

# 7-9 年级语文、数学单科标准化成就测验的编制

范晓玲<sup>1</sup>, 闫春平<sup>1,2</sup>, 柴彩霞<sup>1,\*</sup>, 梁娟<sup>1</sup>

(1.湖南师大教科院心理系,湖南 长沙 410081;2.河南新乡医学院,河南 新乡 453002)

【摘要】 目的:编制 7~9 年级语文、数学单科标准化成就测验。方法:在多次试测基础上形成正式测本,采用分层和整群抽样方法,抽取语文样本 1008 人和数学样本 1018 人作项目分析,另有一定数量的被试用作信、效度估计及学科专家作内容效度评定。结果:语、数 75% 的项目难度为 0.20~0.80,88% 以上的项目鉴别力达到可接受水平,各分测验在各年级和总测验的 D 值均在 0.50 以上。语文测验总分的复本信度、分半信度在 0.74~0.87 之间,数学测验总分的各信度值在 0.73~0.88 之间。语数 A、B 题本总分与学科成绩相关为 0.47~0.69,探索性因素分析语数各抽取 2 因素模型为佳,验证性因素分析结果表明两因素模型拟合较佳。结论:7~9 年级语、数单科成就测验符合测量学的要求。

【关键词】 单科成就测验;标准化;项目分析;信度;效度

中图分类号:G449.5

文献标识码:A

文章编号:1005-3611(2008)05-0449-03

## Development of Single-subject Chinese and Mathematics Achievement Tests for Grade 7-9

FAN Xiao-ling, YAN Chun-ping, CAI Cai-xia, LIANG Juan

Hunan Normal University, Changsha 410081, China

【Abstract】 **Objective:** To develop single-subject Chinese and Mathematics achievement tests for grade 7~9 (CATs, MATs). **Methods:** According to stratified and cluster sampling, a total of 1008 CATs and 1018 MATs samples were tested in the formal study. **Results:** The difficulties of 75% CATs and MATs items ranged from 0.20~0.80. The discriminations of 88% items were acceptable. The parallel reliability, split-half reliability and coefficient of the total scores of Chinese test ranged from 0.73~0.88, and the reliabilities of the total scores of the MATs were 0.73~0.88. The correlations between total scores of CATs or MATs and those of the academic performance were 0.47~0.69. Two-factor models of the CATs and MATs were obtained by exploratory and confirmatory factor analysis. **Conclusion:** The CATs and MATs have good psychometric quality.

【Key words】 Single-subject achievement test; Standardized; Item analysis; Reliability; Validity

标准化的成就测验是教育、教学测评的组成部分之一,它不仅是普通教育和特殊教育进行各种教育决策的依据<sup>[1-3]</sup>,而且也是各行各业进行人才测评及临床评估的常用工具之一。语、数不仅是义务教育的基础学科,而且是贯穿中小学乃至大学的主干课程,因目前尚缺乏较为系统的、针对性较强的初中阶段标准化成就测验,故编制 7~9 年级语文、数学单科标准化成就测验,旨在为语、数综合测验以及与之相应的学习能力倾向测验的联合研究奠定基础。

## 1 方 法

### 1.1 编制策略

采取课程取向和广泛取向并重的原则,依据国家课程标准为依据,运用经验与理论、教师与专家相结合的方法进行测验的编制工作。

### 1.2 测验内容与项目编制

设有语文和数学两个量表,11 个分测验。其中语文量表包括拼音、字型、词汇、语法修辞、文学常

识、阅读理解和言语表达,数学分量表包括基本知识、计算、数推理和数应用。在比较各种教材,搜集已有试卷,参阅相关出版读物及学科专家命题、自编题目的基础上,进行项目的专家评定与筛选并形成试测题本。试题入选标准为基础性、广泛性、出现频率等。经个别测试,形成预测题本;再经预测和项目分析形成正式题本。

### 1.3 抽样

采用整群和分层抽样相结合的方法,分别抽取城市示范与城市一般学校,抽样人数语文 1008 人(其中 A 本 511, B 本 497),数学 1018 人(其中 A 本 516, B 本 502)。另外,语文获复本信度有效样本 96 例,效标样本 A 卷 58 例, B 卷 59 例;数学获复本信度有效样本 118 例,效标样本 A 卷 45 例, B 卷 51 例。

## 2 结 果

### 2.1 项目分析

语文量表项目难度为 0.20~0.80 的 A 本占 77.3%, B 本占 75.0%, 平均为 76.15%。各分测验的

【基金项目】 湖南省“十五”教育规划课题(XJK03BJ005)

\* 山西师大教科院

难度 A 本为 0.38~0.70, 平均难度为 0.57;B 本为 0.33~0.70,平均难度为 0.57。除拼音、字形分测验较容易外,其它分测验难度基本接近 0.50,且呈渐近正态分布。数学量表项目难度为 0.20~.80 的 A 本占 75.5%,B 本占 71.4%,平均为 73.45%,各分测验的难度平均 A 本为 0.33~0.45, 平均难度为 0.40;B 本为 0.30~0.42,平均难度为 0.39。且 A 本正态拟合性较好,B 本略呈正偏态。语、数量表的测试结果呈现出年级特征,即随着年级的增高测验得分逐渐增加。

语文量表的区分度指数属优良水平 ( $r\geq 0.30$ ) 的 A 本为 61.36%,B 本为 60.23%; 属可接受水平

( $r\geq 0.20$ ) 的 A 本为 81.82%,B 本为 86.36%,且 54.0%的题目与总分的相关达到优良水平,88%达到可接受水平。A、B 本各分测验在各年级的区分度指数均在 0.22 以上,总量表分别为 0.34 和 0.35;各分测验在各年级的相关区分度均在 0.50 以上,总体在 0.63~0.74 之间。数学量表的区分度指数属优良的 A 本为 79.6%, B 本为 77.6%。 各分测验的区分度指数 A 本为 0.48~0.58,B 本为 0.49~0.60,A、B 本总分区分度指数均为 0.53。各分测验与总分的相关为 0.57~0.87,结果详见表 1。

表 1 区分度指数与相关区分度

		语文量表					数学量表					
		1	2	3	4	5	6	7	8	9	10	11
A 本	D	0.27	0.36	0.34	0.30	0.37	0.35	0.50	0.52	0.58	0.48	0.56
	r	0.70	0.70	0.72	0.66	0.71	0.70	0.63	0.80	0.87	0.57	0.85
B 本	D	0.27	0.35	0.32	0.31	0.36	0.43	0.47	0.60	0.52	0.49	0.51
	r	0.68	0.72	0.69	0.70	0.73	0.74	0.64	0.81	0.85	0.58	0.87

表 2 各分测验的复本信度、分伴信度及 Cronbach  $\alpha$  系数

		语文量表					数学量表					
		1	2	3	4	5	6	7	8	9	10	11
A 本	$r_{nn}$	0.67	0.64	0.55	0.49	0.60	0.57	0.73	0.62	0.73	0.74	0.70
	$\alpha$	0.66	0.62	0.54	0.43	0.65	0.55	0.77	0.65	0.79	0.80	0.77
B 本	$r_{nn}$	0.64	0.65	0.54	0.43	0.65	0.55	0.77	0.75	0.60	0.77	0.67
	$\alpha$	0.62	0.61	0.51	0.56	0.62	0.72	0.67	0.73	0.76	0.81	0.75
复本信度		0.66	0.69	0.45	0.50	0.48	0.63	0.83	0.50	0.55	0.61	0.64

2.2 信度分析

语文量表各分测验的复本信度为 0.45~0.83,总量表为 0.81; 分半信度 A 本各分测验为 0.49~0.73,总量表为 0.75;B 本各分测验为 0.52~0.79, 总量表为 0.74;Cronbach'  $\alpha$  系数 A 本各分测验为 0.43~0.77, 总量表为 0.85;B 题本各分测验为 0.51~0.77,总量表为 0.87,详见表 2。此外 A、B 本各分测验的标准误在 0.81~1.97 之间。

数学量表各分测验的复本信度为 0.50~0.64,总量表为 0.76。分半信度 A 本各分测验为 0.62~0.74,总量表为 0.73;B 本各分测验为 0.60~0.77, 总量表为 0.75。Cronbach'  $\alpha$  系数 A 本各分测验为 0.65~0.80, 总量表为 0.89;B 本为 0.73~0.81, 总量表为 0.88,详见表 2。此外,各年级各分测验的测量标准误在 0.73~3.54 之间。

2.3 效度分析

2.3.1 内容效度 专家对语文量表各项目 and 分测验评定的平均符合度均达 3.50 以上,且专家评定的一致性程度为 0.80 以上;对数学量表各分测验评定的平均符合度 A 本为 3.23~4.58,B 本为 3.40~4.54,其

中有 3 个分测验达到 4.0 以上;对各项目评定的平均符合度结果在 3.00 以上。

2.3.2 效标效度 语文量表 A、B 题本总分与语文学期末成绩的相关分别为 0.47(58 人)和 0.50(59 人),数学量表 A、B 题本总分与数学期末成绩的相关分别为 0.60(45 人)和 0.69(51 人)。

2.3.3 结构效度 根据随机分层原则,将 A、B 题本总数据按学校类型、年级和性别分为两部分,一部分用于探索性因素分析,一部分用于验证性因素分析。

探索性因素分析结果见表 3 和表 4, 其两因素模型的解释力较强,其中语文量表(n=511,其中 A 本为 254 人,B 本 246 人)的分测验 1~5 主要考查学生语文基础知识,命名为语文基本知识技能;分测验 6~7 则反映阅读理解和言语表达, 命名为语文综合能力;数学量表(n=512,其中 A 本为 252 人,B 本 260 人)A、B 题本因为第一因素均负荷在分测验 1、2、4(数学基础知识,数运算,数学应用题和几何证明题)上,第二个因素均负荷在分测验 3(数推理)上,所以因素 1 命名为数学基本知识技能,因素 2 命名为数推理。验证性因素分析结果见表 5。

表 3 语文量表探索性因素分析 2 因素结构分析结果

题本	因素	FY1	FY2	FY3	FY4	FY5	FY6	FY7	特 征 根		方差贡献率
									1	2	
A 本	F1	0.672	0.815	0.781	0.706	0.781	0.606	0.113	3.56	3.23	50.78
	F2	0.099	0.070	0.184	0.161	0.146	0.460	0.962	0.89	1.22	12.68
B 本	F1	0.759	0.755	0.792	0.609	0.756	0.451	0.053	3.56	2.92	50.78
	F2	0.085	0.261	0.156	0.401	0.177	0.673	0.916	0.95	1.58	13.60

注:1 为旋转前结果,2 为旋转后结果

表 4 数学量表探索性因素分析 2 因素结构分析结果

题本		分测验 1	分测验 2	分测验 3	分测验 4	特征根		方差百分比%	累积贡献率%
						1	2		
A 本	因素 1	0.819	0.917	0.158	0.698	2.500	2.023	50.59	50.59
	因素 2	0.331	0.013	0.958	0.479	0.782	1.258	31.45	82.04
B 本	因素 1	0.861	0.890	0.173	0.763	2.500	2.145	53.62	53.62
	因素 2	0.249	0.021	0.967	0.384	0.790	1.145	28.63	82.24

表 5 A、B 题本验证性因素分析结果

	题本	N	df	X <sup>2</sup>	X <sup>2</sup> /df	RMSEA	NNFI	CFI
语文	A	257	13	19.24	1.48	0.043	0.99	0.99
	B	251	13	23.95	1.84	0.058	0.98	0.99
数学	A	262	26	71.04	2.73	0.084	0.95	0.96
	B	241	26	52.92	2.04	0.068	0.97	0.98

2.4 A、B 题本平行性分析

语文量表和数学量表 A、B 题本的平均数、标准差(语文 A 56.21±13.65, 语文 B 56.68±13.96; 数学 A 25.62±14.46, 数学 B 27.11±15.52)、平均难度及区分度均无统计上的显著差异,且 A、B 题本总难度和区分度也基本接近,表明 A、B 题本是基本平行的。

3 讨 论

本研究编制的测验是在有限时间内的团体测验,在内容设置上主要考虑了四个方面的关系,一是晶体智力理论与成就测验的关系,二是课程标准和生活的关系,三是基础性与广泛性的关系,四是实施的可行性和操作性的关系,综合各种关系而把语文测验分为拼读、字形、词语、语法修辞、文学常识、阅读理解和言语表达七部分,数学测验分为数学基本知识、数运算、数推理知识和数应用四部分。

从项目难度看,语数 A、B 题本项目难度在 0.20~0.80 之间的平均占 76%; 各分测验平均为 0.50;总测验平均为 0.48,接近 0.50,且各年级在分测验上的难度及难度等级呈现明显的年级特点和年级差异,表明项目和分测验难度基本符合常模参照测验的标准<sup>[4]</sup>。从项目区分度看,语数 A、B 题本项目鉴别力指数达到优良水平<sup>[5]</sup>的平均为 71%,符合可接受水平<sup>[6]</sup>以上的平均为 84%以上,各分测验鉴别力指数平均为 0.46;总测验平均为 0.44。题目与总分的相关指标达到可接受水平以上的项目平均 87%以上,各分测验与总分的平均相关 0.68。语文测

验 A、B 本均有 6 个项目与总分相关较低,数学则分别有 7 个和 13 个项目区别度较低。总之,绝大多数项目难度、区分度较好。不过语文的拼读偏易,项目区分度较低,原因有二:一是拼读主要是小学阶段的训练内容,经多年学习,拼读都达到了相当水平;二是测验形式与测试目标上,只检测了再认能力,而未检查回忆能力,从而降低了难度。数学测验比语文稍难,按难度等级分布标准,属中等或中等偏难水平。原因可能有二:一是数学学科本身的阶梯性所致,二是测试时间为秋季,只相当在各年级学习了 3 个月的水平。

语数 A、B 本部分的重测信度、复本信度、分半信度和同质性信度分析总信度值均在 0.80 以上,符合团体测验及标准化成就测验的标准<sup>[7-9]</sup>。值得注意的是,语文测验中言语表达虽只有 4 个项目,但信度值却是所有分测验中最高的均在 0.77 以上,说明影响测验信度的关键因素并不是题量,而是质量。内容效度是评价成就测验的一个非常重要的指标,我们采用专家判断法、经验法和复本法对测验的内容效度进行了评定。语、数测验各分测验及项目的符合度均达到 3.00 以上,说明测验内容符合 7~9 年级教学的主要内容,项目代表性较高;以经验法观察七到九年级学生原始分和项目通过率发现,测验分数呈明显的年级特征。另外,按克龙巴赫的观点,内容效度可由一组被试在取自同样内容范围的两个复本相关来估计,相关高则说明具有内容效度,语数 A、B 题本的复本信度分别为 0.81 和 0.76。语、数的同时效度以语文和数学期末统考成绩为效标,语文 A、B 本总分与语文期末成绩的相关为 0.47 和 0.50,数学 A、B 成就测验与数学期末成绩的相关为 0.60 和 0.69,已有研究证明标准化的成就测验与学科成绩的相关为中等水平的,大约为 0.5~0.6<sup>[10]</sup>,而本研究

(下转第 467 页)

正交旋转,将 20 项条目经过探索性的因子分析,最后获得特征值大于 1 的因子共有三个,包含 18 个条目,累积的解释变异量为 61.47%,说明该量表具有良好的结构效度。同时,也进一步验证了该量表符合研究者当初编制量表的构想。

此次研究调查了全国 5 个城市的吸毒患者,结果反映了目前他们的信心状况。如果某一被试者的得分高于均值,则说明他的信心比较高,对于以后的康复有积极的作用;反之,如果某一被试者的得分低于均值,则说明他的信心比较低,有必要进一步树立信心,为进一步的康复做准备。而且可以根据分量表显示的内容,有重点的针对其中的主要方面进行心理治疗。(致谢:本研究得到教育及社会科学应用研究论文奖计划资助,并得到武汉、杭州、郑州、西安、天津五个城市公安局强制戒毒所干警和医务人员的大力支持,在此一并感谢!)

#### 参 考 文 献

- 1 李峰,黄焕光,宋月晗,等.劳教戒毒人员康复期心理调查与康复对策.中国康复理论与实践,2006,12(1):73-74
- 2 Weiss F, Ciccocioppo R, Parsons LH, et al. Compulsive drug-seeking behavior and relapse. Neuroadaptation, stress, and conditioning factors. Ann N Y Acad Sci, 2001, 937:1-26
- 3 王海琴,蒲东华.518 例阿片类依赖者复吸原因分析.中国药物滥用防治杂志,2004,10(2):85-87
- 4 朱炜良.海洛因依赖者复吸风险因素的研究.中国行为医学科学,2002,11(1):20-22
- 5 刘永有,曾岳峰,王增珍.心理干预预防毒品复吸的效果分析.中国药物滥用防治杂志,2004,10(6):321-325
- 6 曲如杰,林霖,王文忠.吸毒者心理健康状况及与复吸原因的关系.中国临床心理学杂志,2006,14(1):55-57
- 7 铁恩贵,曾恒,金俊,等.阿片类依赖者脱毒后复吸情况及复吸原因的流行病学调查.中国药物依赖性杂志,1999,8(4):305-309
- 8 杨丽,刘盈,吴枫.大学生心理健康调查及心理干预效果评估.中国临床心理学杂志,2005,13(1):102-108
- 9 王玲,叶明志,韩自力,等.自信与医学生心理状态及行为方式关系探讨.中国学校卫生,2003,23(1):53-54
- 10 姚斌,韩卫,吴朝俊,等.毒品依赖者社会支持与心理健康的相关分析.中国临床心理学杂志,2005,13(2):219-220,239
- 11 赵敏,郝伟,杨得森,等.海洛因依赖者复吸相关因素的前瞻性研究.中国临床心理学杂志,2001,9(2):81-83,89
- 12 Gwaltney CJ, Shiffman S, Norman GJ, et al. Does smoking abstinence self-efficacy vary across situations? Identifying context-specificity within the Relapse Situation Efficacy Questionnaire. J Consult Clin Psychol, 2001, 69(3):516-527
- 13 Janekovic K. Comparative research on substance abuse and self-perception among adolescents with physical handicap. Coll Antropol, 2003, 27(2):479-489
- 14 Breslin FC, Sobell LC, Sobell MB, et al. A comparison of a brief and long version of the Situational Confidence Questionnaire. Behav Res Ther, 2000, 38(12):1211-1220
- 15 郑日昌.心理测量.湖南:湖南教育出版社,1987. 320-4251

(收稿日期:2008-02-27)

(上接第 451 页)

结果与学科成绩之间的相关达到中等水平的相关,表明语数 A、B 测验具有良好的实证效度。分别以探索性和验证性因素分析考察测验结构发现,在抽取不同数目时的因素结构模型时,两个因素的结构模型解释性较为合理,其中语文因子为语文基本知识技能和语文的综合能力,数学因子有数学基本知识技能和数推理。

#### 参 考 文 献

- 1 Scott G, Jodie L, Julianne C. Developing Disillusionment: Students' Perceptions of Achievement Test. Issues in Education, 2000, 6:29-30
- 2 Bell, Paul F, Lentz Jr, Francis E. Effects of curriculum-test overlap on standardized achievement test scores: Identifying systematic confounds in educational decision making. School Psychology Review, 1992, 21(4):644-656
- 3 Whitehead, Bruce; Santee, Phillip. Using standardized test results as an instructional guide. Clearing House, 1994, 67(6):317-319
- 4 金瑜.心理测量.上海:华东师范大学出版社,2001. 257-258
- 5 戴海崎,张锋,陈雪枫.心理与教育测量.广州:暨南大学出版社,2002. 117
- 6 Lou MC. Measuring and Evaluating School Learning. Allyn and Bacon, 1994. 220-222
- 7 Edward GC, Richard AZ. Reliability and Validity Assessment. Sage Publications, 1976. 50-51
- 8 吉尔伯特·萨克斯著.王昌海,董奇译.教育和心理的测量与评价原理.南京:江苏教育出版社,2002. 281
- 9 中国就业培训技术指导中心,中国心理卫生协会.国家职业资格培训教程:心理咨询师(基础知识).北京:民族出版社,2005. 352
- 10 William LD, Gerald H, Tim A. Coaching Effect on California Achievement Test Scores in Elementary Grade. Journal of Educational Research, 1987, 80(3):149-155

(收稿日期:2008-01-28)