

SON - R 6 - 40 非言语智力测验在中国的适用性

徐建平^{1,2}, 郭枝^{1,3}, 王茜^{1,4}, Peter Tellegen⁵

(1. 北京师范大学心理学院, 北京 100875; 2. 北京市应用实验心理重点实验室, 北京 100875;

3. 清华大学人因与工效学研究所, 北京 100084; 4. 锦程教育集团, 北京 100080;

5. Department of Psychometrics & Statistical Techniques, University of Groningen, Groningen, The Netherlands 9700 AB)

摘要: SON - R 6 - 40 非言语智力测验是以非文字形式测量个体推理与空间能力的智力评估工具。研究旨在考查 SON - R 6 - 40 在中国的适用性。根据全国人口分布, 在中国六大行政区 6 至 40 岁年龄人群中施测了 1721 人, 通过与荷兰/德国混合样本比较, 分析该测验的跨文化等值水平。结果表明: (1) 测验信度与结构跨样本一致, 说明该测验在两种文化下结构等值; (2) 多样本验证性因素分析证明了该测验的测量单位等值性。因此, SON - R 6 - 40 适合在中国推广使用以及用于智力的跨文化比较研究。

关键词: SON - R 6 - 40; 非言语智力测验; 跨文化等值

中图分类号: B841.2

文献标识码: A

文章编号: 1003 - 5184(2014)01 - 0036 - 07

1 问题提出

斯一欧非言语智力测验 (the Snijders - Oomen Non - verbal Intelligence Test, SON) 是以非文字形式测量个体推理与空间能力等流体智力的评估工具, 起源于荷兰心理学家 Nan Snijders - Oomen 对 4 ~ 14 岁聋童学习能力的评估研究 (Snijders - Oomen, 1943), 随后多次修订 (Snijders & Snijders - Oomen, 1970; Snijders, Tellegen & Laros, 1989; Starren, 1978; Laros & Tellegen, 1991)。SON - R 6 - 40 非言语智力测验 (SON - R 6 - 40 Non - verbal Intelligence Test, SON - R 6 - 40) 作为 SON 测验最新版, 由 Tellegen 与 Laros 编制, 适用于 6 岁至 40 岁的儿童和成人, 于 2011 年在荷兰与德国正式出版 (Tellegen & Laros, 2011), 同时也在英国、巴西及泰国的标准化研究也逐步展开。

作为一种非言语智力测验, SON - R 6 - 40 除了具有非文字形式施测这一优势之外, 它的优势还体现在测量内容全面, 采用了适应性施测程序, 施测过程中主试能够向被试提供结果反馈, 指导语采用了言语和非言语两种相结合的形式等方面 (徐建平等, 2011)。它们保证该测验能够适用于不同的人群, 如正常人群、聋哑群体、听障群体、言语障碍群体、学习障碍群体、智障群体及少数民族群体等 (Tellegen, Winkel, Wijnberg - Williams, & Laros, 1998; Tellegen & Laros, 2011)。此外, 该测验利用精确年龄基于统计拟合程序建立统一而连续的常模 (Tellegen & Laros, 2011), 结果解释更加精确合理, 并且有效利用了非言语智力测验的特点, 给特殊群

体提供了与正常群体平等的评估机会, 与全纳教育思想一脉相承 (徐建平等, 2011)。

因此, 将该测验引入中国具有重要意义。一方面, 中国人口构成复杂多样, 各地区各民族语言与文化差异较大, 也存在各种障碍人群, 使用 SON - R 6 - 40 可较好地保证测验公平性, 并为不同群体的教育与职业发展等提供参考信息与适宜帮助。另一方面, 引进由欧洲心理学家编制的 SON - R 6 - 40 测验, 有助于国内研究者全面了解和掌握国际当代智力测验发展趋势和新技术。

有研究表明, 文化距离常常是跨文化研究中偏差的重要来源之一 (Van de Vijver & Tanzer, 2004), 因此考查某个测量工具在新文化情境下的跨文化等值水平是该测验适用于这种文化情境的前提。跨文化等值包含结构等值、测量单位等值及全量表等值三个水平, 三者并非完全孤立存在, 前一水平往往是后一水平成立的先决条件 (Van de Vijver & Tanzer, 2004)。结构等值是指在不同文化群体中, 无论测量工具是否一致, 但测量结构相同。测量单位等值是指在不同文化群体中测量工具的原点不同, 但测量单位一样。全量表等值是跨文化等值的最高水平, 指在不同文化群体中测量工具的原点与测量单位都相同。不同的等值水平决定了测验的跨文化适用性以及跨文化比较研究的层次。因此, 在中国修订与推广使用 SON - R 6 - 40 非言语智力测验前, 考查该测验在中国文化情境下的跨文化等值水平十分必要。

2 研究方法

2.1 被试

被试由中国样本与荷兰/德国样本两部分组成。

中国样本按地区、教育水平、年龄、性别分层取样。从东北、华北、华东、中南、西南与西北六大行政区,每个行政区选择两至三个省,抽选6岁至40岁各年龄段不同教育水平的被试,并平衡被试的社会经济地位水平,最终入选样本共1721人。其中男性占51.3%,少数民族占4.0%,特殊群体(包括听障和智障)占3.8%。样本中性别、各地区人数、各类教育水平人数的比例与第六次全国人口普查数据对应类别人员所占的比例相吻合。

荷兰/德国样本为荷兰标准化研究的常模样本,共1933人。同样是采取上述分层取样方法分别在荷兰与德国选取了6岁至40岁938人、995人。由于两国被试整体表现的分布一致,因此将荷兰与德国被试合为一个样本,其中男性占49.8%,移民群体占19.0%,特殊群体占5.0%。研究表明,该样本具有代表性(Tellegen & Laros 2011)。

2.2 测量工具

SON-R 6-40 非言语智力测验。鉴于测验材料为图形和拼板,在不同的国家,使用了统一的测量工具,施测材料未做任何改变。该测验包含类比、拼板、归类及连线四个分测验,每个分测验包含两个或三个系列,系列内题目均由易到难排列。类比(Analogies)分测验要求被试根据发生改变的几何图形识别转换规律,并将其应用到另一个几何图形;拼板(Mosaics)分测验要求被试利用红白色正方形拼板在一个外围框内拼出图册中呈现的图形;归类(Categories)分测验要求被试根据所给图画共同属性,从另外候选的五幅图画中选出也包含这一相同属性的两幅图画;连线(Patterns)分测验要求被试根据所给空缺图案中线条的变化规律,用铅笔填补线条的空缺部分使其成为一个连续整体。其中类比与归类分测验采用选择题型,评估个体的推理能力,属于推

理分量表;拼板和连线通过操作评估个体的空间洞察力,属于操作或空间分量表。完成整个测验需要大约1至1.5小时。荷兰标准化研究表明,该测验与韦氏儿童智力量表第四版(WISC-IV)相关为0.77,与韦氏成人智力量表第三版(WAIS-III)相关为0.83(Tellegen & Laros 2011)。

2.3 研究程序

研究分五阶段。第一阶段,组织翻译中文版测验施测手册,并进行交叉校核;第二阶段,集中培训主试;第三阶段,根据抽样分布,进行一对一施测,完成中国各行政区取样和样本数据采集;第四阶段,汇集数据并对其有效性进行核查,确定最终用于统计分析的样本数据;第五阶段,采用SPSS 16.0及AMOS 17.0比较分析处理中国样本与荷兰/德国样本数据。其中荷兰与德国样本数据由合作方提供。

3 结果与分析

3.1 结构等值

考查SON-R 6-40非言语智力测验在中国与荷兰/德国样本之间的结构等值性前,首先比较了两样本中测验的测验信度(Hult et al. 2008)。采用内部一致性系数计算方法,分别从项目水平和分测验水平分析两样本中测验的信度,即Cronbach α 系数与概化系数(Tellegen & Laros 2011)。各分测验与总量表的 α 系数和概化系数结果见表1。一般认为,内部一致性系数大于0.60为可接受水平,0.70以上较好(凌文轻,方俐洛 2003)。由表1可知,两样本中各分测验与总量表 α 系数和概化系数都远大于0.70,表明该测验在中国的信度与在荷兰和德国同样良好。按照Hakstian和Whalen(1976)提出检验群体间测验内部一致性系数差异的程序,进一步考查该测验信度的跨文化一致性。结果见表1,除连线分测验外,各分测验及总量表的信度在两样本间的差异均在0.01水平上不显著。因此可以说,该测验的信度在中国与荷兰/德国两样本间基本具有跨文化一致性。

表1 中国与荷兰/德国样本中的信度

	Cronbach α 系数				概化系数	
	类比	拼板	归类	连线	总量表	总量表
中国样本(N=1721)	0.88	0.85	0.89	0.87	0.95	0.84
荷兰/德国样本(N=1933)	0.87	0.85	0.88	0.89	0.95	0.84
信度差异检验统计值(M)	4.23	0.44	4.57	15.58**	0.31	0.02

注: ** $p < 0.01$ 。

采用探索性因素分析方法,基于标准化后的分测验分数,进一步精确分析SON-R 6-40非言语智力测验在中国与荷兰/德国样本间的结构等值性

(Matsumoto & Van de Vijver 2011)。两样本及合并样本中各分测验间相关系数见表2,所有相关系数都显著,因此可确定适合进行因素分析。

表2 两样本及合并样本中各分测验间的相关系数

	中国($N = 1721$)			荷兰/德国($N = 1933$)			合并样本($N = 3654$)		
	类比	拼板	归类	类比	拼板	归类	类比	拼板	归类
拼板	0.63**			0.59**			0.59**		
归类	0.59**	0.53**		0.53**	0.48**		0.52**	0.53**	
连线	0.64**	0.65**	0.57**	0.59**	0.67**	0.51**	0.61**	0.65**	0.53**

注: ** $p < 0.01$ 。

探索性因素分析适用性的检验结果显示,两样本及合并样本中 KMO 值均大于 0.80 ($KMO_{\text{中国}} = 0.82$, $KMO_{\text{荷兰/德国}} = 0.80$, $KMO_{\text{合并}} = 0.81$),且 Bartlett 球形检验结果均显著,表明两样本及合并样本都适合进行探索性因素分析。首先采用主成分分析法,对合并样本进行因素分析,确定因子数。提取特征根大于 1 的因子,发现只能提取一个公因子,累积解释率为 67.85%。然后分别对两个样本强制提取一个因子进行探索性因素分析。结果发现各分测验的载荷在两样本间仅有微小差异(见表 3)。

表3 两样本单因素模型中各分测验的载荷

	中国	荷兰/德国
类比	0.84	0.83
拼板	0.84	0.84
归类	0.78	0.76
连线	0.84	0.85

为进一步验证两样本中提取的公因子确实一致,需要计算 Tucker's phi 系数。结果显示,phi 系数达到 0.99,大于 0.95,表明两样本中所提取的因子具有高度一致性(Welkenhuysen - Gybels & Van de Vijver, 2001)。因此,验证了 SON-R 6-40 非言语智力测验在中国与荷兰/德国样本间的结构等值性。

3.2 测量单位等值

测验的跨文化结构等值性得到验证后,进一步探讨更高层次的跨文化等值性——测量单位等值。

中国样本中各分测验及总量分数的描述统计结果见表 4,表明各分测验及总量表标准化后的分数均为正态分布。因此采用多样本验证性因素分析的方法(Myers, Calantone, Page, & Taylor, 2000; Wang, Bristol, Mowen & Chakraborty, 2000),对测验结构(如图 1)进行跨样本模型比较。四个待检验和比较的模型如下:(1)模型 1 为非限制模型,允许因子载荷、结构误差(图 1 中的 e5)和测量误差(图 1 中的 e1、e2、e3、e4)在两个样本间自由估计,其他模型以此模型为基础进行比较检验;(2)模型 2 为限制因子载荷模型,固定因子载荷在两个样本间相等;(3)模型 3 为限制结构误差模型,固定因子载荷和结构误差在两个样本间相等;(4)模型 4 为限制测量误差模型,固定因子载荷、结构误差和测量误差在两个样本间相等。

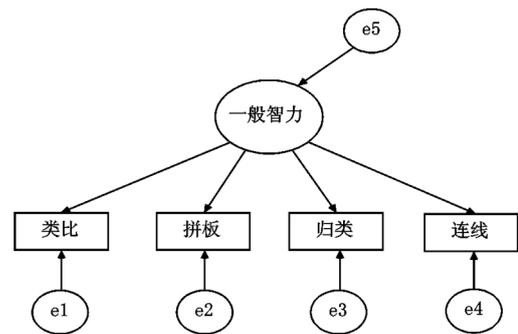


图1 SON-R 6-40 非言语智力测验的单因素结构模型

表4 中国样本中各分测验及总量分数的描述统计结果

	最低分	最高分	M	SD	偏度检验		峰度检验	
					偏度值	标准误	峰度值	标准误
类比	1	18	10.0	3.0	0.01	0.06	-0.12	0.12
拼板	1	19	10.0	3.0	-0.008	0.06	-0.10	0.12
归类	1	19	10.0	3.0	-0.003	0.06	0.01	0.12
连线	1	19	10.0	3.0	0.038	0.06	-0.06	0.12
总量表	55	145	100.0	15.0	-0.008	0.06	-0.10	0.12

考虑到样本量过大会高估 χ^2 值的显著性,因此基于模型中待估计的参数个数,在两个样本中分别随机抽取 20% 被试进行多样本验证性因素分析

($N_{\text{中国}} = 338$, $N_{\text{荷兰/德国}} = 398$; Jackson, 2001; Marsh & Hocevar, 1985) 结果见表 5 和表 6。

表5 多样本验证性因素分析模型拟合结果

模型	χ^2	df	p	χ^2/df	GFI	AGFI	CFI	RMSEA
模型1(非限制模型)	17.38	4	0.002	4.34	0.99	0.94	0.99	0.068
模型2(限制因子载荷模型)	20.74	7	0.004	2.96	0.99	0.96	0.99	0.052
模型3(限制结构误差模型)	21.44	8	0.006	2.68	0.99	0.97	0.99	0.048
模型4(限制测量误差模型)	30.21	12	0.003	2.52	0.98	0.97	0.99	0.045

表6 多样本验证性因素分析模型比较结果

	与模型1比较			与模型2比较			与模型3比较		
	χ^2_{diff}	df	p	χ^2_{diff}	df	p	χ^2_{diff}	df	p
模型2(限制因子载荷模型)	3.36	3	0.339						
模型3(限制结构误差模型)	4.07	4	0.397	0.71	1	0.400			
模型4(限制测量误差模型)	12.83	8	0.118	9.47	5	0.092	8.77	4	0.067

已有研究表明,CFI、GFI与AGFI值大于0.90, RMSEA小于0.08,模型可接受;若这些拟合指标大于0.95, RMSEA小于0.05,表明模型拟合很好; χ^2 值与df值之比小于5表示模型可以接受,小于3模型比较好(温忠麟,侯杰泰,马什赫伯特,2004)。由表5可知,模型1中 χ^2 值显著,但其他不受或较少受样本量影响的模型拟合指标良好,说明测验单因素结构模型在两个样本中都得到了验证。各模型间两两比较的 χ^2 差异值均不显著(见表6),说明各模型对数据的拟合情况相同,因子载荷、结构误差和测量误差在两样本间相同。重新在两个样本中分别随机抽取20%的被试再次进行分析($N_{\text{中国}}=358$, $N_{\text{荷兰/德国}}=375$),得到了相似的结果。因此,SON-R 6-40非言语智力测验在中国与荷兰/德国样本间等值性达到了测量单位等值的水平。

3.3 全量表等值

测验的跨文化测量单位等值性得到验证之后,继续探讨最高水平的跨文化等值性——全量表等值。Cheung(2008)指出,跨文化研究中被试能力相同时,若一个样本中的被试在某些测验项目上的得分系统地高于或低于另一样本中的被试时,该测验则不具有全量表等值性。因此,通过分析SON-R 6-40非言语智力测验的项目难度,初步探讨该测验的跨文化全量表等值性。如果项目难度在两样本间完全一致,可充分说明该测验达到了全量表等值水平。

由于该测验按适应性程序施测,因此每个题目序列中起测题前面跳过的题目得分均记为1,结束

题后面未做的题目得分均记为0,然后通过计算通过率分析测验项目难度。分析发现中国样本与荷兰/德国样本一样,各个分测验中每一系列项目在总体上符合由易到难的排列原则,几个系列之间后一系列总是比前一系列容易,符合适应性施测程序要求。具体而言,中国样本中,124题中只有9题出现了难度反转,即该项目比同一系列中紧邻的后一项目的难度p值小,且大部分题目难度反转的差异约为0.02,属于可接受范围(Tellegen & Laros, 2011)。只有归类分测验a系列第6题(6a)与c系列第4题(4c)差异较大,分别为0.11和0.05。荷兰/德国样本中出现难度反转的题目有7题,其中5题与中国样本一致,包括归类分测验的6a与4c,而差异均在0.02左右。

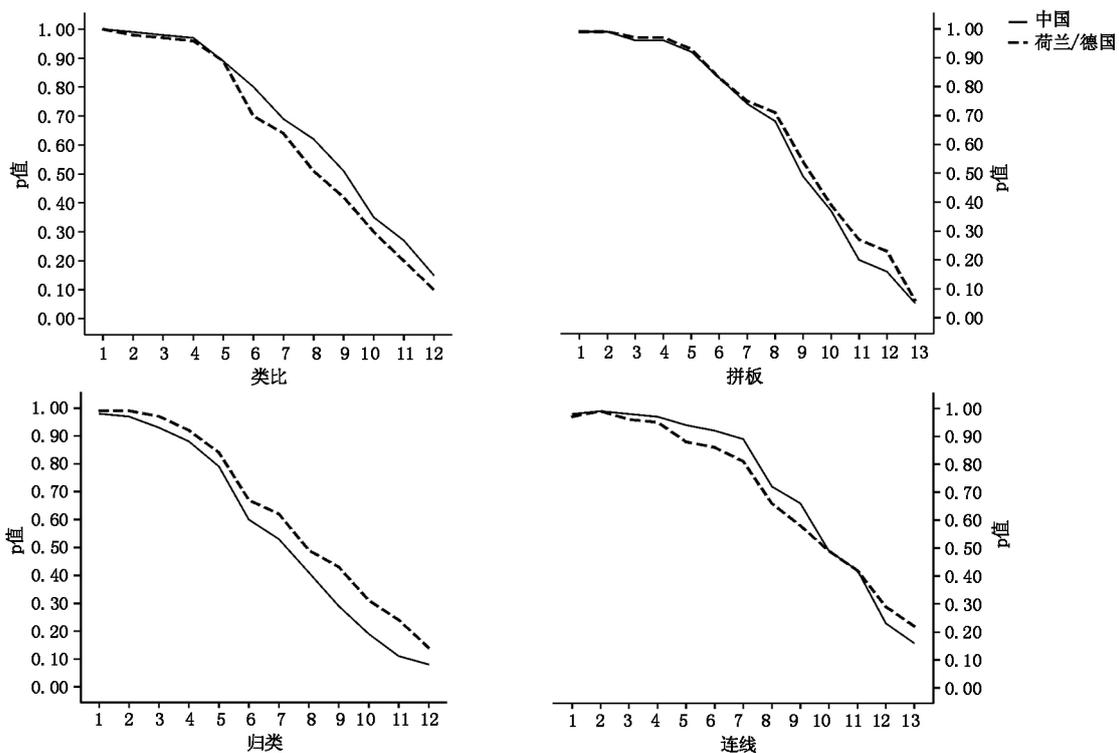
通过比较中国与荷兰/德国样本中各分测验的题目难度(如图2),发现在类比和连线分测验的部分题目上,中国人的表现优于荷兰/德国人,p值差异大于0.04(见表7);而在归类分测验中十个题目,拼板分测验中三个题目以及连线分测验最后两题上则相反,p值差异也大于0.04。Tellegen和Laros(2011)在SON-R 6-40非言语智力测验荷兰版测验手册中指出,p值的差异在0.02以上可以视为难度确实存在差异。由此可以说明,对于中国与荷兰/德国人而言,SON-R 6-40非言语智力测验某些测验项目的难度的确是不一致的。因此,SON-R 6-40非言语智力测验全量表等值水平的证据不充分,需要进一步分析。

表7 中国与荷兰/德国样本各分测验的项目难度(p值)

项目序号	类比		拼板		归类		连线	
	中国	荷/德	中国	荷/德	中国	荷/德	中国	荷/德
1	1.00	1.00	0.99	0.99	0.98	0.99	0.98	0.97
2	0.99	0.98	0.99	0.99	0.97	0.99	0.99	0.99

续表 7

项目 序号	类比		拼板		归类		连线	
	中国	荷/德	中国	荷/德	中国	荷/德	中国	荷/德
3	0.98	0.97	0.96	0.97	0.93	0.97	0.98	0.96
4	0.97	0.96	0.96	0.97	0.88	0.92	0.97	0.95
5	0.89	0.89	0.92	0.93	0.79	0.84	0.94	0.88
6	0.80	0.70	0.83	0.83	0.60	0.67	0.92	0.86
7	0.69	0.64	0.74	0.75	0.53	0.62	0.89	0.81
8	0.62	0.51	0.68	0.71	0.41	0.49	0.72	0.66
9	0.51	0.42	0.49	0.54	0.29	0.43	0.66	0.58
10	0.35	0.30	0.37	0.39	0.19	0.31	0.49	0.49
11	0.27	0.20	0.20	0.27	0.11	0.24	0.42	0.42
12	0.15	0.10	0.16	0.23	0.08	0.14	0.23	0.29
13			0.05	0.06			0.16	0.22

图2 中国与荷兰/德国各分测验项目难度(p 值)比较

4 讨论

按照测验由低到高的跨文化等值水平,依次检验了SON-R 6-40非言语智力测验在中国与荷兰/德国样本间的结构等值性、测量单位等值性和全量表等值性。结果表明该测验在中国与荷兰/德国两样本间的跨文化等值性达到了测量单位等值水平。同时,项目难度分析结果也为东西方个体智力的文化差异提供了证据。

4.1 SON-R 6-40 非言语智力测验的跨文化等值水平

内部一致性系数差异检验结果表明,除连线分测验外,SON-R 6-40非言语智力测验各分测验及

总量表的信度在中国与荷兰/德国两样本间不存在显著差异。事实上,连线分测验的信度在两样本中的差异也较小,差异值只有0.02,之所以显著,一方面可能是因为样本量较大,高估了差异的显著性,另一方面可能是由于两样本原始分数的标准差差异较大。进一步分析发现,各分测验中,连线分测验原始分数的标准差在两样本中的差异最大($\Delta s = 1.13$),而在中国样本中,该分测验原始分数的标准差较小,导致测验信度被低估,从而导致了与荷兰/德国样本在信度上的差异。因此可以说,无论在荷兰、德国还是在中国,该测验各分测验及总量表都具有良好的稳定性与可靠性。

采用探索性因素分析的跨文化结构等值结果表明,SON-R 6-40 测量构想的单因子在中国与荷兰/德国两个样本间具有高度一致性。但这一单因素模型与按完成任务方式将测验分为推理与操作两部分的双因子模型(Tellegen & Laros 2011)不一致。于是,又分别对两个样本强制提取两因子并进行斜交旋转,结果显示两个样本中都只有归类分测验在第二因子上载荷较大,因此并不支持按任务完成方式所划分的双因素模型。由此可见,无论在中国还是荷兰/德国,单因素模型都更为合理。从分测验评估的能力角度进行分析,类比分测验实际上也涉及空间能力,拼板与连线也涉及推理能力,因此很难将测验结构区分为推理与空间能力两个维度,提取一个因子更符合理论构想,该因子实际上就是流体智力(Tellegen & Laros 2011)。

在结构等值性得到验证的基础上,运用多样本验证性因素分析的方法,验证了SON-R 6-40 在中国与荷兰/德国两个样本间的测量单位等值性。随后又进一步探讨了最高水平的测验跨文化等值性——全量表等值,但并未得到充分的证据。因为项目难度分析结果表明部分题目在两样本间差异较大,并非完全一致。事实上,经典测验理论下的项目难度很难将被试的能力与题目参数加以区分。因此样本间存在难度差异,不能精确说明是题目存在偏差还是被试本身的能力不同。而要精确、详细地评估测验的全量表等值性,应采用项目功能差异分析方法(Salzberger, Sinkovics, & Schlegelmilch, 1999; 王旖旎, 1999)。

总体来说,SON-R 6-40 非言语智力测验在两样本中测验信度跨文化一致性及结构一致性系数表明了该测验跨文化结构等值,而进一步的多样本验证性因素分析证明了其跨文化等值水平已达到测量单位等值。SON-R 6-40 在中国、荷兰与德国所测量的结构一致,单位相同,充分说明该测验可以引入中国并推广使用,同时也可以进行水平取向的跨文化比较研究,如个体智力的性别差异程度在不同文化中的大小。而直接进行分数比较的跨文化研究则需要进一步充分地证明该测验可达到全量表等值水平。

4.2 东西方个体智力的文化差异现象

通过比较中国与荷兰/德国两样本中的各分测验题目难度,发现中国人在类比分测验上的表现明显优于荷兰人和德国人。这与以往研究结果一致。在类比推理能力方面,中国人的表现优于西方文化中的个体(Kuwabara & Smith 2009; 张厚粲, 龚文, 孙燕青, 田效勋, 1997),因为东方文化更关注整体及关系结构,而西方文化则更关注分析特定的对象

(Kuwabara & Smith 2009; Masuda & Nisbett 2001)。这也可以解释中国人在连线分测验中部分题目上的表现优于荷兰人和德国人,却在最后两题上的表现较差。因为最后两题中线与线的交织程度非常大,不易分辨,西方文化中的个体由于善于将其分解成单独的对象,所以表现较好,而东方文化中的个体则更容易将其知觉为一个整体,所以表现较差。

而在归类分测验上,中国人的表现却明显不及荷兰人和德国人。这一差异也与前人的研究结果一致(张厚粲等, 1997)。归类分测验中涉及概念形成,虽然该分测验的材料对中国人、荷兰人和德国人不存在文化偏差(Tellegen & Laros 2011),但由于概念形成受过去经验的影响,中国人所形成的概念系统与西方文化中的个体毕竟存在差异,而可能正是这一差异导致了题目难度的增加。比如,施测过程中,主试发现,学过英语与未学过英语的人对该分测验a系列中第6题的反应速度明显不同,前者的反应速度更快。更确切的结论则有待后续有关文化差异的研究来进一步深入探讨。

5 结论

研究发现,SON-R 6-40 非言语智力测验在中国文化情境下满足结构等值水平,且已达到测量单位等值水平,可以在中国推广使用,并适用于水平取向的跨文化比较研究。

参考文献

- 凌文轮, 方俐洛. (2003). *心理和行为测量*. 北京: 北京机械工业出版社.
- 王旖旎. (1999). 教育测评中的不公正问题——项目功能差异. *中国远程教育*, 8, 39-41, 63.
- 温忠麟, 侯杰泰, 马什赫伯特. (2004). 结构方程模型检验: 拟合指数与卡方准则. *心理学报*, 36(2), 186-194.
- 徐建平, 张厚粲, 郭枝, 杜艳婷, 武琳, Peter Tellegen. (2011). 斯-欧非言语智力测验在特殊群体中的应用. *中国特殊教育*, 11, 21-25.
- 张厚粲, 龚文, 孙燕青, 田效勋. (1997). 斯-欧氏非言语智力测验的修订研究. *心理科学*, 20(2), 97-103, 190.
- Cheung, G. W. (2008). Testing equivalence in the structure, means and variances of higher-order constructs with structural equation modeling. *Organizational Research Methods*, 11(3), 593-613.
- Hakstian, A. R., & Whalen, J. E. (1976). A K-sample significance test for independent alpha coefficients. *Psychometrika*, 41(2), 219-231.
- Hult, G. T. M., Ketchen Jr., D. J., Griffith, D. A., Gonzalez-Padron, T., Harmancioglu, N., Huang, Y., et al. (2008). Data equivalence in cross-cultural international business research: assessment and guidelines. *Journal of International*

- Business Studies* 39, 1027 – 1044.
- Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: a Monte Carlo investigation. *Structural Equation Modeling: A Multidisciplinary Journal* 8(2), 205 – 223.
- Kuwabara, M., & Smith, L. B. (2009). Cross – cultural differences in analogical reasoning. In development and learning. *IEEE International Conference* 1 – 6.
- Laros, J. A., & Tellegen, P. J. (1991). *Construction and validation of the SON – R 5 – 17 the Snijders – Oomen non – verbal intelligence test*. Groningen: Wolters – Noordhoff.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self – concept: first – and higher order factor models and their invariance across groups. *Psychological Bulletin* 97(3), 562 – 582.
- Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, 81(5), 922 – 934.
- Matsumoto, D., & Van de Vijver, F. J. R. (2011). *Cross – cultural research methods in psychology* (pp. 179 – 215). New York, USA: Cambridge University Press.
- Myers, M. B., Calantone, R. J., Page, T. J., & Taylor, C. R. (2000). An application of multiple – group causal models in assessing cross – cultural measurement equivalence. *Journal of International Marketing* 8(4), 108 – 121.
- Salzberger, T., Sinkovics, R. R., & Schlegelmilch, B. B. (1999). Data equivalence in cross – cultural research: a comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal*, 7(2), 723 – 738.
- Snijders, J. Th., & Snijders – Oomen, N. (1970). *Snijders – Oomen Non – verbal intelligence scale: SON – '58*. Groningen: Wolters – Noordhoff.
- Snijders, J. Th., Tellegen, P. J., & Laros, J. A. (1989). *Snijders – Oomen Non – verbal intelligence test: SON – R 5 – 17. Manual and Research Report*. Groningen: Wolters – Noordhoff.
- Snijders – Oomen, N. (1943). *The examination of intelligence with deaf – mute children*. Nijmegen: Berkhout.
- Starren, J. (1978). *The development of a new version of the SON for 7 – 17 year olds. Manual and Research Report*. Groningen: Wolters – Noordhoff.
- Tellegen, P. J., & Laros, J. A. (2011). *SON – R 6 – 40 niet – verbale intelligentietest*. Germany: Hogrefe.
- Tellegen, P. J., Winkel, M., Wijnberg – Williams, B., & Laros, J. A. (1998). *Snijders – Oomen Nonverbal Intelligence Test. SON – R 2 – 5 – 7. Manual and Research Report*. Lisse: Swets Test Publishers.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross – cultural assessment: an overview. *Revue Europeenne de Psychologie Appliquée* 4(54), 119 – 135.
- Wang, C. L., Bristol, T., Mowen, J. C., & Chakraborty, G. (2000). Alternative modes of self – construal: dimensions of connectedness – separateness and advertising appeals to the cultural and gender – specific self. *Journal of Consumer Psychology* 9(2), 107 – 115.
- Welkenhuysen – Gybels, J., & Van de Vijver, F. J. R. (2001, May). *Methods for the evaluation of construct equivalence in studies involving many groups*. In the 56th Annual Conference of the American Association for Public Opinion Research, Montreal, Canada.

Applicability of the SON – R 6 – 40 Non – verbal Intelligence Test in China

Xu Jianping^{1,2}, Guo Zhi^{1,3}, Wang Qian^{1,4}, Peter Tellegen⁵

(1. School of Psychology, Beijing Normal University, Beijing 100875;

2. Beijing Key Lab for Applied Experimental Psychology, Beijing 100875;

3. Institute of Human Factors and Ergonomics, Tsinghua University, Beijing 100084; 4. Jincin Education, Beijing 100080;

5. Department of Psychometrics & Statistical Techniques, University of Groningen, Groningen, The Netherlands 9700 AB)

Abstract: SON – R 6 – 40 Non – verbal Intelligence Test (SON – R 6 – 40) is a test for evaluating individual's reasoning and spatial ability in nonverbal form. The aim of present study is to examine whether it can be applied directly in China. Based on the national census a sample of 1721 individuals of different ages from different provinces was selected. By comparing it with a mixed Dutch and German sample the levels of equivalence for the SON – R 6 – 40 were analyzed. The results show that: (1) both the reliabilities and the construct of the test are identical between the two samples, which means good construct equivalence in two cultures; (2) by multiple – group confirmatory factor analysis the metric equivalence of the SON – R 6 – 40 is clearly demonstrated. Therefore, the SON – R 6 – 40 test is adequate for application in China, and also can be used for cross – cultural studies on the intelligence.

Key words: SON – R 6 – 40; nonverbal intelligence test; cross – cultural equivalence