

# 瑞文标准推理测验在我国的修订<sup>1)</sup>

R' SPM-CR 全国协作组

张厚粲 王晓平 执笔

北京师范大学心理系

## 摘 要

为了满足科研与实际工作对团体智力测验的迫切需要,由17个单位组成的全国协作组对瑞文标准推理测验进行了修订。常模团体的建立以1982年全国人口普查资料为依据,按各地区人口分布,各类人员比例取样,共包括年龄跨度从5岁半至70岁以上的5108人。经过标准化的施测手续,测定了1986年中国城市常模。

测验结果的项目分析、信度和效度等技术指标达到或超过了国外同类研究。

研究发现除40岁组外,其他各年龄组均无显著男女差异。本研究还对中国大陆城市青少年的成绩与香港、英国和新西兰相应年龄组的成绩做了比较。

## 一、瑞文推理测验简介与本研究的目的

自一九七九年心理测验研究工作恢复以来,我国心理学工作者修订了一批世界公认较好的智力、性格及职业选拔等方面的量表与问卷,并且已经完成和正在编制一些我国自己的量表。随着科研工作与社会实践的要求,心理测验的应用越来越广泛。

实际应用发现,现有的智力测验大都是个别施测的量表,如韦氏智力量表与比内智力量表等。这些测验对个体做能力优劣的分析是适用的,但用于智力普查与筛选工作或对某一团体智力发展水平进行鉴定就显得费时费力。为了向研究者提供一套方便的智力评价工具我们决定修订瑞文成套测验,首先对瑞文标准推理测验做修订。

瑞文标准推理测验(Raven's Standard Progressive Matrices简称 R. SPM)是英国心理学家瑞文(J. C. Raven)1938年设计的非文字智力测验<sup>[1]</sup>。

瑞文标准推理测验的编制,理论上根源于本世纪初期斯皮尔曼(C. Spearman)对智

1)本文于1987年11月21日收到。

• 本协作组由张厚粲教授领导,各单位与主要负责人如下:华北区:北师大张厚粲,王晓平,天津师大石德澄、张蓉,河北沧州师专吴长亮,山西大学章竞思,太原师专晁连生;东北区:哈尔滨师专刘庆生,哈尔滨32中秦立群;华东区:上海市教科所顾天祐,高德建,江苏淮阴师范张绍京,福建三明师专白永莲,黄草木;中南区:河南大学丁秀峰,河南新乡红旗区教师进修学校杜希敏,新乡新华区教师进修学校杜兴骥,湖北沙市教科所陈光浩;西南区:四川教育学院陈国英,云南教育学院胡育英、唐忠;西北区:西安交大宋合义,甘肃教育学院童长江。

北京师大马世晔、庞理立同学对测验数据处理做了大量工作,姜扬、袁晓京、陈海平及心理系八四、八三级同学参加测验信度、效度搜集工作。心理系冯伯麟、教育系韦小满同志也为该测验修订提供过帮助。

力本质的研究。众所周知,斯皮尔曼主张智力的二因素论,认为任何活动都包含一般因素(即“g”因素)和特殊因素(即“s”因素),通常所谓一个人的智力可以用g因素解释,但是斯皮尔曼也认为笼统地用智力一词还不足以描绘人的多种认知能力,于1927年明确提出存在着两种既对立又协作,有着发生学上内在联系的行为——再生和推断,并且强调分别探讨这两种行为的实质,对于个体能力心理学具有极为重要的意义。依据上述观点,Raven把智力的g因素划分为两种相互独立的能力,一种称再生性能力(reproductive ability),一种称推断性能力(eductive ability)。再生性能力指一个人当前所具备的回忆已获得信息并进行言语交流的能力。它表明一个人通过接受教育达到的水平。显然,它和学校的教学内容有着密切的关系。瑞文认为,测量这种能力用词汇测验最为有效,为此他编制了一套词汇量表称“米尔·希尔词汇量表”(Mill Hill Vocabulary Scale, MHV)。推断性能力指一个人作出理性判断的能力,是一个人智能活动的能量(capacity)。它与一个人的知识多少或受教育水平关系并不紧密,但对于适应社会生活却有重要意义。为了测量推断性能力,他编制了一种非言语测验,全部采用几何图形,这就是著名的“瑞文推理测验”(Raven's Progressive Matrices, RPM)。瑞文推理测验由易到难分成三个不同水平的测验:一是“瑞文彩图推理测验”(Raven's Colored Progressive Matrices, CPM),适用于幼儿和智力水平较低的人;二是“瑞文标准推理测验”,即我们现在修订的SPM,适用于所有年龄在5.5岁以上且智力发展正常的人;三是“瑞文高级推理测验”(Raven's Advanced progressive Matrices, APM),适用于在SPM上得高分或者智力水平较高的人。但从产生年代来讲,APM是SPM向更高智力水平一端的扩展,它最初使用是在1941年,而CPM则是SPM向更小年龄,更低水平一端的扩展,最初使用是在1947年。显然,SPM在瑞文测验系列中处于最重要的地位<sup>[2,7]</sup>。

瑞文测验的优点在于测验对象不受文化,种族与语言的限制,并且可用于一些生理缺陷者。测验既可个别进行,也可团体实施,使用方便,省时省力,结果解释直观、简单。瑞文测验的另一优点是适用的年龄范围宽,有儿童,成人和老人,年龄下限为5岁半,年龄上限不定。

我们的修订工作从1985年10月成立全国协作组开始,第一步先进行瑞文标准推理测验(1958年版)中国城市版的修订。然后延伸到高级推理测验和彩图推理测验的修订以及各测验农村版的修订。这项工作受到原编制人之子,当代瑞文推理测验的负责人,英国心理学家约翰·瑞文(J. Raven)博士的支持。

## 二、测验修订过程与施测方法

测验修订过程包括测试样本标准化与测验标准化两方面。

### 1. 常模团体的建立

这次测验修订的取样遵循两个原则:年龄分组与各年龄组人数基本按照瑞文所取英国样本<sup>[9]</sup>并参考其他国家取样,和每个年龄组内被试来源遵从1982年全国人口普查资料按比例安排。

由于我国地域广阔,人口众多,为了避免城乡文化差异的影响,这次修订不采取城乡混合版而先制定中国城市版。常模团体5,108人。

年龄组是这样划分的, 5 岁半至 16 岁半每半岁为一个年龄组(半岁组从 3 个月 1 天至 8 个月 30 天、整岁组从 9 个月 1 天至下一年 2 个月 30 天), 每组 150 人至 200 人; 20 岁以上每十岁为一个年龄组(如 20 至 29 岁为 20 岁组, 40 至 49 岁为 40 岁组)。与国外研究所不同的是我们补充了 17 至满 19 岁这样一个年龄组, 目的是使各年龄组前后衔接, 共 30 个年龄组, 其中 70 岁以上的老人为一组, 该组人数为 61。各年龄组被试根据人口普查材料的比例从东北、华北、华东、西南、中南、西北六大区城市人口中抽取, 同时考虑到全国大、中、小城市人口的比例。在对成人被试抽取时照顾到文化、职业之人口比例。所取被试男女也基本按比例抽取, 其中男性 2,713 人, 女性 2,395 人<sup>[4]</sup>。

## 2. 施测

修订后的测验指导语基本不变, 施测可以用团体与个别两种方法实施。答题形式分为被试说出正确答案号与写出正确答案号二种。一般对小于 8 岁的和 60 岁以上被试用个别测试法, 主试为被试记录答案, 其余被试(除智力极低者外)均可使用团体纸笔式或个别纸笔式。测验在无时间限制的条件下进行, 大约需要时间为 30 至 40 分钟。

## 3. 测验材料与构成

测验时, 每位被试应用测验图册一本, 答卷纸一张, 自备铅笔 1、2 枝(主试需有备用铅笔)。

瑞文标准推理测验由 60 道题目组成, 分为 5 个系列, 每个系列有其重点与操作水平。每一系列包含 12 个题目。在难度排列上, A 列最易, E 列最难, 各列中也是前边的题较后面的题更容易。A 列由一幅缺一小块的大图案和 6 个供选择补全所缺部分的小图案组成(见图 1)。B 列是  $2 \times 2$  的图形矩阵和 8 个可选答案。直观上看 A 列的任务是完成图形, 主要测查对图形的观察、比较与想象能力; B 列主要测对图形的类比能力; C 系列测对图形系统变化的认识; D 系列测对图形排列与系统变化的认识; E 系列测对图形分析与综合的能力。被试的任务是认识图形变化规律, 在备选答案中找出合适的小图案以补充主题图中的缺失部分, 并把该图案的序号填入答卷纸上相应的格子里。

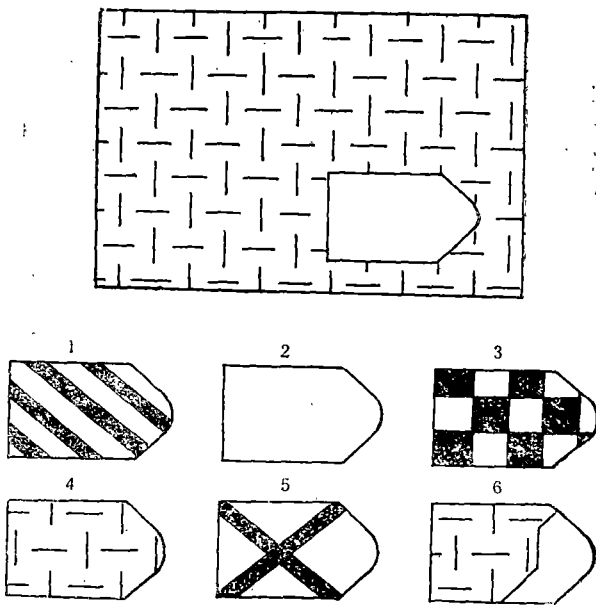


图 1 例图 A<sub>1</sub> 正确答案是 4

在测验修订过程中, 我们分析了各种被试答题的倾向与系统错误并且统计了各题目的难度值(结果见表 1)。

## 4. 记分与分数解释

记分过程分为记原始分和转换为标准分两步。记分键提供了对答案做 1 或 0 记分的

标准,我们根据各年龄组被试的分数分布制定出常模表,测试者可以依据常模表了解每个被试的量表分。量表分以百分等级表示,每个被试又可从所得百分等级了解自己所属智力水平。I级——测验成绩等于或高于同龄组95%的人,具有高水平的智力;II级——测验成绩在75%与95%之间,智力水平良好;III级——成绩在25%与75%之间,为中等智力水平;IV级——成绩在5%与25%之间,智力水平较差;V级——成绩在最低的5%内,为智力缺陷。

测验修订过程中除取得常模表外还取得各年龄组的平均数与标准差等数据(结果见表2)。

### 三、结果与讨论

#### 1. 题目分析

我们对来自六大区各年龄组的3,000名被试的答卷,做了题目的难度分析,得到各题难度值表(如表1)。

表1 瑞文标准推理测验(中国城市版)难度分析

题目 系列	1	2	3	4	5	6	7	8	9	10	11	12
A	.99	.98	.97	.97	.96	.94	.84	.88	.84	.70	.65	.50
B	.97	.88	.88	.77	.71	.68	.69	.57	.62	.69	.57	.49
C	.83	.83	.81	.65	.73	.66	.71	.56	.67	.51	.47	.21
D	.76	.77	.74	.72	.74	.62	.57	.56	.56	.53	.27	.15
E	.64	.55	.51	.51	.54	.49	.39	.40	.38	.24	.13	.16

从表1可见,60个测验题在各系列内程度递增,各系列间难度也递增,而且难度顺序与原题顺序基本一致,只有少数题目难度值与位置不符,但变化不大,所以我们使用的测验仍保留原来顺序。

我们进行题目分析的另一方面是对被试选择答案倾向作分析,发现年龄小的儿童的

相当一部分人在选择答案时,挑与缺失部分相临的图案一样的图案为答案,而不能发现系列变化规律,一些成人被试的错误常常由于推理错误造成,尽管他们选错了,但仍然能说出他们自以为是的道理。

#### 2. 各年龄组平均成绩与男女差异问题

我们除根据各年龄组被试分数分布计算出常模表外,还统计了各年龄组的平均数与标准差,了解该测验反映的所测特质发展趋势,对各年龄组分男女统计,比较男女差

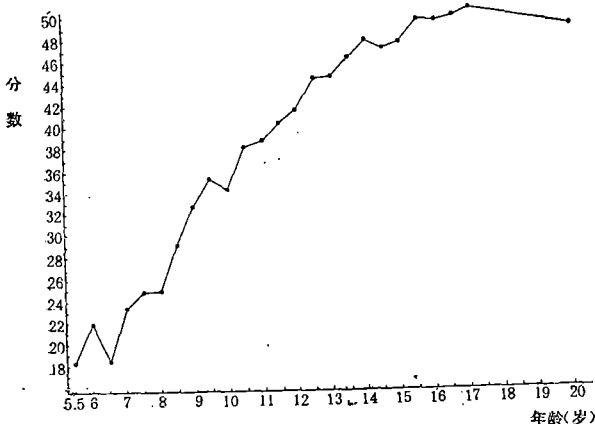


图2 各年龄组的平均结果

异。结果见表 2 和图 2。

从表 2 可见, 除 6 岁组被试平均分明显高于 6 岁半组外, 5 岁半至 17 岁 24 个年龄组被试平均分呈递减趋向。这种分数上升与下降的趋势、转折点均与国内外智力发展研究的结果接近, 尤其与美国心理学家布卢姆(B. S. Bloom)1964 年报告的智力发展速度与趋势一致。另外, 从 10 岁半组开始, 平均成绩呈较快发展至 15 岁半组成绩接近高原的状况与英国常模反映的趋势相一致<sup>[8]</sup>。上述情形在图 2 上表现得更为直观。

表2 SPM中国城市版各年龄组平均数、标准差、男女差异

AGE	M	S	D	t	P
5.5	18.23	6.55	-2.51	-2.32	<.05
6	21.89	8.94	1.04	0.80	>.05
6.5	19.70	8.50	2.86	2.17	<.05
7	23.30	10.85	2.70	1.60	>.05
7.5	24.93	10.10	0.36	0.53	>.05
8	24.93	10.44	1.53	0.90	>.05
8.5	29.29	10.54	0.72	0.42	>.05
9	32.69	10.62	2.41	1.46	>.05
9.5	35.18	9.77	-0.57	-0.37	>.05
10	34.25	11.71	-1.72	-0.92	>.05
10.5	38.13	9.57	1.58	1.07	>.05
11	38.69	9.34	-1.15	-0.82	>.05
11.5	40.25	9.43	2.62	1.76	>.05
12	41.55	9.27	1.70	1.23	>.05
12.5	44.32	7.65	1.83	1.61	>.05
13	44.44	7.62	2.51	2.07	<.05
13.5	46.13	6.26	-0.53	-0.51	>.05
14	47.69	7.28	0.36	0.32	>.05
14.5	47.07	7.90	1.36	1.14	>.05
15	47.51	7.85	0.13	0.36	>.05
15.5	49.68	6.65	0.35	0.85	>.05
16	49.46	5.94	0.10	0.10	>.05
16.5	49.92	5.92	0.23	0.26	>.05
17	50.58	6.68	-0.86	-0.90	>.05
20	48.85	8.34	-0.52	-0.54	>.05
30	46.42	9.09	-0.06	-0.04	>.05
40	45.68	9.01	3.49	2.76	<.01
50	40.39	9.85	2.32	1.53	>.05
60	37.66	10.87	2.46	1.50	>.05
70	34.55	10.89	-4.49	1.62	>.05

• 表中D、t列的负值代表女高于男, 正值代表男高于女。

为了了解不同年龄被试完成速度与答案准确性的关系, 我们在被试答题 20 分钟时要求他们标记正在完成的题目号, 以计算正确作答数与正确率, 结果如表 3。

表 3 的结果表明, 由 5 岁半至 17 岁组在 20 分钟内完成题数接近 (即操作速度接近), 而正确题数与正确率则表现为自 5 岁半至 16 岁直线上升, 保持到 16 岁半, 然后从 20 岁开始, 则无论是 20 分钟作答题数或正确率均呈下降趋势, 只是正确率的下降速度却较缓慢。

表3 瑞文标准推理测验(中国城市版) 20分钟通过率

年龄组	20分钟完成题数	20分钟得分	通过率	年龄组	20分钟完成题数	20分钟得分	通过率
5.5	45.07	15.66	.39	13	45.81	37.75	.83
6	39.86	19.95	.58	13.5	45.55	38.07	.84
6.5	44.67	20.27	.48	14	47.5	40.42	.85
7	48.64	20.36	.44	14.5	48.87	40.79	.84
7.5	43.28	22.51	.55	15	47.30	40.28	.86
8	45.97	22.67	.52	15.5	45.61	39.34	.87
8.5	45.01	26.19	.60	16	45.29	40.19	.89
9	49.41	29.93	.06	16.5	47.42	41.70	.89
9.5	46.76	30.75	.67	17	44.27	39.51	.89
10	40.47	28.37	.71	20	42.61	37.28	.88
10.5	45.27	33.23	.74	30	39.60	34.05	.87
11	45.16	33.74	.76	40	38.45	32.75	.86
11.5	44.66	34.15	.78	50	34.74	28.24	.82
12	44.36	34.94	.79	60	34.90	27.33	.79
12.5	47.40	38.12	.81	70	28.35	21.84	.79

注:完成题数和得分都是各年龄组的均值。

值得分析的是6岁组,该组平均分数21.89显著高于6岁半组与5岁半组(见表2)。从表3可见,6岁组20分钟完成题数比6岁半组和5岁半组少,通过率却远比这两个组高。可见这组儿童平均分高不是靠速度而是靠准确取得的。是该组儿童确实具备如此良好的品质还是偶然原因,这有待于进一步资料分析和个案了解。

从表2看,30个年龄组中只有40岁组男女被试成绩有显著差异( $P < .01$ ),男性组平均分非常显著地高于女性平均分数。另外5岁半组、6岁组和13岁组也有比较显著的男女差异( $P < .05$ ),其余26个年龄组的成绩没有明显男女差异。如果分析差异原因,我们认为40岁组的被试年龄包括40至49岁,在这一年龄阶段多数妇女处于更年期,男性成绩高于女性的原因很可能与女性更年期情绪不稳定有关。情绪不稳定使被试不能长时间集中精力思考,致使成绩不高,但这并不意味着女性智力不如男性。至于13岁组、5岁半组和6岁半组的男女差异可能出自随机误差。许多国家对于瑞文测验在使用中的男女分数差异进行过研究,根据从瑞文测验问世的30年代起至80年代近20个国家的70多次研究进行分析,结果发现,瑞文标准推理测验的男女成绩很少呈现规律性变化,一般表现为男女组无明显差异或在少数年龄组中可发现一定的差异。与世界研究比较,我们的结果同样显示了这种现象。

### 3. 我国常模与国外常模的比较

我们将瑞文标准推理测验的中国城市常模与香港、英国<sup>[3]</sup>和新西兰<sup>[6]</sup>近年来的常模做了比较,结果如表4。发现中国大陆青少年的发展水平较香港少年稍慢;与新西兰青少

年发展水平相当;比英国同龄组青少年发展水平高些。然而由于英国常模是于 1979 年取得的,所以在解释差异时,时间发展的因素影响,也还需要加以考虑。

表 4 中国大陆城市常模与香港和其他国家常模的比较

常模	人数	时间	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5
中国 (大陆城市)		1986	18	19	21	23	29	33	35	35	39	39	42	42	45	45	46	48	48	48	49
香港	4,500	1982	19	22	26	29	33	36	40	43	44	44	47	48	48	50	51	51	51	52	53
英国	3,464	1979	16	19	22	25	31	33	36	38	39	40	41	41	41	43	44	45	46	47	47
新西兰	3,174	1984	—	—	—	30	31	33	35	37	39	40	42	42	43	44	47	47	48	49	—

注:比较的是第50百分位值,不是平均分。

#### 4. 测验的信度与效度

①信度: 信度表示测验结果的一致性或稳定性。单份测验的分半信度与复测信度通常作为测验信度的代表。我们在小团体范围内搜集了这两类指标,取得了瑞文标准推理测验中国城市版的分半与复测信度。结果见表 5。

表 5 中的分半信度是从全国样本中随机抽取 158 份答卷计算而得,对奇偶两半相关用斯皮尔曼——布朗公式做了校正。

复测信度分别从北京一所普通小学与普通中学取得,间隔 15 天的信度资料从小学(11—13 岁)儿童中测得。由于我们所用样本很小,得到这样的结果已比较令人满意了。

表 5 瑞文标准推理测验中国城市版信度测定

信度类型	人数 N	间隔时间	$r_{ss}$	P
分半	158		0.95	<.01
复测	35	15 天	0.82	<.01
复测	52	30 天	0.79	<.01

②效度: 效度是对测验准确性的衡量,它表示测验对想测的目标测得有多么好。效度资料的取得也是在小样本中进行的。表 6 显示了我们取得的一份同时效度资料 and 一份预测效度资料。

表 6 中同时效度资料是在北京一所普通中学的 45 名 12 至 15 岁学生中测得。我们对这群学生同时施测瑞文标准推理测验和韦氏儿童智力量表。计算瑞文测验成绩与韦氏儿童智力量表的言语智商,操作智商和总智商等结果之相关。

预测效度是对北京市二所中学高三的 69 名学生施测瑞文标准推理测验,并搜集他们

表 6 瑞文标准推理测验中国城市版效度

效度类型	人数 N	$r_{xy}$	$r_{xy}$	$r_{xy}$
同时效度	45	与 WISC-CR 言语智商 0.54(P<.01)	与 WISC-CR 操作智商 0.70(P<.01)	与 WISC-CR 总智商 0.71(P<.01)
预测效度	69	与高考语文分 0.29(P<.02)	与高考数学分 0.54(P<.01)	与高考总分 0.45(P<.01)

3个月后的大学入学考试成绩,计算瑞文测验成绩与高考语文、数学和总分的相关。

由于搜集数据和个别施测韦氏量表的困难,我们的效度样本也偏小。可喜的是无论是信度与效度资料均达到或超过国外同类研究的相关系数值。说明该测验在中国城市人口中使用是可靠的,有较高的准确性。

#### 四、小 结

瑞文测验是省时、省力又可减少文化差异影响的团体智力测验。瑞文标准推理测验在中国城市的修订中得到较高的信度和效度。该常模的建立将为心理学、教育学界的研究与社会实践工作提供方便。从实验结果看,该测验可以较好地鉴别智力发展水平。一般说来,常模团体成绩没有显著的男女差异。

#### 参 考 文 献

- [1] Raven, J. C. (1941). Standardisation of Progressive Matrices, Brit. J. Med. Psychol., 19, 137-50.
- [2] Raven J. (1987). The Raven Progressive Matrices: US Norms and Their Ethnic and Socio-economic Variation in a Cross-cultural Context (in press).
- [3] Holmes, B. J. (1983). British Columbia Norms for Wechsler Intelligence Scale for Children-Revised, Peabody Picture Vocabulary Test, Slosson Intelligence Test, Standard Progressive Matrices, and Mill Hill Vocabulary Scale. University of British Columbia, Faculty of Education.
- [4] 张厚粲等,瑞文标准推理测验中国城市修订版手册。
- [5] Court J. H. Perspectives, Sex Differences in Performance on Raven's Progressive Matrices: A Review, The Alberta Journal of Educational Research 1983, Vol 29, No. 1, 8.
- [6] New Zealand Council for Educational Research, (1984). Standard Progressive Matrices: New Zealand Norms Supplement. Wellington, New Zealand: NZCER.
- [7] Raven, J. C., Court, J. H. & Raven, J. (1986). Manual for Raven's progressive Matrices and Vocabulary Scales, General Overview. London: H. K. Lewis & Co. LTD.



## STANDARDIZATION RESEARCH ON RAVEN'S STANDARD PROGRESSIVE MATRICES IN CHINA

Zhang Hou-can      Wang Xiac-ping  
*Beijing Normal University*

### Abstract

Raven's Standard Progressive Matrices (R. SPM) was revised by a research group collaborated with people came from 17 places all over the country. Based upon the data of 1982 census, a group of 5108 people aged from 5-to over 70 was sampled for establishing the norm. Through a series of standardized procedures, the Chinese norm (city version) of R. SPM was accomplished with appropriate reliability and validity. It meets the need of group test for measuring intelligence. The result shows that there is no significant sex difference except the group aged 40-49. The Chinese norm was also compared with those came from H. K., U. K. and New Zealand.