

斯—欧氏非言语智力测验的修订研究^①

张厚粲 龚文 孙燕青 田效勋

(北京师范大学心理学系)

摘要 斯—欧氏非言语智力测验是 Snijders-Oomen 1965 年发表的智力测验。它具有施测过程不受言语知识的限制等优越性。我们于 1995 年着手斯—欧氏非言语智力测验在中国的修订研究。在研究中,我们选取了 6 岁组、11 岁组、14 岁组三个年龄段来自北京、天津、山东和河北的被试进行预测。施测总人数为 334 人。通过对 SON-R 测验在我国施测所得数据资料的分析,证明该测验具有较好的内部结构和效标关联效度。同时,我们也发现少量题目由于文化背景的影响,不适于中国的儿童。并对这些题目进行了必要的更换或调整。

关键词: SON 测验 智力测量 修订研究

1 前言

目前,在我国较为常见的智力测验主要有韦氏儿童测验、韦氏成人测验、瑞文推理测验等。因其各自的优点和教育实践的需要,这些测验在我国应用日益广泛^[1]。但随着时代的发展,这些测验已经不能满足日益增长的社会需要。又随着对版权保护的日益重视,未经版权拥有者的许可,我们不能再继续修订国外的心理测验。这就要求我们应加强与国外测量学家的合作。此时,荷兰的心理学家 Laros, J. A. 和 Tellegen, P. J. 与我们联系,并希望在中国介绍“SON-R5.5—17”测验。由于其施测过程及测验结构不受语言知识的限制等优越性,我们于 1995 年 3 月着手修订 SON-R 测验中文版的工作。经过近一年的研究,我们取得了一些初步的成果。

斯—欧氏非言语智力测验^[2](The Snijders-Oomen Non-verbal Intelligence Test)最早发表于 1965 年,简称 SON。“SON-R5.5—17”是 SON 测验最新的一个修订版本,适用于 5.5 岁至 17 岁的儿童,于 1989 年完成。作为一个非言语智力测验,它包括 7 个分测验——“类别”、“积木”、“隐蔽图形”、“模式”、“情境”、“类比”、“故事”,分别考查具体推理、抽象推理、空间、知觉等四个方面的能力,从而对智力作出较为全面的评估。

在 7 个测验中,除“隐蔽图形”外,其余 6 个分测验均由 2—3 个平行系列的题目组成,每个系列中,题目由易到难排列。这种结构是与其独特的施测程序——适应性程序(adaptive procedure)相对应的。这种适应性程序不仅可以节省时间,而且可以使被试不必反复做那些对他们来说过于简单或困难的题目,避免收到过多消极的反馈,从而可以使被试在施测过程中始终维持较高的积极性。该测验规定依据被试回答问题的实际情况给予反馈。而传统智力测验一般不予反馈或只给予支持性反馈,即无论回答正确与否,都给予“很好”或类似的反馈。这种方式显得很 unnatural,被试往往会认为实验者对他们的回答漠不关心,或认为他们的结果是对的,因此当他们没有理解任务要求或使用了错误的策略时,不能及时发现并纠正错误或变换解决问题的策略。SON-R 测验中给予反馈

① 本课题得到国家教委八五重点博士点基金的资助。与此文有关的交流请与张厚粲(100875 北京师范大学心理学系)联系。

的方式,可以避免上述误导被试等问题。

最重要的是,由于该测验是非言语测验,在施测过程中,不受语言能力的束缚;测验内容也不必经过翻译;较少地受到言语知识以及学校知识的影响;也适用于聋哑儿童及其它语言交流存在障碍的情况。它不仅可以作为测量儿童智力的理想工具,也可以用来进行不同民族、种族、宗教信仰等等之间的跨文化研究。

从我国的国情考虑,大力推广和使用 SON-R 测验更具特殊意义。首先,我国城乡之间的经济水平、教育程度等方面差异显著,城市儿童和农村儿童有着不同的知识背景。一般的言语智力测验中包含了大量与学校知识相关的题目,从而难以保证城乡之间的公平性。而 SON-R 测验以操作性题目为主,不受学校知识及语言知识的影响,可以较好地保证公平性。其次,我国幅员辽阔,民族众多,各地区、各民族之间语言差异很大,为施测带来困难。使用 SON-R 测验可以解决这个问题。而且该测验对于聋哑儿童同样适用,扩大了智力测验使用的范围,也是它的一大特点。此外,SON-R 测验在内容、施测过程、解释结果等方面都有它的独到之处,对于我国智力测验的研究工作具有借鉴意义。

2 研究方法和结果

2.1 被试取样

SON 测验的适用对象为 5.5~17 岁的儿童。在修订工作中,我们选取 6 岁组(6 岁 0 个月—12 个月)、11 岁组(11 岁 0 个月—12 个月)、14 岁组(14 岁 0 个月—12 个月)3 个年龄段的被试进行预测,被试分别取自北京、密云、天津、山东广饶县、河北任丘等大、中、小城市。6 岁组共 114 人,男 52 人,女 62 人;11 岁组共 115 人,男 64 人,女 51 人;14 岁组共 105 人,男 56 人,女 49 人。施测总人数为 334 人,男 172 人,女 162 人。

2.2 施测过程

主试均经过严格培训。施测程序有 2 种:1. 按照原测验中规定的适应性程序施测,共有被试 263 人;2. 为分析测验题目,令被试逐一完成各个题目,共有被试 71 人。

2.3 测验结果及分析

2.3.1 难度分析

按 3 个年龄段对测验题目进行难度分析,对于适应性程序中未完成的题目按缺失值处理。表 1 至表 7 为分测验的难度结果。

表 1 类别测验的难度分析

系列	6 岁组			11 岁组			14 岁组		
	a	b	c	a	b	c	a	b	c
1	0.85	0.74	0.69	1.00	0.88	0.82	1.00	0.56	0.85
2	0.39	0.49	0.44	0.96	0.81	0.83	0.96	0.72	0.74
3	0.29	0.34	0.34	0.69	0.64	0.49	0.82	0.70	0.42
4	0.21	0.37	0.15	0.21	0.64	0.65	0.28	0.83	0.79
5	0.25	0.00	0.24	0.51	0.64	0.42	0.53	0.78	0.47
6	0.16	0.15	0.44	0.64	0.53	0.80	0.77	0.65	0.81
7	0.00	0.00	0.08	0.13	0.66	0.37	0.30	0.86	0.61
8	0.00	0.09	0.00	0.35	0.25	0.24	0.59	0.46	0.63
9	0.00	0.18	0.09	0.08	0.23	0.37	0.27	0.50	0.59

由以上分析可以看出,除“隐蔽图形”外,其余 6 个分测验的题目在排列上基本符合由易到难排列的构想。这也证明了其适应性程序的合理性。由于中外文化背景的差异,有个别题目在难度系列中的位置与我们施测的结果不一致,我们将根据施测情况,调整这些题目的排列顺序。“隐蔽图形”分测验中,被试在 4 个题目上的得分没有明显的差别,证明 4 个题目的难度相近,这也是与测验的结构构想相符的。

表 2 积木测验的难度分析

系列	6 岁组		11 岁组		14 岁组	
	a	b	a	b	a	b
1	0.97	0.96	1.00	1.00	1.00	1.00
2	0.93	0.92	0.99	1.00	0.99	0.75
3	0.52	0.71	0.96	0.98	1.00	0.90
4	0.23	0.51	0.92	0.87	0.96	0.86
5	0.14	0.24	0.78	0.86	0.87	0.80
6	0.06	0.22	0.82	0.87	0.85	0.80
7	0.07	0.00	0.60	0.84	0.85	0.91
8	0.00	0.00	0.33	0.39	0.72	0.59
9	0.00	0.00	0.24	0.51	0.56	0.70
10	0.00	0.00	0.53	0.40	0.84	0.73

表 3 模式测验的难度分析

系列	6 岁组		11 岁组		14 岁组	
	a	b	a	b	a	b
1	0.91	0.88	1.00	1.00	1.00	1.00
2	0.72	0.75	1.00	1.00	1.00	1.00
3	0.35	0.46	0.97	0.98	0.97	1.00
4	0.35	0.31	0.94	0.95	0.97	1.00
5	0.20	0.22	0.79	0.89	0.81	0.74
6	0.03	0.28	0.68	0.85	0.83	0.87
7	0.06	0.05	0.64	0.53	0.88	0.77
8	0.00	0.00	0.05	0.13	0.17	0.16
9	0.00	0.00	0.06	0.15	0.09	0.21

表 4 隐蔽图形测验的难度分析

	6 岁组	11 岁组	14 岁组
HID1	0.174	0.496	0.594
HID2	0.229	0.561	0.69
HID3	0.273	0.546	0.643
HID4	0.219	0.552	0.641

表 5 情境测验的难度分析

系列	6 岁组			11 岁组			14 岁组		
	a	b	c	a	b	c	a	b	c
1	0.90	0.59	0.33	1.00	0.95	0.92	1.00	0.80	1.00
2	0.63	0.55	0.53	0.95	0.82	0.92	0.97	0.57	1.00
3	0.44	0.60	0.62	0.83	0.95	0.81	0.89	0.89	0.14
4	0.37	0.27	0.45	0.85	0.92	0.91	0.89	0.92	0.67
5	0.22	0.17	0.28	0.61	0.83	0.92	0.78	0.73	0.83
6	0.21	0.00	0.26	0.62	0.62	0.62	0.76	0.68	0.75
7	0.28	0.08	0.06	0.57	0.71	0.43	0.62	0.78	0.69
8	0.15	0.00	0.00	0.46	0.48	0.41	0.54	0.49	0.57
9	0.09	0.18	0.00	0.44	0.45	0.64	0.68	0.55	0.75
10	0.00	0.00	0.09	0.23	0.11	0.27	0.48	0.19	0.46
11	0.00	0.00	0.00	0.07	0.09	0.13	0.40	0.27	0.42

表 6 类比测验的难度分析

系列	6 岁组			11 岁组			14 岁组		
	a	b	c	a	b	c	a	b	c
1	0.84	0.77	0.61	0.99	1.00	0.97	1.00	1.00	1.00
2	0.59	0.61	0.66	0.96	1.00	1.00	0.98	1.00	0.50
3	0.45	0.55	0.56	0.99	1.00	1.00	0.95	0.80	1.00
4	0.65	0.43	0.50	0.96	0.94	0.91	0.94	0.89	0.83
5	0.32	0.31	0.32	0.95	0.86	0.91	0.91	0.75	0.82
6	0.44	0.48	0.30	0.73	0.69	0.81	0.85	0.82	0.68
7	0.27	0.18	0.35	0.73	0.75	0.76	0.70	0.83	0.84
8	0.10	0.35	0.42	0.56	0.67	0.66	0.71	0.66	0.75
9	0.17	0.24	0.15	0.51	0.38	0.58	0.68	0.55	0.73
10	0.00	0.00	0.25	0.36	0.38	0.58	0.66	0.60	0.66
11	0.00	0.18	0.08	0.26	0.51	0.58	0.54	0.62	0.58

表 7 故事测验的难度分析

系列	6 岁组		11 岁组		14 岁组	
	a	b	a	b	a	b
1	0.73	0.67	0.97	0.96	0.96	0.82
2	0.35	0.40	0.96	0.93	0.95	0.80
3	0.48	0.44	0.72	0.91	0.82	0.93
4	0.26	0.24	0.71	0.77	0.84	0.78
5	0.18	0.24	0.57	0.74	0.76	0.59
6	0.11	0.00	0.54	0.46	0.74	0.54
7	0.08	0.00	0.57	0.70	0.60	0.83
8	0.00	0.00	0.39	0.28	0.75	0.45
9	0.00	0.00	0.47	0.39	0.56	0.77
10	0.00	0.00	0.00	0.05	0.16	0.13

2.3.2 区分度分析

按 3 个年龄段对测验题目进行区分度分析(计算每个题目与其余题目总分的相关)。为了计算方便,将完成全部题目的数据转换为与适应性程序一致的数据(经检验,数据转换前后差异不显

著)。

3个年龄段的区分度分析结果见表8、表9、表10。

表8 6岁组区分度分析结果

	区分度 >0.30 的题目及其百分率	区分度为零的题目及其百分率
类别	41%(a1~a6b1~b4c1)	29%(a7~ab5b7~69c8)
积木	25%(a3a4b3~b5)	35%(a8~a10b7~b10)
模式	44%(a1~a4b1~b4)	11%(a8a9)
情境	21%(a1a5b2~b5c4)	24%(a8~a11b8~b11)
类比	61%(a3a4a6a9b1~b8c1~c8)	15%(a10a11b10b11c11)
故事	30%(a1a2a4b1~b3)	45%(a7~a10b6~b10)

表9 11岁组区分度分析结果

	区分度 >0.30 的题目及其百分率	区分度为零的题目及其百分率
类别	41%(a5,b1,b4~b7,c2~c6)	4%(a1)
积木	60%(a5~a10,b5~b10)	15%(a1,b1,b2)
模式	33%(a6,a7,b4~b7)	22%(a1,a2,b1,b2)
情境	42%(a4,a5,a7,b4~b9,c6,c8~c11)	6%(a1,c1)
类比	30%(a5,a6,a8,b7~b9,b11,c8~c10)	15%(b1~b3,c2,c3)
故事	45%(a6,a9,b3~b7)	5%(a10)

表10 14岁组区分度分析结果

	区分度 >0.30 的题目及其百分率	区分度为零的题目及其百分率
类别	59%(a2,a5,a6,b3,b5~b9,c3~c9)	4%(a1)
积木	60%(a5~a10,b5~b10)	15%(a1,a3,b1)
模式	44%(a5~a7,a9,b5~b7,b9)	33%(a1,a2,b1~b4)
情境	58%(a5~a11,b5~b9,c4,c5,c6,c8~c11)	9%(a1,c1,c2)
类比	82%(a2~a11,b4~b11,c2,c4~c11)	30%(a1,b1~b3,c1,c3)
故事	75%(a3,a5~a9,b1~b9)	

总起来看,除“隐蔽图形”分测验外,其余6个由平行系列构成的分测验题目的区分度变化,在3个年龄段相同的呈现相同的趋势,即对于6岁组,位于系列末端的题目区分度较低,位于系列开头的题目区分度较高;对于11岁组和14岁组,位于系列开头的题目区分度较低,位于系列中、后部的题目区分度较高;随着年龄的增长,区分度最高的题目在系列中的位置逐渐后移。同时,3个年龄段中,具有中等难度的题目区分度较高,过难或过易的题目区分度较低。“隐蔽图形”分测验中,4个题目的区分度没有明显的差别。同时,我们也发现在各个分测验中,均有个别题目的区分度不符合上述趋势;参考表1—表7,会发现其中大部分题目的难度也不符合趋势。因此可以认为,造成这些题目的区分度不符合趋势的原因,主要是由于题目的排列顺序不当或由于不同文化背景的影响而导致儿童不能正确理解测验内容。为使测验内容更适合中国儿童,我们除了更换部分图片外,将根据通过率对测验题目重新排序。

2.3.3 与荷兰各分测验平均分的比较

通过统计6岁组、11岁组、14岁组在7个分测验中的平均分、标准差,并与荷兰常模中相应年龄段进行差异显著性检验,结果见表11。

表 11. 两国儿童的比较

	6 岁组			11 岁组			14 岁		
	中国 平均	荷兰 平均	差异 检验	中国 平均	荷兰 平均	差异 检验	中国 平均	荷兰 平均	差异 检验
类别	4.98	5.7	* *	13.21	15.3	* *	16.13	18.0	* *
积木	6.08	7.5	* *	14.11	14.3		16.58	16.5	
隐蔽图形	6.69	6.7		16.12	16.0		19.26	19.6	* *
模式	4.83	3.7	* *	12.56	11.2	* *	13.39	13.3	
情境	7.20	9.2	* *	19.43	20.5	* *	22.23	23.7	* *
类比	9.01	6.8	* *	23.68	20.0	* *	26.26	23.7	
故事	3.40	4.0	* *	10.62	11.2	* *	12.87	13.6	* *

(中国 6 岁组被试 89 名,11 岁组被试 102 名,14 岁组被试 104 名;荷兰各组被试均为 150 名。* 代表在 0.05 水平显著,* * 代表在 0.01 水平显著。)

由上表我们可以看出,在 3 个年龄段的“类别”、“情境”、“故事”3 个分测验中,荷兰的平均分高于中国的平均分,并都达到 0.01 的显著水平;“模式”、“类比”两个分测验中,中国的平均分显著高于荷兰的平均分,并都达到 0.01 的显著水平;“隐蔽图形”分测验中,中国与荷兰的平均分无显著差异。“积木”分测验在 6 岁组、11 岁组,荷兰的平均分高于中国的平均分,分别达到 0.01 和 0.05 的显著水平,14 岁组,荷兰与中国的平均分无显著差异。

在“类别”、“情境”、“故事”的 3 个分测验中,中国儿童得分普遍偏低。根据施测过程中观察到的被试的反应推测,这种情况很可能是由于这 3 个分测验本身的内容特点造成的:这 3 个分测验中使用的图片内容涉及生活、文化等领域,中国儿童对图片内容不熟悉,故出现较多的错误反应。为此,我们便就较易引起歧义或儿童不认识的图片进行调整,并征得测验原作者的同意,可以在以后的施测中更换。经检验,通过调整题目顺序及更换部分图片,该测验更适于在中国应用。

在“模式”、“类比”分测验中,中国儿童得分普遍较高。这两个分测验的测验材料均为几何图形,受文化背景的影响较小,中国儿童能够较好地理解测验要求。相关的研究^[3]也发现中国青少年的推理能力发展水平高于美国同龄组的青少年。

2.3.4 效度分析

由于瑞文测验是国内外公认的较为成熟的测验,并且它与 SON-R 测验都是非文字测验,我们分别对 11 岁组和 14 岁组进行 SON-R 测验与标准瑞文推理测验的相关分析。在两个年龄段,SON-R 测验总分与瑞文测验分相关显著(0.01 水平),SON-R 中的抽象推理测验总分、具体推理测验总分分别与瑞文测验相关显著(0.01 或 0.05 水平)。

由于智力因素是影响学习成绩的一个重要方面,我们通过考查 11 岁组和 14 岁组在 SON-R 测验中的得分与最近一次期中考试成绩的相关,发现 11 岁组的 SON-R 测验总分与数学、语文及其总分相关显著(在 0.01 水平);14 岁组的 SON-R 测验总分与语文、英语、代数、几何、物理等成绩的 z 分数之和相关显著(在 0.05 水平),与几何、物理成绩相关显著(在 0.05 水平)。

从以上结果,可以认为 SON-R 测验具有较好的效度。

2.3.5 3 个年龄段的平均分比较

我们对 6 岁组、11 岁组、14 岁组的被试在 SON-R 测验中各分测验中的平均分进行差异显著性检验,检验结果见表 12、表 13。

表 12 平均分差异显著性检验(6 岁组和 11 岁组)

	6 岁组(N=89)		11 岁组(N=102)		T 检验
	平均	标准差	平均	标准差	
类别	4.89	2.94	13.21	5.07	47.15***
积木	6.08	2.04	14.11	3.32	103.73***
隐蔽图形	6.69	2.98	16.12	3.98	73.94***
模式	4.83	2.82	12.56	2.07	117.69***
情境	7.20	3.51	19.43	4.93	64.93***
类比	9.01	5.39	23.68	4.43	55.25***
故事	3.40	2.07	10.62	3.42	88.69**

表 13 平均分差异显著性检验(11 岁和 14 岁组)

	11 岁组(N=102)		14 岁组(N=104)		T 检验
	平均	标准差	平均	标准差	
类别	13.21	5.07	16.13	5.88	3.82***
积木	14.11	3.32	16.58	3.65	5.08***
隐蔽图形	16.12	3.98	19.26	4.77	5.13***
模式	12.56	2.07	13.39	2.13	2.84***
情境	19.43	4.93	22.23	5.48	3.86***
类比	23.68	4.43	26.26	4.75	4.03***
故事	10.62	3.42	12.87	4.34	4.14***

由表 12 和表 13 可以看出,3 个年龄段的被试在 SON-R 测验的各分测验的平均分有显著差异,被试的得分随着年龄的增长而上升。这与儿童的发展趋势是一致的。

3 结论

总起来看,与目前常用的智力测验相比,SON-R 测验在内容、结构、施测程序、解释测验结果等方面突破了以往的限制,能够从多个维度较为全面地评估智力。测验内容新颖、丰富,并且在施测过程中注重激发儿童的兴趣,因此尽管施测时间较长,但儿童始终能够维持较高的积极性。最重要的是,该测验是非言语测验,施测过程不受言语知识、技能的束缚,因此不仅适用于正常儿童,而且适用于聋哑儿童及其它有言语交流障碍、情绪障碍的儿童,应用范围 较广。并且由于测验不包含与学科教学相关的题目,故可以在不同的经济水平、教育程度的地区间保证跨文化公平性。总之,SON-R 测验是一个适合我国国情的比较理想的智力测量工具。

通过对 SON-R 测验在我国施测所得数据资料的分析,证明该测验具有较好的结构和效度。同时,我们也发现少量题目由于文化背景的影响,不适于中国的儿童或在难度系列中的位置与施测情况不符,因此对这些题目进行了更换或调整。经检验,修订后的测验更适于在中国施测。

4 参考文献:

1 张厚粲,龚文. 当前心理测量学的发展和现状. 心理学探新,1995;15(1)

2 Laros J A,Tellegen P J. Construction and validation of the SON-R 5 1/2-17,the Snijders-Oomen non-verbal intelligence test. Woltes-Noordhoffbv Groningen,The Netherlands,1991

3 王晓平,张厚粲,林传鼎. 韦氏儿童智力量表中美两种版本的因素分析:两国样本的年龄组间比较. 心理科学,1993;17(5):11-16

ENGLISH ABSTRACTS

REVISING SNIJDERS-OOMEN NONVERBAL INTELLIGENCE TEST-REVISION IN CHINA

Zhang Houcan, Gong Wen, Sun Yanqing, Tian Xiaoxun
(Department of Psychology, Beijing Normal University)

Developed by Dr. Snijders-Oomen in 1965, Snijders-Oomen Nonverbal Intelligence test, is widely used in Europe, especially for exceptional children. Because of its advantages, we conducted the revised SON in China. In our case, 334 subjects aged 6, 11 and 14 were selected from northern China. Besides the item analysis, we conducted a criterion-relative validity analysis, and got satisfied results. Although a few items of SON-r are not suitable for Chinese children, we think the SON is quite useful to educators and other test-users.

Key words: SON, intelligent measurement, test revising.

A MULTI-METHOD COMPLEX ANALYSIS OF THE VALUE ORIENTATIONS AMONG URBAN AND SUBURBAN TEENAGERS

Wang Zhongming, Lu Xinhai
(Psychology Department, Hangzhou University)

With the new multi-method and multi trait approach, this study adopted multiple measures to analyze the value orientations of collectivism and individualism among teenagers under different cultural settings. The study attempted to achieve some progress in both theoretical constructs and research methodology on value orientation. The results showed that 1) subjects of different age groups and from different regions had significantly different value orientations of collectivism and individualism; 2) there were significant different correlations among the four methods (measures), indicating the feasibility of multi-method multi-trait analysis; 3) the comparison between the Chinese and overseas' results of value orientations of collectivism and individualism revealed different pat-

terns. The present study provided useful evidence and theoretical/methodological implications for further research.

Key words: value orientation, teenager.

HOW MANUSCRIPTS ARE HANDLED BY ACTS PSYCHOLOGICAL SINICA

Wei Mingxiang, Zhuang Jingchun
(Institute of Psychology, Academia Sinica, Beijing)

The article introduces how manuscripts are selected, examined, approved and typeset, and what requirements are made by the editorial board. It also analyzes some common errors in the manuscripts.

Key words: manuscript, Chinese expression, document.

SOURCES OF AGE DIFFERENCES IN INFORMATION PROCESSING STAGES

Wo Jianzhong, Shen Jiliang, Lin Chongde
(Institute of Developmental Psychology, Beijing Normal University, Beijing)

140 people aged 7-19 were tested in four different tasks. Except for the 7-year-olds, the spatial regulating speed of the same figures was found to be slightly different; the regulating speed of different figures, the translating speed and the translating diversity of different letters in people aged 11 and above were basically constant. The other processing stages speeded gradually with age. The age-related variable characteristics in the altering, reacting stages were well described by exponential functions, but not those in the letter matching stages.

Key words: information processing, age, reaction time.

SELF-RELATIVE CHINESE CHARACTERS: A STUDY OF IMPLICIT MEMORY

Gao Xiangping
(Department of Psychology, East China Normal University)

Han Zhengjue