110000	12	2.0%	7	2.9%	5	1.3%
9	110001	18	2.9%	17	7.1%	1	0.3%
10	110010	7	1.1%	5	2.1%	2	0.5%
11	110011	53	8.6%	33	13.8%	20	5.3%
12	110111	9	1.5%	3	1.3%	6	1.6%
13	111011	161	26.3%	67	28.0%	94	25.1%
14	111110	1	0.2%			1	0.3%
15	111111	324	52.9%	92	38.5%	232	62.0%

表6 MS-DINA模型下的规则掌握模式

序号	规则掌握模式	频次	比例	序号	规则掌握模式	频次	比例
1	000000	8	1.3%	9	110101	1	0.2%
2	010000	5	0.8%	10	110111	2	0.3%
3	011001	1	0.2%	11	111000	1	0.2%
4	100000	17	2.8%	12	111001	131	21.4%
5	100001	3	0.5%	13	111011	31	5.1%
6	110000	16	2.6%	14	111101	20	3.3%
7	110001	77	12.6%	15	111111	297	48.5%
8	110011	3	0.5%				

根据Mix-DINA模型的分析结果,研究样本(N=613)中倾向于使用分析策略的被试239人(39%)。倾向于使用分析策略的被试能力均值为-0.254,倾向于使用视觉空间策略的被试能力均值为0.186,二者差异显著($t_{df=437.633}=-7.022, P<0.001$)。倾向于使用分析策略的被试中,各规则的掌握比例分别为C

(98.3%)、PP(95.4%)、D3(66.9%)、D2(39.8%)、M(84.5%)、A/S(90.0%);而倾向于使用视觉空间策略的被试中,各规则的掌握比例分别为C(97.9%)、PP(97.3%)、D3(87.4%)、D2(64.0%)、M(95.7%)、A/S(95.2%)。可见,除规则C之外,倾向于使用视觉空间策略的被试各规则上的掌握比例更高。

使用Mix-DINA模型分析瑞文高级推理测验可得到15种规则掌握模式(表5),其中1表示掌握,0表示未掌握,规则顺序与表1相同。可见,总人群中111111模式人数最多,且多倾向使用视觉空间策略(62%)。由表4可知规则D3和D2较难掌握,所以模式111011(26.3%)和110011(8.6%)也较多。

MS-DINA模型下得到了15种规则掌握模式(表6)。尽管模式111111数量与Mix-DINA模型结果相似,但Mix-DINA模型所得到的010011、011011、100010、110010、111110,MS-DINA模型下的011001、110101、111000、111001、111101等规则掌握模式在另一个模型诊断结果均未出现。可见,两多策略模型的诊断结果差异比较大。在MS-DINA模型下,占比例较大的错误模式是111001(21.4%)和110001(12.6%)两种,这与其高估了规则M的难度的情况是一致的(见表4)。

3 讨 论

研究构建了包括分析与视觉空间两种策略的瑞文高级推理测验的Q矩阵,数据分析显示,基于多策略假设的Mix-DINA拟合数据效果及内部效度最佳,MS-DINA次之,而基于单策略假设的DINA模型最差。研究结果表明,允许多种解题策略、基于混合分布的Mix-DINA模型在分析图形推理测验时更有优势,可同时报告个体的规则掌握状态与策略使用倾向,提供了非常有价值的诊断信息。

MS-DINA模型虽然允许各题目上的潜在反应变量不同,但事实上只有少数题目潜在变量上存在差异,其忽视了更多存在多种解题策略的测验题目,高估了规则M的难度。所以其拟合程度逊于Mix-DINA模型,但在部分题目上允许多种潜在变量,所以其拟合仍优于DINA模型。de la Torre和Douglas分析小学分数数据时,信息准则显示DINA模型的拟合程度优于MS-DINA模型^[17]。注意到他们研究中的所有测验题目均使用不同策略解决,而瑞文高级推理测验中存在使用相同认知过程解决的题目。考虑到Q矩阵对参数估计结果影响较大^[1],并且MS-DINA模型强制策略间题目参数不变,这可能是

导致我们的研究结果与其不同的原因。

工作记忆通过不同的解题策略影响个体完成图形推理任务的表现^[18],所以解题策略不同是个体差异研究不可忽视的一个方面。研究中使用的是瑞文高级推理测验,但结果可推广至其他存在多种解题策略的测验任务。基于混合分布的Mix-DINA以不同题目参数拟合多策略作答是可行的,通过解题策略分析可以更深入地了解个体的作答过程。同时,题目参数增加也需要较大的样本量,理论上,加大样本量会更有利于Mix-DINA模型。然而,研究中对三种诊断模型拟合程度仅进行了相对比较,使用绝对指标的比较结果是否与此一致需要进一步研究。多策略认知诊断的研究远未成熟,如何恰当地诊断策略使用、鉴别策略转换等复杂心理现象给研究方法带来了新的挑战。任务类型、策略数量和Q矩阵形态等因素对模型识别的影响,以及如何结合实验方法采集测验外部效度证据等,都是更准确地分析多策略认知作答过程亟待解决的新课题。

参 考 文 献

- 1 Vakil E, Lifshitz-Zehavi H. Solving the Raven Progressive Matrices by adults with intellectual disability with/without Down syndrome: Different cognitive patterns as indicated by eye-movements. *Research in Developmental Disabilities*, 2012, 33(2): 645-654
- 2 Carretti B, Belacchi C, Cornoldi C. Difficulties in working memory updating in individuals with intellectual disability. *Journal of Intellectual Disability Research*, 2010, 54(4): 337-345
- 3 Carpenter PA, Just MA, Shell P. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 1990, 97(3): 404-431
- 4 Loesche P, Wiley J, Hasselhorn M. How knowing the rules affects solving the Raven Advanced Progressive Matrices Test. *Intelligence*, 2015, 48: 58-75
- 5 DeShon RP, Chan D, Weissbein DA. Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence*, 1995, 21(2): 135-155
- 6 林崇德,沃建中,陈浩莺.小学生图形推理策略发展特点的研究. *心理科学*, 2003, 26(1): 2-8
- 7 陈德枝,戴海琦,周骏.同结构标准图形推理测验的编制与验证. *中国临床心理学杂志*, 2011, 19(3): 319-322
- 8 Primi R. Developing a fluid intelligence scale through a combination of Rasch modeling and cognitive psychology. *Psychological Assessment*, 2014, 26(3): 774-788

(下转第469页)

- tion and Emotion, 2002, 16: 127-141
- 9 Sundberg ND, Latkin CA, Farmer RF, et al. Boredom in young adults: Gender and cultural comparisons. *Journal of Cross-Cultural Psychology*, 1991, 22: 209-223
 - 10 Wang W, Wu Y, Peng Z, et al. Test of sensation seeking in a chinese sample. *Personality and Individual Differences*, 2000, 28: 169-179
 - 11 Mercer-Lynn K, Flora DB, Fahlman SA, et al. The measurement of boredom: Differences between existing self-report scales. *Assessment*, 2013, 20: 585-596
 - 12 Byrne BM, Shavetson RJ, Muthen B. Testing for the equivalence of factor covariance and mean structure issue: The issue of partial measurement invariance. *Psychological Bulletin*, 1989, 105: 456-466
 - 13 Byrne BM. *Structural Equation Modeling with Amos. Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum Associates, Publishers, 2001
 - 14 Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 1993, 58(4): 525-543
-
- (上接第463页)
- 9 Leighton JP, Chu M-W. First Among Equals: Hybridization of Cognitive Diagnostic Assessment and Evidence-Centered Game Design. *International Journal of Testing*, 2015. 1-17
 - 10 Wang C, Shu Z, Shang Z, et al. Assessing Item-Level Fit for the DINA Model. *Applied Psychological Measurement*, 2015, 39(7): 525-538
 - 11 Huo Y, de la Torre J. Estimating a cognitive diagnostic model for multiple strategies via the EM algorithm. *Applied Psychological Measurement*, 2014, 38(6): 464-485
 - 12 刘铁川. Mix-DINA 模型功能开发及其与 DINA、MS-DINA 模型的模拟与实证比较. 江西师范大学博士学位论文, 2012
 - 13 Biemer PP. *Latent Class Analysis of Survey Error*. New Jersey: Wiley, 2011
 - 14 Burnham KP, Anderson DR. Multimodel inference. *Sociological Methods & Research*, 2004, 33(2): 261-304
 - 15 白新文,陈毅文. 测量等价性的概念及其判定条件. *心理科学进展*, 2004, 12(2): 231-239
 - 16 Schmitt N, Kuljanin G. Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 2008, 18(4): 210-222
 - 17 王济川,王小倩,姜宝法. *结构方程模型: 方法与应用*. 北京: 高等教育出版社, 2011
 - 18 Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 2002, 9(2): 233-255
 - 19 Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 2000, 3(1): 4
 - 20 Andy HN, Yong L, Jian-zhi C, et al. Culture and state boredom: A comparison between European Canadians and Chinese. *Personality and Individual Differences*, 2015, 75: 13-18
- (收稿日期:2015-12-01)
-
- 15 Kim A-Y. Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, 2015, 32(2): 227-258
 - 16 DiBello LV, Roussos LA, Stout W. A Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models. In Rao CR and Sinharay S, *Handbook of statistics*, Vol. 26: Psychometrics. Amsterdam: Elsevier Science, 2007. 979-1030
 - 17 de la Torre J, Douglas J. Model Evaluation and Multiple Strategies in Cognitive Diagnosis: An Analysis of Fraction Subtraction Data. *Psychometrika*, 2008, 73(4): 595-624
 - 18 Gonthier C, Thomassin N. Strategy use fully mediates the relationship between working memory capacity and performance on Raven's matrices. *Journal of Experimental Psychology: General*, 2015, 144(5): 916-924
- (收稿日期:2016-01-08)