

事前分析的拉丁方任务的复杂性及其对儿童表现的影响

张丽^{①②}, 辛自强^②, 林崇德^{②*}, 李红^①

① 西南大学心理学院, 重庆 400715;

② 北京师范大学发展心理研究所, 北京 100875

* 联系人, E-mail: linchongde@263.net

2008-10-06 收稿, 2008-12-17 接受

国家自然科学基金资助项目(批准号: 30500162)

Zhang L, Xin Z Q, Lin C D, et al. The complexity of the Latin Square task and its influence on children's performance.

Chinese Science Bulletin, 2009, 54(5): 766—775, doi: 10.1007/s11434-009-0079-5

摘要 任务复杂性的分析和评估是心理测量学和认知心理学都非常关注的重要主题。以往有研究者根据关系复杂性理论对拉丁方任务的复杂性进行了分析, 不过其分析存在一些不足。本研究以 195 名二至五年级儿童为被试, 考察了对拉丁方任务事前分析的合理性, 并探讨了复杂性对不同年级儿童表现的影响。结果表明: (i) 对拉丁方任务事前分析的关系复杂性以及任务解决至少需要的加工步骤均能显著预测 Rasch 模型分析的项目难度和儿童正确回答时的反应, 说明对拉丁方任务复杂性的分析具有一定的合理性; (ii) 二至五年级儿童能较好地解决关系复杂性为二元关系或三元关系的拉丁方任务, 但在四元关系任务的解决上仍存在困难。本研究最后探讨了任务复杂性事前分析的特点和问题。

关键词

关系复杂性

加工步骤

拉丁方任务

Rasch 模型分析

任务复杂性是影响个体认知能力测量和评估的重要因素, 因此其分析和评估是心理测量学和认知心理学都非常关注的主题。任务复杂性的分析通常有两种思路: 第一种思路是事后分析, 这是测量学家们通常采取的思路, 如使用通过率和反应时作为任务复杂性的指标等; 第二种思路是事前分析, 目前有不少认知心理学研究者采取这种思路^[1-3]。其中, Halford及其同事^[2]提出的关系复杂性理论(relational complexity)是最具代表性的, 该理论认为可通过分析任务的关系复杂程度事先确定任务的复杂性。

关系复杂性理论从完成任务需要同时加工的关系复杂程度来界定任务复杂性。关系复杂性, 指关系的数量, 即相关的实体或变量的数量。一元关系只有 1 个变量, 如类概念中黑狗是狗的一种, 即狗(黑狗), 这里括号外的表示关系或操作, 括号里表示关系的一个例证。而二元关系则有 2 个变量, 如大象和老鼠哪个比较大, 即较大(大象, 老鼠)。三元关系则有 3 个变量, 如 2, 3, 5 等 3 个数字相加, 即相加(2, 3, 5)。

四元关系则有 4 个变量, 如比例问题 $2/3 = 6/9$ 就有 4 个相互作用的成分。Halford及其同事^[4-7]使用关系复杂性理论分析了大量任务, 这既包括一些经典的皮亚杰任务, 如类包含、传递性推理、等级分类、平衡秤任务等, 也包括诸如句子理解、心理理论、推理、代数运算等各种任务^[8-12]。其中, 他们也对拉丁方任务的复杂性进行了分析^[13]。

拉丁方任务来自古老的智力游戏, 即如何处理方格中的拉丁字母排列方式, 使每个字母在每行每列都出现并且仅出现 1 次。在Birney等人^[13]的研究中, 任务要求 1, 2, 3, 4 四个数字在每行每列都要出现并且只能出现 1 次, 让儿童根据已知方格中的数字推断某一未知方格中的数字是什么。任务共有 3 种难度水平。难度最小的是二元关系任务(图 1), 这类任务只需对单列的数字进行整合就可得到答案, 可表达为: 和(C1(4), C3(2), C4(3)) $\rightarrow$ C2(1)。一条下画线代表一个前提或结论, “和”表示操作。此任务中, 表面上已知 3 个元素, 但实质上问题解决并不需要处理 3 个元素之间的关系,

因此 3 个元素实质上是一个组块, 是一个整体, 构成了一个前提. 由于这类任务涉及一个前提和一个结论, 因而称为二元关系. 中间难度的为三元关系任务(图 1), 未知单元可通过 B 列的数字 1 和 3 及第 2 行的数字 2 推断出来, 表达为: 和(B1(1), B4(3), D2(2)) $\rightarrow$ B2(4). 其中 B 列的数字 1 和 3 是一个整体, 这类任务涉及 2 个前提和 1 个结论, 因此称为三元关系. 难度最大的是四元关系任务(图 1), 可表达为: 和(A1(1), C3(1), D4(3)) $\rightarrow$ B4(1), 涉及 3 个前提和 1 个结论, 称为四元关系. 具体讲, 即 A1 为 1 推论出 A4 不能为 1, C3 为 1 推论出 C4 不能为 1, 加上 D4 为 3, 3 个前提推论出 B4 为 1, 因为每行或者每列都必须有一个 1. 关于任务复杂性的确定, 关系复杂性理论还认为, 如果任务完成需要多个步骤, 每个步骤都需要界定其复杂性. 例如, 某任务需要 3 个步骤才能完成, 第 1 步的加工复杂性为二元关系, 第 2 步为三元关系, 第 3 步为二元关系. 这时任务的关系复杂性由复杂性最大的那个步骤决定, 即该任务的复杂性为三元关系.

	A	B	C	D
1			4	
2	4		?	2
3	1		2	
4			3	

三元关系

	A	B	C	D
1		1		4
2		?		2
3				3
4		3		

三元关系

	A	B	C	D
1	1			
2		3		1
3			1	
4		?		3

四元关系

图 1 Birney等人^[13]设计的 3 类拉丁方任务

然而, Birney等人^[13]对 3 类拉丁方任务难度序列合理性的检验表明, 二元关系中有些项目的难度超过了三元关系中的项目, 四元关系中有些项目的难度低于三元关系中的项目, 这与他们的设想不完全一致. 对此, 他们用加工步骤的多少来解释. 不过, 加工步骤的多少有时是很难确定的, 同一个题目不同被试采取的方法可能不同, 尤其是拉丁方任务中已知数字较多的情况下, 采取不同的思路可能都能解决问题, 这给确定每个题目的加工步骤带来一定困难. 举例来讲, 他们设计的四元关系任务就可以使用策略来解决, 即每行每列有且只有一个 1, 2, 3 或 4, 意味着所有方格中每个数字各有 4 个. 图 1 的四元关系任务中, 3 个不同行列中都有了数字 1, 因此最后一个行列中必然是数字 1. 对于掌握了此策略的被试来说, 此题目的难度必然比较小, 甚至于和二元关系任务的难度相当.

基于以上考虑, 本研究对拉丁方任务进行了以下几方面的改进: (i) 四元关系任务排除了策略的使用; (ii) 所有任务中空白方格的数量均为 11 或 12 个. 实质上, 对于拉丁方任务不管难度如何, 只要合理安排 3~5 个数字就能解决. Birney等人^[13]的研究中基本上已知数字都多于必需的数字, 这使得一个题目可能采用多种思路都能解决, 同时给任务复杂性的分析增加了不确定性. 因此本研究中所有任务方格中的数字都为 4 或 5 个, 基本上都是必需的数字, 这样有利于对加工步骤的分析. 不过本研究分析的是任务解决至少需要的加工步骤. 虽然如前所述同一任务不同被试的加工步骤可能不同, 不过加工步骤的多少确实是任务复杂性的一个重要影响因素^[10,14]. 因此本研究分析了任务至少需要的加工步骤, 这样就能排除因个体策略不同导致的加工步骤的不确定性. 根据关系复杂性和任务至少需要的加工步骤本研究设计了复杂性不同的五类拉丁方任务, 其复杂性特点见方法部分.

然而, 这样的分析合理性如何呢? 这是本研究的第一个研究目的, 即探讨根据关系复杂性和任务至少需要的加工步骤设计的拉丁方任务的难度序列分析是否具有合理性. 对此, 我们采取了两条检验标准: 第一条标准是任务的复杂性是否能较好地预测 Rasch模型分析得到的任务难度; 第二条标准是任务的复杂性是否能较好地预测儿童正确解决任务时的反应时. Rasch模型分析能够确定出任务难度和个体认知能力序列, 而且 Rasch分析得到的难度值具有等距量表的性质, 可以实现在同一个尺度上分析个体的能力和项目的难度^[15]. 与通常作为任务复杂性指标的通过率相比, Rasch分析通过将原始分数转换成 logit 分数克服了通过率分数的一些缺点, 如天花板和地板效应^[16]. 此外, 它也能用来考察项目与理论设想的等级序列的一致性^[17]. 目前有很多心理学研究者使用 Rasch模型分析任务的复杂性^[4,17-20]. 因此本研究将使用 Rasch模型分析而不是通过率来考查任务复杂性.

此外, 本研究也要探讨复杂性对不同年级儿童表现的影响. Andrews和Halford认为^[4], 2 岁儿童能够解决二元关系任务, 5 岁儿童能解决三元关系任务, 11 岁儿童能够解决四元关系任务. 据此本研究将探讨 7~11 岁的小学儿童是否能够很好地解决二元关系、三元关系的拉丁方任务, 而在四元关系任务上仍存在困难.

1 研究方法

(i) 被试. 以整班为单位选取北京某公立小学二至五年级儿童共 195 人为被试. 包括二年级儿童 51 名, 其中男生 25 名, 女生 26 名; 三年级儿童 53 名, 其中男生 32 名, 女生 21 名; 四年级儿童 41 名, 其中男生 22 名, 女生 19 名; 五年级儿童 50 名, 其中男生 28 名, 女生 22 名. 二至五年级儿童平均年龄分别为 7.91, 9.08, 9.78 和 10.98 岁. 所有被试来自不同的家庭.

(ii) 任务和材料. 5 类复杂性水平的拉丁方任务例子见图 2, 其复杂性特点见表 1.

	A	B	C	D
1			4	
2		2	?	
3			2	
4			3	
第1类				
	A	B	C	D
1		1		
2	3	?		2
3				
4		3		
第2类				
	A	B	C	D
1			1	
2	4			
3		2	?	
4			2	
第3类				
	A	B	C	D
1	4			2
2		2		
3	3			?
4		3		
第4类				
	A	B	C	D
1			2	
2	2	4		
3			1	
4		?		3
第5类				

图 2 本研究设计的 5 类任务

表 1 拉丁方任务的复杂性特点^{a)}

项目	二元关系	三元关系	四元关系	至少需要的加工步骤	任务的关系复杂性
第 1 类	1	0	0	1	二元
第 2 类	0	1	0	1	三元
第 3 类	1	1	0	2	三元
第 4 类	0	2	0	2	三元
第 5 类	2	0	1	3	四元

a) 二元关系、三元关系、四元关系三列中的数字表示其数量

第 1 类任务中, 未知单元可通过 C1 为 4, C3 为 2 和 C4 为 3 推断出来, 表达为: 和(C1(4), C3(2), C4(3)) → C2(1). 此任务的完成只需要一个加工步骤, 关系复杂性为二元关系.

第 2 类任务中, 未知单元至少需要一个加工步骤才能完成, 表达为: 和(B1(1), B4(3), D2(2)) → B2(4), 这是一个三元关系. 当然, 此任务的完成也可分为两步来完成, 首先, B1 为 1 和 B4 为 3 推论出 B2 为 2 或者 4, 表达为: 和(B1(1), B4(3)) → B2(2 或者 4), 这是

个二元关系. 接下来, 第一步的结论和 D2 为 2 可推论出 B2 只能为 4, 表达为: 和(B2(2 或者 4), D2(2)) → B2(4), 这是一个三元关系. 根据关系复杂性理论的观点, 多个步骤时任务的复杂性为复杂性最高的那个步骤的复杂性, 因此这时仍为三元关系.

第 3 类任务中, 未知单元可通过 C 列的数字 1 和 2 及第 2 行的数字 4 推断出来, 它至少需要两个步骤: 首先, A2 为 4 推出 C2 不为 4, 表达为: A2(4) → 不是(C2(4)), 这是一个二元关系; 其次, C2 不为 4 结合 C1 为 1 和 C4 为 2, 推出 C3 为 4, 表达为: 和(C1(1), C4(2), 不是(C2(4))) → C3(4), 这是一个三元关系. 对于此任务, 被试也可能采取其他思路来解决, 但不管哪种思路, 都会涉及两个前提和一个结论, 因此此类任务的关系复杂性为三元关系.

第 4 类任务中, 未知单元至少需要两个步骤才能推断出来. 具体地讲, B2 为 2, B4 为 3 和 A1 为 4 推论出 B3 为 4, 即和(B2(2), B4(3), A1(4)) → B3(4), 这是一个三元关系; B3 为 4 结合 A3 为 3, D1 为 2, 推论出未知方格 D3 只能为 1, 即和(A3(3), B3(4), D1(2)) → D3(1), 这也是一个三元关系. 概括来讲, 第 4 类任务包含 2 个三元关系, 其任务复杂性也为三元关系. 同样, 对于此题目, 被试也可能采取其他思路来解决, 但每种思路都会涉及两个前提和一个结论, 因此此类任务的关系复杂性也为三元关系.

最后, 第 5 类任务的完成至少需要 3 个步骤: 首先, A2(2) → 不是(A4(2)), 这是一个二元关系, 其次, C1(2) → 不是(C4(2)), 这也是一个二元关系. 最后, A4 不为 2, 加上 C4 不为 2, 和 C4 为 3, 根据规则每行每列必须有且只能有一个 2 的原则, 推论出 B4 为 2, 可表达为: 和(不是(A4(2)), 不是(C4(2)), D4(3)) → B4(2), 这是一个四元关系. 整体上, 5 类任务的难度是依次增加的.

拉丁方任务中每类复杂性任务有 3 个项目, 共 15 个项目. 任务采用计算机随机呈现. 指导语强调既快又准确地完成任务. 任务呈现时, 同时给儿童呈现可供选择的答案, 即 1, 2, 3, 4 四个选项. 每道题目的内容和备选答案在相同页面出现. 计算机记录儿童的反应时和回答选项. 每个项目儿童做出正确回答, 记为 1 分, 错误回答记为 0 分. 这样儿童在每个复杂性水平任务上的得分就处于 0~3 之间.

(iii) 研究程序. 首先根据学生的学习成绩筛选掉那些存在严重的学习困难的儿童. 采取小团体集体施测, 每次施测 15~20 个儿童, 由 4 个心理学专业的研究

生担任主试. 正式施测前, 由一名主试统一讲解每个任务的题目要求, 并要求儿童练习 3 个题目, 以便儿童熟悉题目要求和作答方式. 任务是在儿童的一次计算机课堂上施测的, 完成任务大约需要 15~20 min.

2 结果与分析

2.1 拉丁方任务的难度分析

使用 Winsteps 3.63.0 软件^[21,22]进行了 Rasch 模型分析. 首先, 对数据进行检查看是否有在所有题目上回答均正确或者均错误的被试和是否有所有被试均回答正确或均错误的项目. 因为 Rasch 模型认为这些被试和项目无法提供任务难度和个体能力的信息, 因此这些题目或被试将被排除在分析之外. 本研究中有 1 名被试在所有项目上回答均正确, 因而进入 Rasch 分析的数据实质上有 194 个.

Rasch 模型分析与一般的模型分析不同, 其主要任务是产生拟合 Rasch 理论设想的数据, 而不是产生拟合数据的模型^[15]. 关于 Rasch 模型的理论设想和事实上收集到的数据之间的拟合程度, 通常有两种不同的方法: 内拟合(infit)和外拟合(outfit). 内拟合对

于接近项目难度值的个体分数给予了更多权重, 而外拟合没有对数据进行权重, 对极端数据的影响更敏感. 因此, 多数研究者更为关注内拟合值^[15]. 使用 Rasch 模型分析拉丁方任务的项目难度, 结果见表 2. 表中 MNSQ 为均方值, 预期值为 1. 拟合值较高, 说明实际的数据变异比理论模型期望的要高; 拟合值低, 说明实际的变异比期望的低, 数据过于拟合. 可接受的范围根据测验类型的不同也存在一些差异, 对多项选择题一般接受范围是 0.70~1.30^[23]. 本研究中拉丁方任务正是要求儿童进行多项选择, 因此我们就根据拟合值是否在 0.70~1.30 判断数据是否和模型拟合较好. 除了 MNSQ 值, 很多研究者也使用 t 值来检验项目的拟合程度. t 值代表均方标准化残差值, 预期值为 0, 通常认为可接受的范围是 -2~+2. 因为 t 值是标准化残差值, 更容易受样本的影响, 通常随着样本的增加, t 值大于 2 的项目数量会有很大增加. 因此本研究主要依据 MNSQ 结果做出结论. 从表 2 中可以看出, 所有项目均与 Rasch 模型拟合较好. 此外, 项目分离信度值为 0.98, 表明 15 个项目测量了一个潜在的变量即复杂性水平, 可靠性比较高.

表 2 拉丁方任务的 Rasch 分析和传统分析

项目	Rasch 分析结果				传统分析			
	难度	SE	内拟合 (MNSQ)	通过率	平均反应时		正确回答反应时	
					M	SD	M	SD
第 1 类 1	-1.54	0.24	0.80	0.88	6.98	6.87	7.28	6.77
第 1 类 2	-1.92	0.27	0.89	0.91	5.68	5.11	5.56	3.92
第 1 类 3	-2.07	0.28	0.97	0.92	6.77	10.83	6.88	11.09
平均	-1.84	0.16						
第 1 类 1	-1.99	0.27	0.78	0.92	8.42	8.34	8.19	7.24
第 1 类 2	-1.49	0.23	0.94	0.88	7.26	6.83	7.55	7.13
第 1 类 3	-1.66	0.24	0.83	0.89	8.64	8.45	8.56	8.14
平均	-1.71	0.15						
第 3 类 1	2.36	0.19	1.05	0.22	8.66	10.56	14.51	17.49
第 3 类 2	2.00	0.18	0.94	0.28	10.34	16.09	13.00	17.69
第 3 类 3	2.26	0.18	0.90	0.24	9.87	20.63	11.93	9.62
平均	2.20	0.11						
第 4 类 1	-0.06	0.17	1.16	0.68	13.38	27.49	14.18	32.04
第 4 类 2	-0.87	0.20	1.05	0.81	12.70	22.58	12.33	22.77
第 4 类 3	0.24	0.16	1.18	0.62	10.33	10.95	9.64	12.39
平均	-0.23	0.33						
第 5 类 1	1.76	0.17	1.00	0.32	13.95	12.44	14.94	12.56
第 5 类 2	0.97	0.16	0.97	0.47	16.49	22.37	15.54	16.76
第 5 类 3	2.03	0.18	0.93	0.27	20.89	50.29	22.39	22.04
平均	1.59	0.32						

个体拟合方面(person fit), 一个重要指标是分离信度, 这与经典测量理论中的信度值相当. 本研究中分离信度为 0.42, 可见Rasch模型对个体能力的测量有中等程度的可靠性. 进一步对个体拟合值的分析显示: 所有被试中内拟合值MNSQ处于 0.70~1.30 的约有 44.3%, 还有 34.0%的被试MNSQ小于 0.70, 21.7%的被试MNSQ值大于 1.30. 然而, 值得注意的是, 虽然MNSQ可接受的范围是 0.70~1.30, 不过拟合值小于 0.70 的被试, 其反应心理意义是很明确的且合理的, 与格特曼模式拟合, 即在所有容易的项目上都正确, 在所有困难的项目上都错误^[15]. 而大于 1.30 的数据则不太合理、偶然性较强, 因为这些被试在困难的项目上回答较好, 而在容易的项目上回答较差. 可以说这些被试的反应没有得到很好的测量. 因此, 随后关于复杂性对儿童成绩影响的研究中没有包括这 21.7%的被试, 有 11 名二年级被试, 10 名三年级被试, 6 名四年级被试和 15 名五年级被试, 还剩余 153 名被试.

2.2 拉丁方任务的复杂性与项目难度、反应时

关系复杂性、任务至少需要的加工步骤与 Rasch 模型分析的项目难度、反应时的相关分析结果见表 3, 其中反应时指正确回答的反应时, 以下类同.

表 3 相关矩阵^{a)}

	1	2	3	4
1 关系复杂性	—			
2 任务至少需要的加工步骤	0.85**	—		
3 项目难度	0.64*	0.81***	—	
4 反应时	0.81***	0.90***	0.80***	—

a) *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$

从表 3 可以看出, 关系复杂性、任务至少需要的加工步骤与 Rasch 分析的项目难度和反应时都有显著相关. 进一步回归分析如下: (i) 以关系复杂性和任务至少需要的加工步骤为自变量, 项目难度为因变量进行回归分析, 共线性诊断结果表明两个自变量存在一定的共线性, 容忍度为 0.29, 方差膨胀因子为 3.50, 因此没有继续进行多元回归分析, 而是分别以关系复杂性和任务至少需要的加工步骤为自变量进行了一元回归分析. 结果显示任务至少需要的加工步骤对任务难度的回归系数显著, $\beta = 0.81$, $P < 0.001$, $R^2 = 0.66$, 关系复杂性对任务难度的回归系数也显著, $\beta = 0.64$, $P < 0.05$, $R^2 = 0.41$; (2) 分别以关系复杂性和任务至少需要的加工步骤为自变量以反应

时为因变量进行了一元回归分析. 结果显示任务至少需要的加工步骤对反应时的回归系数显著, $\beta = 0.90$, $P < 0.001$, $R^2 = 0.80$, 关系复杂性对反应时的回归系数也显著, $\beta = 0.81$, $P < 0.001$, $R^2 = 0.66$. 从这些结果可以看出, 不管是对任务难度还是对反应时的预测作用, 任务至少需要的加工步骤比关系复杂性的预测作用均较大. 此外, 任务至少需要的加工步骤和关系复杂性对反应时的预测作用比它们对任务难度的预测作用均较大.

关系复杂性对任务难度的预测作用在本研究中只有 41%, 而Birney等人^[13]的研究中关系复杂性对任务难度的预测作用达到 64%. 对此, 我们做了一个简单的对应图以更直观地观察本研究中项目难度和关系复杂性的关系(图 3). 可看出, 第 3 类任务中有些项目难度要大于第 4 类和第 5 类中的项目. 对儿童在第 3 类任务上的错误反应模式进行分析发现(以图 2 中的任务为例), 很多儿童认为问号所在方格即 C3 应为 3. 这些儿童根据 A2 为 4, 推论出 C2 不能为 4, 同时看到 C1 为 1 和 C4 为 2, 知道 C3 只能是 3 或者 4. 然而, C2 不能为 4 是个否定前提, 需要和 C3 只能是 3 或者 4 这个肯定前提相结合, 才能做出推论. 正是这个否定前提和肯定前提相结合的问题给很多儿童造成了困难. 甚至于在我们对几个成人被试进行施测时, 也有被试在这儿出现了错误, 对此现象我们将在讨论部分进行详细解释. 概括来讲, 主要是第 3 类任务的实际难度与关系复杂性的设想不一致. 如果去掉此类任务的 3 个项目, 项目的关系复杂性与其实际难度的相关则有了提高($r = 0.85$, $P < 0.001$). 进一步回归分析显示关系复杂性对项目难度的预测作用达到 72%, $P < 0.001$.

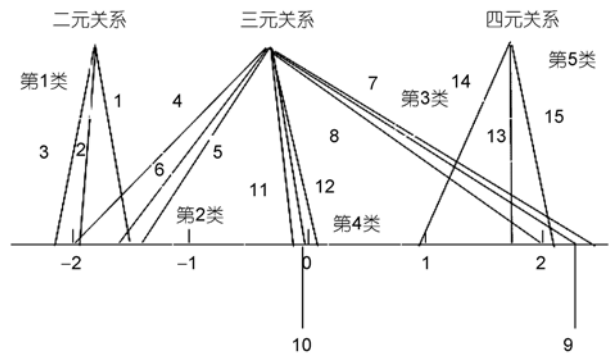


图 3 基于 Rasch 模型的项目难度估计(Logit)

另外,为什么关系复杂性和任务至少需要的加工步骤对反应时的预测作用比它们对任务难度的预测作用均较大?是不是任务难度和反应时没有都反映任务复杂性的特点?一般来讲,如果儿童完成任务时存在速度-准确性权衡倾向,那么不管儿童是速度倾向还是准确性倾向,两者不可能同时都准确反映任务的复杂性特点.因此我们分析了儿童在完成本研究中的任务时是否存在速度-准确性权衡倾向.结果表明,15 个项目的通过率和反应时(正确回答的反应时)存在显著负相关, $r = -0.78$, $P < 0.001$,即整体上通过率较低的任务其反应时也较长,这意味着儿童在完成这些任务时并没有显示出速度-准确性权衡倾向.不过,从图 4 可以看出,项目 7, 8, 9 三个题目,即第 3 类任务中的 3 个项目,通过率很低,同时反应时较短,说明儿童在这 3 个项目上存在一定的速度倾向.当然,这种倾向有可能是无意的,有可能儿童觉得任务非常简单或者不知如何作答,所以迅速做出反应.因此,可能是因为儿童在完成第 3 类任务中的项目时存在一定的速度倾向影响了任务至少需要的加工步骤和关系复杂性对反应时、项目难度的预测作用.去掉第 3 类任务后,关系复杂性和任务难度、反应时的相关分别为 0.85 和 0.83,且关系复杂性对任务难度和反应时的解释率分别为 72% 和 69%;任务至少需要的加工步骤和任务难度、反应时的相关分别为 0.97 和 0.90,且任务至少需要的加工步骤对任务难度和反应时的解释率分别为 94% 和 81%.这时,不管是关系复杂性还是任务需要的加工步骤对任务难度和反应时的作用均比较接近.

以上是从项目水平进行的分析,下面将通过分

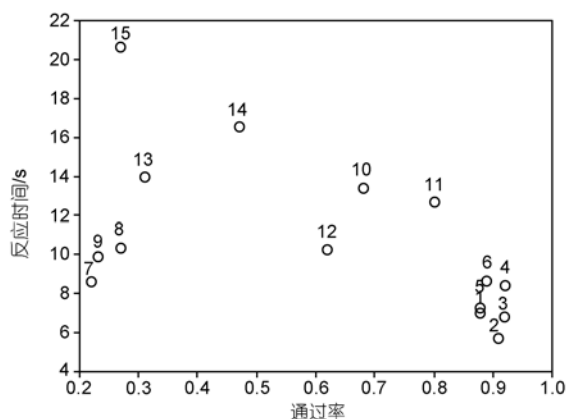


图 4 通过率和反应时的相关散点图

析 5 类任务的难度之间是否存在显著差异来检验关系复杂性对任务难度序列的设想是否合理.在 Rasch 分析的基础上,进行了相关样本的 t 检验,每类任务的平均数和标准误差见表 2.结果表明,除了第 1 类任务和第 2 类任务的难度不存在显著差异外,其他几类任务之间的差异均显著,难度最大的为第 3 类任务,其难度显著高于第 1 类、第 2 类、第 4 类和第 5 类, t 值分别为 21.26, 20.58, 6.39 和 2.10, $P < 0.05$; 其次是第 5 类任务,其难度显著高于第 1 类、第 2 类和第 4 类任务, t 值分别为 10.39, 10.00 和 3.71, $P < 0.05$; 接下来是第 4 类任务的难度,其难度显著高于第 1 类和第 2 类, t 值分别为 4.74 和 4.63, $P < 0.05$.

2.3 任务复杂性对儿童表现的影响

本研究以儿童正确回答的项目数占项目总数的百分比作为儿童在每类复杂性水平任务上的表现或成绩.以年级和性别为自变量,儿童在拉丁方任务上的成绩为因变量进行多因素方差分析,描述性统计见表 4.结果表明: (i) 复杂性最大的第 3 类任务上存在显著的年级差异, $F_{(3,145)} = 5.80$, $P < 0.01$, $\eta^2 = 0.11$,事后检验表明五年级学生的成绩显著优于二、三、四年级学生, $P < 0.01$,性别主效应和交互作用均不显著; (ii) 第 5 类任务上存在显著的年级差异, $F_{(3,145)} = 5.81$, $P < 0.01$, $\eta^2 = 0.11$,事后检验表明五年级学生的成绩显著优于二、三、四年级学生, $P < 0.01$.性别主效应和交互作用均不显著; (iii) 第 4 类任务上女生的成绩显著优于男生, $F_{(1,145)} = 4.09$, $P < 0.05$, $\eta^2 = 0.03$,年级主效应和交互作用均不显著; (iv) 复杂性最小的第 1 类、第 2 类任务上年级主效应、性别主效应和交互作用均不显著.概括来讲,儿童在复杂性较大的第 3 类、第 5 类拉丁方任务上成绩有较大变化,这主要表现在五年级儿童的成绩相对于二、三、四年级儿童有很大提高,说明在这两类拉丁方任务上二~四年级儿童的成绩比较平稳,不过从四年级到五年级儿童成绩迅速提高.

值得注意的是,虽然儿童在复杂性较大的第 3 类和第 5 类任务上的成绩有很大提高,但儿童并没有能够很好地解决这两类任务.第 3 类任务上即使五年级儿童也只能解决约 30% 的项目;第 5 类任务上即使五年级儿童也只能解决约 47% 的项目.可以说儿童在这两类任务的解决上还存在一定困难.而第 1 类和第 2 类任务上儿童能解决 92%~98% 的项目,第 4 类任务上能解决 70%~80% 的项目.看来小学儿童已能较好地解决复杂性较小的这 3 类任务.

表4 儿童在拉丁方任务上的成绩

	二年级		三年级		四年级		五年级	
	M	SD	M	SD	M	SD	M	SD
第1类	0.93	0.14	0.92	0.14	0.98	0.08	0.98	0.08
第2类	0.93	0.16	0.95	0.14	0.97	0.12	0.95	0.14
第3类	0.12	0.21	0.12	0.20	0.13	0.25	0.31	0.30
第4类	0.70	0.31	0.82	0.23	0.81	0.23	0.73	0.28
第5类	0.23	0.23	0.26	0.25	0.32	0.29	0.47	0.28

3 讨论

研究结果显示,设计的5类拉丁方任务与 Rasch 模型拟合较好,说明所有项目测量了潜在的复杂性变量。另一重要结果是任务的关系复杂性、至少需要的加工步骤都能显著预测 Rasch 分析的项目难度以及儿童正确回答任务时的反应时,说明我们对任务复杂性的分析具有一定的合理性。

与 Birney 等人^[13]的研究相比,本研究中拉丁方任务的关系复杂性对 Rasch 分析的项目难度预测作用相对较小,分析发现其原因主要在于第3类任务的实际难度与我们对其复杂性分析的设想存在不一致。我们预测第3类拉丁方任务的难度应小于第4类和第5类,同时大于第1类和第2类任务,而实际结果表明第3类任务是5类任务中难度最大的。如前所述,这可能是第3类任务中否定前提和肯定前提相结合的问题给儿童造成了困难,也就是一些认知操作因素给儿童带来了困难。具体来讲,原因可能有两方面:首先,类似于传递性推理中发现的“一致性效应”^[24],若推理的前提都是肯定或者都是否定其难度要小于一个前提是肯定另外一个前提是否定;其次,类似于传递性推理中发现的“转化效应”^[24],当两个前提没有共同项时难度较大,这时需要转换成含有共同项的命题。以图2中的任务为例,不是直接将两个前提NOT(C2(4))和C3(3或者4)(C1为1和C4为2推论出的结果)进行整合,而是首先将NOT(C2(4))和C2(3或者4)两个前提结合起来,共同项是C2,得出C2为3。然后C2为3意味着C3不能为3。最后,将C3不能为3和C3(3或4)进行整合得出C3为4,其中共同项是C3。

不过,除了以上认知操作因素给儿童解决拉丁方任务带来了困难,工作记忆能力的不足可能也对儿童解决拉丁方任务有很大影响。以往有很多研究表明,工作记忆对推理有很大影响,比如信息的储存、监控、抑制和协调等都会影响个体在推理任务上的成绩^[25,26]。前述提到我们对几个成人进行施测时也有被试出现了错误,不过只要对他们稍加提示,如要求他们再仔细看

一遍,他们就能立即纠正过来。这提示我们有可能工作记忆尤其是执行控制对儿童在第3类拉丁方任务上的成绩也有影响。进一步的研究可以探讨工作记忆和儿童在第3类任务上成绩的关系。此外,空白方格的数量可能也对第3类任务的难度有影响。第4类和第5类任务的解决都包含有否定前提和肯定前提,但是这两类任务的已知数字较第3类任务要多,这一方面有助于儿童采取不同的思路完成任务,另外一方面已知数字较多会使儿童对信息的加工时间增多,而且儿童会觉得数字较多的题目复杂性高,因而会仔细深入地思考。

关于任务复杂性对儿童成绩的影响,本研究表明,二至五年级儿童能够较好地解决第1类、第2类和第4类任务,而第3类和第5类任务的解决还存在一定困难,尽管从四年级到五年级儿童解决这两类任务的能力有了很大提高。第1类、第2类和第4类任务的关系复杂性分别为二元关系和三元关系,第5类任务的关系复杂性为四元关系,因而此结果说明二年级儿童已能解决关系复杂性为二元关系和三元关系的任务,不过直至五年级儿童还不能很好地解决四元关系任务,这和 Andrews 和 Halford 的观点^[4]是一致的,即11岁以下的儿童还不能很好地解决四元关系任务。而第3类任务的关系复杂性也是三元关系,之所以直至五年级多数儿童还不能很好地解决此任务,正如前述讨论所言,是认知操作因素给儿童造成了很大困难。

整体上,本研究对拉丁方任务的复杂性分析具有很大程度的合理性,这丰富了关系复杂性理论。不过对任务复杂性的分析有几点需要说明:(i)这种分析是以能够比较清楚地了解任务的加工过程为基础的。对此,Halford 等人^[24]也曾提到,任务复杂性分析的第一步就是发展任务的过程模型,并从经验上证实它。本研究对拉丁方任务的分析就是如此,不仅分析了关系复杂性,还分析了任务至少需要的步骤。分析任务至少需要的解题步骤,是针对有多种解决思路和方法的任务,如果任务只有一种解决方法,则直接分析其加工步骤即可^[27]。这样的事先对任务复

杂性的分析也更准确。(ii) 关系复杂性和至少需要的加工步骤是从不同角度对任务复杂性进行的分析, 不过两者也存在一定的关系。一个任务如果有多种解决方法, 通常关系复杂性较大的加工步骤较少, 关系复杂性较小的加工步骤较多。本研究中我们是根据关系复杂性较大、加工步骤较少的解法来界定任务的复杂性。(iii) 本研究旨在提供任务复杂性分析的一个角度。实际上还有其他因素会对任务复杂性有影响, 如任务涉及关系的等级和嵌套层次、认知操作性质、知识经验等^[1,10,14], 这些因素本研究尚未深入探讨。这也是本研究的第3类拉丁方任务和我们的预期有些出入的原因, 因为此任务涉及了认知操作的复杂性。将来的研究可以深入考察各种因素对任务复杂性的影响。(iv) 实际生活中很少有关系复杂性大于四元关系的任务, 这一点Halford等人^[2]通过引用很多文献论证了人类能够加工的最多是四元关系, 如果超过了四元关系必然会被分解成多个步骤以降

低关系复杂性。另外, 关系复杂性有很大的发展特点, 正如Andrews和Halford认为2岁儿童能够解决二元关系任务, 5岁儿童能解决三元关系任务, 11岁儿童能够解决四元关系任务^[4]。这说明在探讨儿童认知发展时任务的关系复杂性分析具有重要意义。

最后, 关于Rasch分析有一点需要说明, 即Rasch分析与通常的项目反应理论不同, 其任务是判断数据是否拟合Rasch模型的理论设想, 而不是产生拟合数据的模型, 因此常常会有数据不拟合Rasch模型。正如本研究有一些被试的能力拟合值不在理论范围内。他们的反应不太合乎常理, 即在有些困难的题目上回答正确, 但在容易的题目上却回答错误。这种结果有两种原因: 其一, 这些儿童凭借猜测做出回答, 比较幸运在有些困难的题目上回答正确; 其二, 这些儿童具有特定的知识, 表现在不具备容易项目的知识而具备困难项目的知识^[15]。不过, 将来的研究尚需对这部分被试进行深入的研究。

参考文献

- 1 Commons M L, Trudeau E J, Stein S A, et al. Hierarchical complexity of tasks shows the existence of developmental stages. *Dev Rev*, 1998, 18: 237—278[DOI]
- 2 Halford G S, Wilson W H, Phillips S. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behav Brain Sci*, 1998, 21: 803—831
- 3 辛自强. 关系-表征复杂性模型的检验. *心理学报*, 2003, 4: 504—513
- 4 Andrews G, Halford G S. A cognitive complexity metric applied to cognitive development. *Cogn Psychol*, 2002, 45: 153—219[DOI]
- 5 Andrews G, Halford G S. Children's ability to make transitive inferences the importance of premise integration and structural complexity. *Cogn Dev*, 1998, 13: 479—513[DOI]
- 6 Halford G S, Andrews G, Jensen I. Integration of category induction and hierarchical classification: One paradigm at two levels of complexity. *J Cogn Dev*, 2002, 3(2): 143—177[DOI]
- 7 Halford G S, Andrews G, Dalton C, et al. Young children's performance on the balance scale: The influence of relational complexity. *J Exp Child Psychol*, 2002, 81: 383—416[DOI]
- 8 Andrews G, Halford G S. Complexity effects are found in all relative clause sentence forms. *Behav Brain Sci*, 1999, 22: 95[DOI]
- 9 Andrews G, Halford G S, Bunch K M, et al. Theory of mind and relational complexity. *Child Dev*, 2003, 74: 1476—1499[DOI]
- 10 Birney D P, Halford G S. Cognitive complexity of suppositional reasoning: An application of the relational complexity metric to the knight-knave task. *Think Reason*, 2002, 8 (2): 109—134 [DOI]
- 11 Halford G S, Andrews G. The development of deductive reasoning: How important is complexity? *Think Reason*, 2004, 10 (2): 123—145[DOI]
- 12 English L D, Halford G S. Mathematics education: Models and processes. Hillsdale: Erlbaum, 1995
- 13 Birney D P, Halford G S, Andrews G. Measuring the influence of complexity on relational reasoning: The development of the Latin Square Task. *Educ Psychol Measure*, 2006, 66(1): 146—171[DOI]
- 14 辛自强. 关系-表征复杂性模型. *心理发展与教育*, 2007, 22(3): 122—128
- 15 Bond T G, Fox C M. Applying the Rasch model: Fundamental measurement in human sciences. Mahwah: Erlbaum, 2001
- 16 Green K E, Smith R M. A comparison of two methods of decomposing item difficulties. *J Educ Statist*, 1987, 12: 369—381[DOI]
- 17 Dawson-Tunik T L, Commons M, Wilson M, et al. The shape of development. *Eur J Dev Psychol*, 2005, 2 (2): 163—195[DOI]
- 18 Wellman H M, Liu D. Scaling of theory-of-mind tasks. *Child Dev*, 2004, 75(2): 523—541[DOI]
- 19 Goodheart E A, Dawson T L. A Rasch analysis of developmental data from the Laundry Problem Task Series. Poster presented at the 11th Annual Adult Development Symposium. Boston, MA. 1996
- 20 Müller U, Sokol B, Overton W. Developmental sequences in class reasoning and propositional reasoning. *J Exp Child Psychol*, 1999, 74: 69—106[DOI]
- 21 Linacre J M. A user's guide to WINSTEPS: Rasch-model computer software program. <http://www.winsteps.com/winsteps.htm>
- 22 Linacre J M, Wright B D. Multiple-choice, rating scale, and partial credit Rasch analysis (computer software). Chicago: MESA Press, 2000
- 23 Wright B D, Linacre J M. Reasonable mean-square fit values. *Rasch Measurement Transactions*, 1994, 8(3): 370
- 24 张仲明, 李红. 传递性推理心理效应的研究. *南京师大学报(社会科学版)*, 2006, 1: 111—114
- 25 张清芳, 朱滢. 工作记忆和推理. *心理学动态*, 2000, 8(1): 12—17
- 26 白学军, 臧传丽, 王丽红. 推理与工作记忆. *心理科学进展*, 2007, 15(1): 16—21
- 27 Xin Z Q. Fourth through sixth graders' representations of area-of-rectangle problems: Influences of relational complexity and cognitive holding power. *J Psychol: Interdisciplin Appl*, 2008, 142(6): 581—600[DOI]