

托尼非文字智力测验(TONT-2)的初步修订*

张雨青** 查子秀

龚正行 程念祖

刘彭芝

(中国科学院心理研究所心理健康重点实验室,北京,100101)

(北京市第八中学)

(中国人民大学附属中学)

1 问题的提出

托尼非文字智力测验(Test of Nonverbal Intelligence,简称TONI,1990年修订后简称TONT-2)是以非文字的方式评价被试者的问题解决能力和智力水平的一个智力测验。美国原版测验适用于5至85岁被试。该测验不受被试者的文化背景和教育水平的影响,适合测量的年龄范围较广、筛查较细,是应用于团体和个人的一个鉴别效果较好的潜能评估测验。由于国内的非文字类智力测验种类很少,而当前的教育、教学等实际应用急需对一般和特殊儿童的智力水平提供具有较高鉴别力的评估,因此,很有必要引进和修订这个测验。

托尼非文字智力测验是由L. Brown, R. J. Sherbenon和S. K. Johnsen编制的。该测验主要是以非文字测验的七项原则为根据编制的,这七项原则是:1:用作业的方式取代纸笔测验;2:以手势而非以口语或文字方式施测;3:正式施测前设有例题帮助被试者熟悉答题的过程;4:测验时间不作限制;5:测验题型具有抽象性;6:着重推理或问题解决而非具体知识的回忆;7:测验题目应该避免过去学习经验的影响。

此外,在编制的基本指导思想,该测验是以评价被试者对抽象图形的问题解决能力为目的的,编制者认为:问题解决是在认知过程中的一种普遍、共通性的活动,足以反映出个人的智力水平。强调以问题解决能力作为智力评价的核心成份的重要性。在这方面,有些著名心理学家也提出了类似的观点:“智力是指个体解决问题所需的能力”,“问题解决能力是智力的共同成分或智力结构的核心,而不是一种独特分离的技能,或是次要的认知成分”,此外,在编制本测验时,编制者还仔细研究了瑞文标准推理测验(SPM)、彩色推理测验(CPM),以及韦氏智力测验各版本的操作测验等。和这些测验的主导思想的不同之处在于,瑞文氏十分强调推理能力,认为推理能力包括有效解决问题的能力,而本测验的编制者认为,推理只是在解决问题时的一种思考过程,推理能力只是问题解决能力中的一种。

因此,“问题解决能力”可以说是智力结构中最重要的一部分。在托尼测验中,是以简单配合、相似性、分类、交叉、渐进五种解决抽象图形问题的基本策略来界定“问题解决能力”结构的。在施测过程中,被试者需要鉴别、分析抽象图形之间的关系,并据此解决问题。本测验的每一个题目,都呈现出一个刺激组型,其中缺少一个(或一组)图形,被试者必须从选项中选出一个正确的答案来完成此组型。这些解决抽象图形问题的策略是由易到难、循序渐进。

该测验的施测一般不作时间限制,通常在15分钟内即

可完成(中文修订版宽限至30分钟)。测验题目均为非文字图形方式的题型。受试者的回答反应过程可以不涉及听、说、读、写的能力,只需要运用简单的动作反应。并且,测验题册一页一题,依难易程度排列。中文版设有七个例题可以让受试者熟悉题型和回答方式。施测时,主试可以用手势说明,受试者则以简单的动作指出答案(尤其适用于低年龄或身心障碍者),或在答案纸上划记作答。

该测验原适用年龄为5至85岁,全部题目均为黑白设计的抽象图形。原版测验的A、B两个题本各有50题,但TONI-2又各加了5题,成为各有55题的测验。台湾修订版在美国原测验55题的基础上,根据被试的年龄确定了一个包含45题的幼儿测验版和包含63题的普通版。本研究是以该台湾修订版的普通版甲式测验为基础。此外,托尼测验的施行方式主要以个别施测为主,也可以小组和集体方式进行。

有关托尼测验与其他测验的相关研究以及适用性等方面已有研究报告。Vance(1988)以6至16岁学生为被试,用托尼非文字智力测验和Quick神经性筛查测验作为预测变量,来预测韦氏智力测验(WISC-R)的得分,发现这两个测验均可以较好地预测韦氏智力测验的得分。其中,如果托尼测验测得的智商为100,那么能够预测WISC-R的全量表智商分为:95.3 ± 9.1(68%置信区间)。

Haddad(1986)比较了51名有学习障碍学生在托尼测验和韦氏测验(WISC-R)上得分的相关,发现托尼测验和韦氏测验全量表分数和操作量表分数均达到显著性相关。但是,控制了操作量表分数的影响后,言语量表和托尼测验之间的相关则不显著($r = .13, p < .05$);控制了言语量表的影响后,操作量表和托尼测验择优显著的相关($r = .30, p < .05$)。该结果说明托尼测验的确为一个非文字智力测验。

吴武典等以台湾1668名小学生为被试,建立了一至六年级学生常模。詹秀美以854名小学生为被试,比较了托尼测验和创造性思考和解决问题的能力之间的关系。发现:托尼测验和威廉斯创造能力测验中的开放能力相关为.21($p < .01$);与独创能力的相关为.13($p < .01$);与精密能力相关为.18($p < .01$);与标题能力相关为.15($p < .01$)。托尼测验和问题解决能力的相关为.44($p < .01$)。

特别是,肖金土和杞昭安还将托尼测验应用于具有听觉和视觉障碍学生的智力鉴定和作为制定个别化教学方案的依据,结果显示该项测验在评价特殊儿童能力方面具有独特价值。

2 研究的方法

根据社会经济和教育发展水平,本研究选择了北京市的

* 本研究属于中国科学院知识创新工程重要方向项目(批准号:KSCXZ-SW-221)

** E-mail: zhangyq@psych.ac.cn

三个城区,每个区分别选择程度中等的小学 and 中学各一所,然后在每所学校的每个年级中随机分别抽取一个普通班作为样本,进行施测。最后选取的样本总数为 1658 人,确定样本的年龄分布为 6 至 18 岁。常模样本的抽取分布及百分比详见表 1 所示。

本研究用团体测验法进行施测,即主试将测验题本和测验纸发放给学生,由学生在班上集体作答。之前,先要由主试者带领做七道练习题,以熟悉测验的题型合作答方式。然后正式开始作答,时间为 30 分钟。

表 1 常模样本的分布及百分比

城区	学校	人数分布	百分比
海淀区	双榆树小学	260	15.7
	育英中学	287	17.3
西城区	自忠小学	235	14.2
	丰盛中学	296	17.8
石景山区	古城五小	289	17.4
	古城中学	291	17.6
合计	1658		

3 结果

3.1 基本常模的建立

本次修订所选择的样本在托尼测验(TONT-2)中测验分数的平均数和标准差如表 2 所示。

表 2 各年龄组样本的平均数和标准差及内部一致性系数

年龄组(岁)	平均数	标准差	人数	系数
6	24.15	8.62	48	.88
7	28.07	9.75	107	.89
8	34.25	9.03	107	.88
9	36.16	8.77	138	.87
10	38.06	9.25	155	.89
11	42.27	8.82	150	.88
12	45.22	6.80	161	.80
13	47.04	5.67	178	.77
14	46.85	7.02	191	.84
15	47.33	5.27	149	.77
16	47.18	7.02	114	.84
17-18	46.88	6.54	160	.82
总体			1658	.89

3.2 不同性别样本的结果

在本研究的被试中,男生为 863 人,占全部被试的 52%,女生为 795 人,占全部被试的 48%,不同性别被试在 TONI 测验的总分差异不显著, $t = -1.53, p = .128 > .05$ 。

3.3 信度和效度检验

3.3.1 信度

同质性信度:根据常模样本中各年龄组被试的测验分数计算其内部一致性系数(即 a 系数)评估,发现各年龄组在 .77 到 .89 之间,总体是 .89,平均数是 .85,符合心理测量学的标准,可以用于进行团体比较和个人鉴别,各年龄组的 a 系数见表 2 所示。

分半信度:根据测验的奇数题和偶数题的得分计算分半信度为 .88。

3.3.2 效度

建构效度:图 1 所示为本测验的分数与年龄之间的关系曲线,本测验的分数平均数均随年龄的增加而稳定递增,根据单因素方差分析和 Post Hoc 检验(Scheffe)表明:12 个年龄组的测验分数具有显著性差异($F = 110.69, p < .001$)。在不同年龄组中,6 岁组和 7 岁组的测验分数除了相互之间差别

不显著外,这两个年龄组的测验分数与其他 10 个年龄组的分数均有显著性差异;8,9,10 三个年龄组除了相互之间差异不显著外,和其他 9 个年龄组的测验分数差异均达到显著性;11 岁组除了和 12 岁组的差异不显著外,和其他 11 个年龄组的测验分数均有显著性差异;12 岁组和 6 至 10 岁组的测验分数有显著性差异,但是和 11,13 至 17-18 岁组的差异均不显著;到了 13 岁以上的年龄组,则均与 6 至 11 岁组的分数呈显著性差异,但是和 13 到 17-18 岁组之间的分数差异均不显著。这个结果表明,托尼测验分数从 6 岁至 13 岁呈现出随着年龄的提高而逐渐增加的现象,但是,从 13 岁起逐步呈现出变化不大的平原状态。这个结果似乎说明了随着年龄的增加,儿童的问题解决能力也在逐渐提高,但是,从 13 岁以后则逐渐趋于平稳,同时说明本测验具有较好的建构效度。

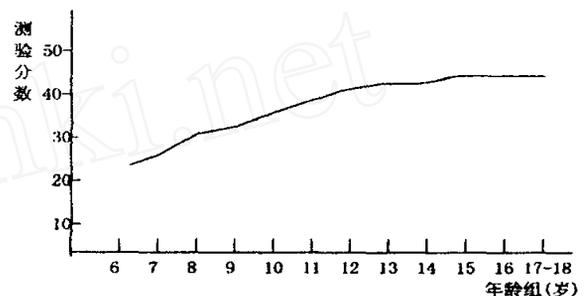


图 1 各年龄组测验分数曲线

同时效度:

本测验分数与瑞文标准推理测验的相关为 .65 ($p < .0001, N = 28$)。该结果表明,同是非文字智力测验的两个测验具有较高的相互关系。然而,托尼测验的难度相对来说明显要高于瑞文测验。

为了检验该测验对学习成绩的预测效度,对小学一年级和四年级两个班级 79 名小学生的语文和数学期末考试成绩计算了相关,发现:本测验与数学测验成绩的相关为 .45,和语文测验成绩的相关为 .34,说明该测验具有一定的预测效度。

表 4 北京儿童和台湾儿童的测验分数比较

年龄组	北京儿童			台湾儿童*			d 值
	人数	M	SD	人数	M	SD	
6 岁	48	24.15	8.62	314	21.43	6.77	0.35
7 岁	107	28.07	9.75	322	27.05	9.49	0.11
8 岁	107	34.25	9.03	284	32.64	9.64	0.17
9 岁	138	36.16	8.77	271	36.20	8.41	0.00
10 岁	155	38.06	9.25	304	38.83	8.31	-0.09
11 岁	150	42.27	8.82	307	41.38	6.64	0.12
12 岁	161	45.22	6.80	343	42.29	7.14	0.42
13 岁	178	47.04	5.67	346	42.79	7.28	0.66
14 岁	191	46.85	7.02	365	44.27	6.89	0.37
15 岁	149	47.33	5.27	327	46.89	6.05	0.08
16-18 岁	274	47.00	6.74	732	47.41	6.34	-0.09
合计	1658		3915				

注:台湾儿童数据来源于吴武典等的研究。表中的 d 值为效果大小指标 Cohen's d。

4 北京儿童和台湾儿童的分数比较

将本次测验的分数与台湾儿童的同样测验分数进行比较,结果见表 4 所示。在 6 至 18 岁 11 个年龄组中,北京儿童和台湾儿童的测验分数均没有显著性差异。

(下转第 333 页)

可行性,当然有些题目需要重新归类或淘汰。

2001年又一次在上海地区采取分层随机的方式抽取220名成人及部分中学生作为被试。回收有效问卷205份,其中男性99人,女性106人。文化程度仍为最低初中,最高博士,年龄从15—55岁,平均年龄25岁。职业有教师,学生,技术人员,管理人员,工人,医护人员,军人,待业者。因素分析的结果仍证明存在着两个主因素。

根据两次预备测验的结果把能归入两个主因素的题目保留组成一个量表,命名为《竞争态度量表》,而其它无法归入该两个因素的题目被淘汰。总量表下再分成“良性竞争态度分量表”和“过度竞争态度分量表”。

3.2. 正式抽样测试 抽取上海地区330人作为常模样本。抽样的范围、年龄、职业类别、学历等与预测样本团体相似。女性被试141人,男性161人。平均年龄26.53($SD = 7.90$)。测验前主试对测验要领稍作讲解,并打消被试的顾虑,请他们如实回答。测验不限时。测验完毕,主试汇总数据,对五种回答分别按5至1来记分。然后将原始数据输入电脑,电脑自动对数据进行评分和储存,并用SPSS10.0对数据进行各种处理。

3.2 信效度检验及常模的编制

3.2.1 信度

(1) 系数 良性竞争态度分量表的内在一致性系数为0.86,过度竞争态度分量表为0.71。

(2) 分半信度 良性竞争态度分量表的Guttman分半信度系数为0.8365,斯皮尔曼—布朗校正后的分半信度系数为0.8373。过度竞争态度分量表的Guttman分半信度系数为0.7152,斯皮尔曼—布朗校正后的分半信度系数为0.7219。

(3) 项目—总分相关 良性竞争态度分量表的测题与总分间相关在0.41—0.70之间,全部具有0.01水平上的显著相关。过度竞争态度量表的测题与总分间相关在0.34—0.58之间,也全部具有0.01水平上的显著相关。

(4) 重测相关 取样本中28名中学生(18女,10男)和

20名成人(17女,3男)在两周后进行重测,结果为良性竞争态度分量表($r = 0.76(p < 0.01)$),过度竞争态度分量表($r = 0.71(p < 0.01)$)。成人组良性竞争态度分量表($r = 0.74(p < 0.01)$),过度竞争态度分量表($r = 0.71(p < 0.01)$)。

以上几项信度测验的结果表明该测验的稳定性都在可接受的范围内。

3.2.2 效度

分别从两个分量表随机挑选两个项目,求它们与两个分量表总分的相关。结果每个项目都与本分量表呈高度正相关,而与另一个分量表呈负相关。(见下表)

项目	良性题1	良性题2	过度题1	过度题2
良性	0.67	0.46	-0.14	-0.08
过度-0.11	-0.05	0.47	0.57	

3.2.3 方差分析

对被试的性别、年龄(5组)、职业(5种)、文化程度(7组)进行多因素方差分析。结果显示在过度性竞争态度方面这四个因素都无显著影响。而在良性竞争态度方面文化程度、年龄和职业在个别组别之间表现出显著性的差异。另外性别、年龄和职业三者之间交互作用显著($F = 3.014, p = 0.018 < 0.05$)

3.2.4 常模的制定

分别根据两个量表的原始分数计算出它们的平均数和标准差,由于原始分数的分布呈常态,所以直接把它们转换成Z分数,然后确定相对应的七个等级分作为常模分数。

4 小结

过度竞争和良性竞争的态度反应了一个人的不同人格特征,对个体的学习和工作都会造成不同的影响,如能对此加以科学的测量具有重要的现实意义。目前国内尚无这类量表,我们修订此量表想填补这方面的空白。通过初步的试用,效果令人满意。当然如能得到进一步的测验验证就更好了。目前的常模仅限于上海市,如有可能应尽快制定全国常模。

(接第331页)

讨论

一般来说,智力测验的型式大多与语文和动作技能的认知表现有关,其内容可能受文化因素的影响。这类测验主要用于认知能力的评价、认知组型的分析以及学业成就的预测。但是,对于有身心障碍的学生来说,由于他们在听、说、读、写、动作等能力上的缺陷或限制,这类测验的型式和内容就会在施测和使用上存在困难。其次,对一般学生或求职人员而言,在评价他们的性向或成就潜能时,也需要一个较少受特定语文及社会经济背景影响的智力测验。基于以上考虑,介绍和引进托尼非文字智力测验这样一个受语言、动作技能和文化因素影响的智力测量工具就变得十分重要。

本次初步修订该测验的结果显示,该测验具有较高的信度和效度,符合心理测量的标准,可以用来检测我国五至十八岁儿童的非文字智力水平。然而,需要注意的是,由于该测验的测题内容偏重于问题解决和图形推理,所测得的被试能力为一般能力而非特殊能力。另外,在本测验的修订结果中,由于从15岁以后基本进入了高原状态,智商变化不大,虽然从理论上来说这可能符合人类智力的发展规律,幼年时发展很快而青春期以后逐渐趋于平缓。但是,这也可能说明

本测验对于15岁以上儿童智力预测能力有限,故本测验的最佳使用年龄应为6至15岁。另外,由于本次修订的样本范围取自北京地区,样本的成绩也许会高于其他边远省份地区的儿童,因此可能存在低估或高估现象。而且,由于目前北京市的初中毕业生分流较大,进入高中的学生可能具有自己的独特性,因此,本研究选择普通高中中被试可能限制了15岁以上样本的代表性。然而,由于本次测试的结果与台湾同样年龄被试的比较显示差距均不显著,这似乎对本此测试的代表性提供了支持。

此外,在实际应用中,该测验主要有以下多种用途:1:可以用于预测智力水平和认知功能、评估儿童发展潜能的工具,并且可以作为超常儿童鉴别和选拔的指标之一。2:确定学生的学业性向,帮助学生选择课程或学习项目。对于本测验分数较高的学生倾向于在空间和图形推理方面较为优势,应该增加数学物理方面的课程学习。3:检测智能缺陷,特别是在无法分辨是智能缺陷还是语言、动作能力不足时。但是应该参照韦氏智力测验等其他个别测验的结果。4:可以作为一般智力测验的工具。根据测验成绩可以作为是否转介以进一步评价或心理咨询或治疗的参考依据。5:班级教学中素质分组的依据,教师可以根据测验分数将学生划分成不同的能力小组,进行分组教学。6:用于研究。