

中学生学业能力倾向测验的初步编制

张月娟^{1,2}, 龚耀先¹

(1. 中南大学湘雅二医院心理中心, 湖南 长沙 410011;

2. 河北师范大学教育科学学院, 河北 石家庄 050091)

【摘要】 目的: 编制一个适用于中学生的学业能力倾向测验。方法: 采用理论与工作分析相结合的编制策略而成的中学生学业能力倾向测验包括 A、B 两个复本, 均由三个分量表含七个分测验组成, 分别测量语文、数理及图形等能力。研究样本由河北省石家庄市 1424 名中学生组成, 并对其中 468 人分别进行信效度检验。结果: ①各分测验原始分基本随年级升高而增长; ②各分测验难度在 0.46~0.54。鉴别指数在 0.38~0.52; ③各分量表及总分的重测相关系数在 0.68~0.92, α 系数在 0.64~0.89, G 系数在 0.64~0.92; ④探索性因素分析表明各分测验对基本能力因子均有中度以上的负荷, 抽取的三个因子解释了总方差的 72.18% (A 式) 和 73.53% (B 式); ⑤测验分数与长-鞍团体智力测验、与教师评价均呈显著正相关。测验分数在不同年级之间存在显著差异。学习成绩的回归分析可见, 测验分数与考试成绩显著正相关, 语文及数理分量表进入考试总成绩的回归方程。结论: 中学生学业能力倾向测验的条目质量及信效度符合心理测量学的基本要求。

【关键词】 学业能力倾向测验; 编制; 难度; 区分度; 中学生

中图分类号: G449.5

文献标识码: A

文章编号: 1005-3611(2004)01-0001-05

The Development of the Academic Aptitude Test for Middle School Student

ZHANG Yue-juan, GONG Yao-xian

The Second Xiangya Hospital, South Central University, Changsha 410011, China

【Abstract】 **Objective:** To develop a new instrument, the Academic Aptitude Test for Middle School Student (AATMS), for assessing academic aptitude of middle school student. **Methods:** AATMS compose of AATMS-A and its parallel form AATMS-B, both consisted of three sub-scales, namely, language, math, figure there were totally seven sub-tests. A total of 1424 middle school students were tested in the formal study. **Results:** ①The higher grade was, the higher raw score of AATMS was; ②The difficulty of sub-tests ranged from 0.46~0.54, while the discrimination of sub-tests ranged from 0.38~0.52; ③The test-retest stability coefficients, Cronbach α coefficients, and G coefficients for sub-scales and the whole test ranged from 0.68~0.92, 0.64~0.89, 0.64~0.92, respectively; ④Three factors were obtained in exploratory factor analyses, accounting for over 70% of the total variance; ⑤Correlations between AATMS and Changsha-Anshan Intelligence Test in Group, teacher's rating, and exam scores were substantially positive. There were significant differences between grades. Liner stepwise regression analysis showed that language and math sub-scales were important predictors for the total final exam scores. **Conclusion:** AATMS developed in the present study attains acceptable psychometric properties.

【Key words】 Academic aptitude test; Development; Difficulty; Discrimination; Middle school student

能力倾向 (aptitude) 是指一个人经过适当训练或置于适当环境下完成某项任务的可能性^[1], 或者说是指个体现实的能表明未来在相应活动中的作业成绩的能力状况^[2]。能力倾向测验通过对现有能力的评量预测未来的业绩水平, 具有诊断和预测两种功能。在西方国家, 能力倾向测验于二十世纪三十年代, 在社会的需要、传统智力测验的局限及智力特质理论和因素分析技术三大因素刺激下应运而生并迅速发展, 研究者采用理论法、工作分析法以及理论和实践相结合的编制策略编制了各种能力倾向测验, 主要用于职业和教育两大领域^[3]。

在教育领域中, 学业能力倾向测验旨在通过对

基本学业能力的测量来评估学生不同学业能力方面的优劣, 预测学生的学业成绩或新的学习情景中成功的可能性^[4,5]。虽然它与成就测验及智力测验同处于能力测验的连续谱中, 但相比有着不同的特点: 不象成就测验其内容依赖具体的知识和特定的教育训练经验, 其目的是对学生完成某种课程或训练后情况的最终评估, 能力倾向测验主要是对学生通过校内外各种经验积累而长时间发展的学业能力进行测量, 其目的是前瞻性的, 故比成就测验更能预测学生未来学习成功的可能性; 也不象智力测验旨在测量更为广泛的影响各种活动的基本能力, 能力倾向测验不必局限于某种智力理论, 而是把与学

习活动关系最密切的能力因素作为测验的内容,它对学业能力的评估及学业指导更具针对性^[6,7]。所以,学业能力倾向测验在许多西方国家应用广泛。在《心理测量年鉴第九版》收录的1409种心理测验中,智力测验和学业能力倾向测验(被归为一类)占7.1%^[8]。美国著名的学业能力倾向测验有:鉴别能力倾向测验(Differential Aptitude Tests,简称DAT)、学业能力倾向测验(Scholastic Aptitude Tests,简称SAT)、认知能力测验(Cognitive Ability Test,简称CogAT)等。在我国台湾地区1992年的统计,修订和编制的能力倾向测验占了心理测验总数(225个)的16%,这些测验同成就测验、兴趣测验等一起在学校中用得非常普遍^[9,10]。学业能力倾向测验主要用于学习问题的鉴别与诊断、学业及职业的咨询与指导以及高一、二级学校录取依据等。

相比之下,在我国的中小学里,用学业能力倾向测验这种客观的测量工具评估学业能力几乎还是一个空白。随着我国的教育由应试教育向素质教育的转变,客观地了解学生的能力状况而因材施教,发展学生的学习潜能,指导学生选择更适合其能力倾向的专业和职业等日益受到重视,这势必需要有科学的评估工具^[11]。遗憾的是,除林传鼎和张厚粲教授等于1979年根据澳大利亚学习能力测验(TOLA)改编的适用于小学毕业生的少年儿童学习能力测验外^[12],自行编制的适用于中学生的学业能力倾向测验少之又少。鉴于此,本研究试图编制“中学生学业能力倾向测验”(Academic Aptitude Test for Middle School Student,简称AATMS)以敷上述之需。本文拟对量表的编制过程及区域性测试结果做一总结。

1 研究方法

1.1 总体构想

AATMS旨在评估中学生基本学业能力倾向,以期为中学生的学习辅导、学业及职业发展咨询以及相关的教育和临床科研提供一个有效的测量工具。

1.2 编制策略

采用理论和工作分析相结合的编制策略:①以能力的层次结构理论为理论基础,吸收学业能力的研究成果,借鉴国外相关测验;②访谈了数名富有经验的中学教师和心理测量专家,并对一重点和一普通中学的90名平均教龄11.13年的各科教师就各科学习所要求的基本能力进行了开放式问卷调查;③根据测验的总体构想及评估目标,AATMS以测量与学习数学、语文等密切相关的基本学业能力进行建构;④为测验实施的方便性和评分的客观性,

AATMS采用可团体施测的纸笔测验形式,并建立A、B两平行复本;⑤测验耗时限制在60分钟左右。

1.3 确定量表结构

AATMS含A、B两式,各式均由三个分量表各含七个分测验构成:①语文分量表。包括填词分测验,拟测量言语的类比推理能力;填句分测验,拟测量在语境中运用词汇的能力以及语言表达能力;阅读分测验,拟测量对深层语义的理解以及对文字段落综合分析和概括能力;②数理分量表。包括算术分测验,拟测量数字操作和运算以及对基本数量关系的理解能力;填数分测验,拟测量数字间关系的概括和推理能力;③图形分量表。包括平面分测验,拟测量对平面图形的知觉和推理能力;立体分测验,拟测量立体图形的空间关系的知觉和想象能力。

1.4 草本预试、条目筛选及实测版本的形成

建立条目库。绝大部分条目自行编制,部分条目借鉴了国外相关测验。经小范围试验对条目进行反复修改,还特别注意了条目中措辞和语句表达的准确性以及图形绘制的规范性,以加强表面效度。用测验草本对三所城市中学初一至高三的107名学生进行预试。对预测结果主要进行条目难度和区分度分析,另外还通过预测对每个分测验的时限进行了确定。将一些不合要求的条目淘汰或修改,然后把条目按难度排序,奇偶分半,将题本分成难度相当,条目形式数量对等的A、B两式,各式均含7各分测验128个条目,形成本研究的实测版本。

1.5 取样及资料分析

按比例分层整群取样共取河北石家庄市6所中学初一至高三29个教学班的1424名中学生(A式713,B式711),重点、一般和较差等三类学校的人数比例约为1:3:1。全部取样工作均由作者一人完成,在每个教学班施测时,A、B式随机隔列发放,施测指导语一致,均用秒表控制时间。另外,重测样本105人,效标效度样本363人。全部取样于两月内完成。测验条目均为0、1记分,答对的条目总数即该分测验的原始分,各分测验原始分转换成10为平均数,3为标准差的标准分数。所有数据使用SPSS8.0和Amos4.0统计软件处理。

2 结 果

2.1 样本构成

样本的年级、性别等构成情况见表1。经 χ^2 检验,各年级男女比例、A式和B式的实测比例均无显著差异。另外,我们对被试父母的文化程度及职业

状况进行了统计,经 χ^2 检验,父、母亲的文化程度各
 年级之间无显著差异,职业状况各年级之间差异显著。

表 1 样本的年级、性别分布

		初一	初二	初三	高一	高二	高三	χ^2	P
A 式	男	68	53	60	68	59	55	1.604	> 0.05
	女	62	56	49	72	54	57		
B 式	男	60	62	52	69	67	51	6.085	> 0.05
	女	58	56	57	76	44	59		
A 式总体		130	109	109	140	113	112	1.058	> 0.05
B 式总体		118	118	109	145	111	110		

表 2 被试父母的文化程度及职业状况

		初一	初二	初三	高一	高二	高三		
父亲文化								$\chi^2= 18.64$ $P> 0.05$	
小学		7	3	5	9	8	5		
初中		42	27	45	63	36	40		
高中		101	87	101	114	99	93		
大学		87	94	64	97	76	80		
母亲文化								$\chi^2= 24.27$ $P> 0.05$	
小学		17	10	14	27	11	8		
初中		43	40	45	59	50	49		
高中		106	91	108	137	109	112		
大学		69	71	48	60	54	49		
父亲职业								$\chi^2= 108.72$ $P< 0.05$	
工人		72	59	96	120	106	90		
农民		8	12	3	33	6	6		
知识分子		42	37	22	41	27	24		
干部		42	52	49	59	50	56		
商人		67	43	34	23	27	28		
军人		7	8	8	6	5	7		
母亲职业								$\chi^2= 102.94$ $P< 0.05$	
工人		87	84	117	152	117	118		
农民		12	20	14	41	13	9		
知识分子		51	42	26	46	32	30		
干部		31	25	23	27	29	38		
商人		47	36	25	11	23	17		
军人		1	2	2	1	1	0		
无业		9	3	5	6	3	1		

2. 2 基本测试结果

A、B 式各年级各分测验的原始分均数列于表
 3。表中数据可见,各分测验原始均分随年级升高而
 呈稳定增长。

2. 3 测验条目分析结果

2. 3. 1 条目 难度 分别按年级计算 A 式、B 式各个
 分测验条目的通过率,作为条目的难度。将难度划
 分为 ≤ 0.10 、 $0.11\sim 0.29$ 、 $0.30\sim 0.69$ 、 $0.70\sim 0.$
 89 、 ≥ 0.90 五个等级,A、B 式合并统计,各个难度
 等级内包含的条目数及百分率列于表 4 中。表中数
 据可见,各分测验难度在 $0.46\sim 0.54$ 之间,难度在
 0.30 至 0.69 之间的条目占条目总数的 $40\sim 66\%$ 。

表 3 A、B 式各年级各分测验原始分(\bar{x})

	分测验	初一	初二	初三	高一	高二	高三
A 式	算术	4.58	4.35	5.61	6.85	7.02	7.41
	填词	9.42	9.64	10.55	12.66	12.64	14.11
	填数	5.14	5.04	5.83	6.52	6.57	6.86
	填句	7.05	7.50	9.87	11.30	11.53	12.38
	平面	12.48	12.58	13.42	14.86	15.29	14.93
	阅读	5.77	6.53	7.24	8.51	8.73	8.92
B 式	立体	7.40	7.24	7.91	9.00	9.15	9.67
	算术	5.33	5.53	6.11	6.95	7.41	7.40
	填词	8.52	10.36	12.12	13.97	15.01	15.55
	填数	5.36	5.71	6.25	6.70	6.87	6.73
	填句	7.66	9.03	10.16	11.34	11.56	13.03
	平面	11.57	12.16	12.58	13.38	14.50	14.16
	阅读	5.24	5.97	6.65	8.10	8.52	9.67
	立体	7.53	7.88	8.46	9.23	10.23	9.85

* 未列出标准差

表 4 各分测验条目的难度,区分度和各测验之间的相关

	指数等级	算术 ($n= 24$)	填词 ($n= 48$)	填数 ($n= 24$)	填句 ($n= 40$)	平面 ($n= 50$)	阅读 ($n= 32$)	立体 ($n= 38$)
难度	≤ 0.1	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
	$0.11\sim 0.29$	4(16.7)	14(29.2)	7(29.2)	7(17.5)	11(22.0)	9(28.1)	12(31.6)
	$0.30\sim 0.69$	16(66.7)	23(47.9)	8(33.3)	25(62.5)	20(40.0)	17(53.1)	19(50.0)
	$0.70\sim 0.89$	3(12.5)	7(14.6)	7(29.2)	7(17.5)	14(28.0)	6(18.8)	6(15.8)
	≥ 0.9	1(4.2)	4(8.3)	2(8.3)	1(2.5)	3(6.0)	0(0.0)	1(2.6)
	平均难度	0.52	0.50	0.51	0.51	0.54	0.47	0.46
鉴别指数	< 0.2	0(0.0)	7(14.6)	4(16.7)	6(15.0)	1(2.0)	0(0.0)	1(2.6)
	$0.2\sim 0.29$	3(12.5)	8(16.7)	2(8.3)	6(15.0)	6(12.0)	7(21.9)	6(15.8)
	$0.3\sim 0.39$	2(8.3)	14(29.2)	5(20.8)	12(30.0)	12(24.0)	7(21.9)	14(36.7)
	≥ 0.4	19(79.2)	19(39.6)	13(54.2)	16(40.0)	31(62.0)	18(56.3)	17(44.7)
	d 均值	0.52	0.39	0.45	0.38	0.45	0.40	0.42
相关系数	< 0.2	0(0.0)	1(2.1)	1(4.2)	6(15.0)	0(0.0)	0(0.0)	0(0.0)
	$0.2\sim 0.29$	2(8.3)	13(27.1)	3(12.5)	5(12.5)	1(2.0)	7(21.9)	5(13.0)
	$0.3\sim 0.39$	3(12.5)	18(37.5)	7(29.2)	20(50.5)	17(34.0)	14(43.8)	18(47.3)
	≥ 0.4	19(79.2)	16(33.3)	13(54.2)	9(22.5)	32(64.0)	11(34.4)	15(39.5)
	r 均值	0.45	0.35	0.40	0.34	0.43	0.36	0.39

注:分测验下为条目数,括弧内为%。

2.3.2 条目区分度 区分度分析包括计算条目的鉴别指数及条目与该测验总分的点二列相关系数。鉴别指数定义为按总分高低排列的上 27% 人群与下 27% 人群的平均通过率之差。根据美国测验学家 Ebel 提出的条目鉴别指数的评价标准^[7], 将鉴别指数分为 < 0.2 、 $0.2 \sim 0.29$ 、 $0.3 \sim 0.39$ 、 ≥ 0.4 四个等级, 分别按年级计算 A 式、B 式各个分测验条目的鉴别指数。A、B 式合并统计, 各等级内包含的条目数及百分比列于表 4 中。表中数据显示, 各分测验的平均鉴别指数在 $0.38 \sim 0.52$, 93% 的条目的鉴别指数大于 0.2。条目与总分的相关系数在 $0.34 \sim 0.45$ 。见表 4。

2.4 信度研究结果

对初一年级 105 名被试(A 式 51 例, B 式 54 例)平均间隔约 25 天重测。重测相关系数: 各分量表在 $0.68 \sim 0.92$, 测验总分 A 式为 0.87 , B 式为 0.92 。各分量表的 α 系数在 $0.64 \sim 0.84$, A、B 式总分的 α 系数分别是 0.88 和 0.89 。A、B 式总分的概化系数分别为 0.91 和 0.92 , 各分量表的概化系数在 $0.64 \sim 0.85$ 。

2.5 结构效度研究结果

2.5.1 内部相关分析 七个分测验与量表总分之间的相关: A 式为 $0.65 \sim 0.78$, B 式为 $0.62 \sim 0.80$ 。各分量表与总分的相关: A 式为 $0.78 \sim 0.89$, B 式为 $0.78 \sim 0.88$ 。语文分量表分与语文分量表中各分测验的相关($0.82 \sim 0.85$)明显高于与其它两个分量表中各分测验的相关($0.40 \sim 0.56$); 数理分量表分与数理分量表中各分测验的相关(0.88)明显高于与其它两个分量表中各分测验的相关($0.39 \sim 0.53$); 图形分量表分与图形分量表中各分测验的相关(0.84)明显高于与其它两个分量表中各分测验的相关($0.39 \sim 0.51$)。三个分量表之间约 0.5 的中度相关。

2.5.2 因素分析 采用主成分分析与方差极大正交旋转分别对 A 式和 B 式进行因素分析, 结果一致。各分测验均对一个基本能力因子中度以上的负荷, 三因子因素分析发现, 三因子模型恰好与 AATMS 结构吻合, 因子 I 负荷填词、填句、阅读分测验, 这些分测验都与语言能力有关, 命名为语文因子, 因子 II 负荷算术、填数分测验, 这些分测验都与数的能力有关, 命名为数理因子, 因子 III 负荷平面、立体分测验, 这些分测验都与图形的知觉和想象有关, 命名为图形因子。抽取的三个因子解释了总方差的 72.18% (A 式), 73.53% (B 式)。

2.6 效标关联效度研究结果

2.6.1 测验分数的年级差异比较 经协方差分析,

各分测验、分量表及测验总分在不同年级之间存在显著差异。两两比较发现, 许多相邻年级之间测验分数差异显著, 若相隔两个年级, 其测验分数几乎均有显著差异。测验得分基本上随年级升高而增长。

2.6.2 与长-鞍团体智力测验结果的相关分析

AATMS 的各分测验与龚耀先教授编制的长-鞍团体智力测验的各分测验有不同程度的相关^[13]; AATMS 的语文分量表、数理分量表与长-鞍团体智力测验的言语理解因素、知觉推理因素有显著正相关($0.38 \sim 0.56$); 两个测验的总分之间的相关 A 式为 0.73 , B 式为 0.52 。

2.6.3 与教师评定的相关分析 语文分量表分与教师对学生语文能力的评定、数学分量表分与教师对学生数学能力的评定均显著正相关; 测验总分与教师对学生能力的总体评定中度或中度以上相关($0.51 \sim 0.67$)。

2.6.4 与学习成绩的相关及回归分析 AATMS 总分与施测 AATMS 约两个月后各科统考成绩显著正相关, 与统考总分的相关为 0.68 (A 式) 和 0.53 (B 式)。分别以统考语文、数学、英语、物理成绩及总成绩为因变量, 以 AATMS 各分测验或各分量表分数为预测变量, 进行多元线性逐步回归分析。结果: 填词分测验能有效预测语文和英语成绩, 算术分测验能有效预测数学成绩, 填词和算术分测验对统考总分也具有有效预测作用。语文分量表能有效预测语文和英语成绩, 而数理分量表能有效预测数学成绩。语文和数理分量表能有效预测统考总成绩, 以统考总成绩为因变量, 以语文和数理分量表为预测变量的回归方程的决定系数为 0.46 (A 式) 和 0.29 (B 式)。

2.6.5 A、B 式平行性的考察 两式的信度指标相当, 结构相同, 各种效标效度也比较一致。有些分测验 A 式较 B 式难些, 其余分测验难度无差异。综合起来看, A、B 式大体平行。

3 讨 论

3.1 关于 AATMS 的建构

编制 AATMS 的基本目的是要评量中学生的基本学习能力倾向, 为此应围绕完成中学各科学学习所需要的基本能力来考虑量表的建构, 但实际上, 完成中学各科学学习所需要的能力很多, 一一考察, 既不可能也无必要。鉴于语文和数学是中学各科中最重要也是最基础的学科, 所以我们是与语文和数学学习密切相关的能力为基点建构量表。围绕语文和数学能力结构国内外有一些研究^[14-16]。我们也就此

进行了专家咨询和开放式问卷调查, 这些研究和调查基本上都把数字运算能力、逻辑推理能力和空间想象能力作为数学能力结构中的重要因素。把听、说、读、写能力即语言的理解、表达和应用能力作为语文能力结构的重要因素。另外, 我们查阅了一些国外和台湾地区的学习能力测验^[8, 10], 这些量表基本上是以上述能力为主要内容进行建构的。综合借鉴和考量我们提出了学习能力结构模型, 并确定了 AATMS 的语文、数理、图形三个分量表共含七个分测验的量表结构。从测量结果看, 各分测验与测验总分有较高相关, 每个分量表与其所含分测验的相关明显高于与其它分量表所含分测验的相关, 三个分量表之间有中度相关。说明 AATMS 的建构比较合理。

3.2 条目质量

一个测验的信度和效度在很大程度上取决于组成测验的条目的性能, 所以, 条目分析是编制和修订标准化测验的重要环节, 基本目的在于选取难度适中并具有鉴别力的条目, 以使测验能有效区分被试在所测特质上的差异。从理论上讲, 难度为 0.5 左右的条目具有最大区分能力, 但实际编制测验时, 应使条目难度分布广些, 最好为常态分布, 条目的平均难度应为 0.5 左右^[7]。对 AATMS 条目难度分布的统计结果表明, 绝大多数分测验的条目难度分布较为理想, 难度在 0.30~0.69 的条目占条目总数的 40~66%, 各难度等级的条目比例也较为适宜, 近似正态分布, 各分测验的平均难度均在 0.5 左右。但填数分测验中等难度条目比例稍小, 立体分测验较难条目比例偏大, 这些问题有待于进一步修改测验时改进。从不同年级看, 测验难度对初一、初二年级偏高, 对高中学生较为适宜。这种年级间的难度差异, 一方面说明测验分数随年级升高而增长, 成为测验的一种效度支持, 另一方面也符合编制测验的初衷, 即该测验主要作为中学高年级学业发展咨询的工具, 也可作为中学低年级选拔学习潜能优秀学生的工具。测量学专家认为条目的鉴别指数达 0.2 即可接受^[7], 0.3 以上较为理想^[17]。从总体看, 93% 的条目鉴别指数在 0.2 以上, 96% 的条目与分测验总分的相关系数在 0.2 ($P < 0.01$) 以上, 表明绝大多数条目均具有有效区分度。各分测验的鉴别指数在 0.38~0.52, 均较理想, 算术、平面、立体和阅读分测验中几乎所有条目的鉴别指数均在 0.2 以上, 但填词、填句和填数分测验中鉴别指数低于 0.2 的条目比例稍高, 这是需要进一步修改测验时考虑改善的地方。

3.3 测验的信度和效度

信度是测验稳定性和可靠性指标。对 AATMS 的重测信度及内部一致性信度的检验结果表明, 与同类团体测验相比^[8, 10, 12], AATMS 的信度指标接近或优于其相应指标。可以说, 整个测验作为测量学业能力的工具是可靠的。概化系数的研究也说明整个测验稳定性较好, 测量误差较小, 质量较高^[18]。

结构效度用来考察一个测验是否测量了设计时的理论构想, 测得程度如何。对 AATMS 因素分析结果表明, 各分测验均对一个基本能力因子中度以上的负荷, 抽取的语文、数理及图形三个因子共解释了总方差的 72.18% (A 式), 73.53% (B 式), 恰好解释了 AATMS 的结构。AATMS 具有理想的结构效度。

效标关联效度是对一个测验对于特定情境中的个体行为进行估计的有效性的考察, 学习成就、智力测验分数、教师评定、年级差异是学业能力倾向测验常用的效标。本研究分别以长-鞍团体智力测验、教师评定等级、对不同年级的区分以及学习成绩为效标对 AATMS 的效标效度进行了初步检验, 结果表明: ①AATMS 总分与长-鞍团体智力测验总分之间呈中度或以上的相关; ②测验分数与教师评定之间具有相当程度的一致性; ③AATMS 各分测验及各分量表及总分年级间差异显著, 测验分数随年级增长呈稳定上升趋势, 且测验分数年级间的差异特点与林崇德教授等对学习力发展的实验研究结果基本吻合^[19]; ④AATMS 总分与统考的各科成绩显著正相关。不同的分测验或分量表对不同的学科成绩的预测力是有差异的, 其总分对各科及统考总分均具有显著的预测力。上述结果均可对 AATMS 有利的效度支持。

综上所述, 本研究初步完成了中学生学业能力倾向测验的初步编制, 对该量表的各项心理测量学指标检验表明基本达到了编制目的, 尽管如此, 对该量表的质量尚需进行更广泛而深入的研究和改善。

参 考 文 献

- 1 Brown HE. Principles of educational and psychological testing (3rd ed). New York: Hold, Tinehart, & Winston, 1983
- 2 Aroll JB. The aptitude-achievement distinction: The case of foreign language aptitude and proficiency, In Green DR (Eds.), The aptitude-achievement distinction: proceedings of the second CTB/ McGraw-Hill conference on issues in Educational Measurement. Monterey, California: CBT/ McGraw-Hill, Inc., 1974

究获得 6 个和 9 个一阶因素, 解释的方差只有 28% 和 34%, 说明还有一些有意义的因子未被提取。强制性四因素(特征根大于 2) 分析显示, N 量表的项目(除第 67 项外) 和 E 量表的项目(除第 29 项外) 都分别符合与 N 因子和 E 因子, 例外的 1 个条目在原维度上也有一定的负荷, P 量表的项目中有 6 个不负荷于 P 因子, 其中有 4 个负荷于 N 因子, L 量表的项目中有 5 个不负荷于 L 因子, 其中有 4 个负荷于 P 因子。Barrett 和 Kline 所做的四因素分析也有类似的现象, 如 L 量表有 4 个条目负荷其它因子, P 量表有 11 个条目负荷于其它因子^[2, 3]。这种现象说明 N 和 E 量表的因素结构比较稳定, 在不同的样本中能得到很好的重复, 具有较高跨文化效度, 而 P 和 L 量表稳定性差些, 同一项目对不同人群的意义不一样, 测量到的人格特质也不同, 这提示我们在解释 P 和 L 量表得分的意义时要考虑被试的文化背景。

查阅文献没有发现 EPQ 验证性因素分析的论文, 我们用验证性因素法对 EPQ 四因素理论模型做了验证。与独立模式相比, 研究模式 χ^2/df 和 RMSEA 明显降低, RMSEA 值低于通常的 0.05 的标准, 达到比较满意的水平, χ^2/df 虽然比较小, 但仍然达到显著性水平, 这可能与样本量比较大有关, 其它验证性因素分析也有类似情况; MECVI 接近饱和模式, 表示有较好的交互效度; 增值指数接近 0.95, 省俭指数在 0.85 以上, 按有关标准来判断, 该模式基本上可以接受。从各项目在相应因子上的负荷上来看, N 量表各项目在 N 因素上负荷在 0.31~0.61 之间, E 量表各项目在 E 因素上的负荷在 0.30~0.67 之间, 说明这两个量表具有较好的结构效度; P 量

表中有 6 个项目在 P 因素上的负荷小于 0.20, L 量表中有 3 个项目在 L 因素上的负荷小于 0.20。从探索性因素分析也可以看出这些项目负荷于其它因素, 说明这两个量表的某些项目有待于改进。

各因子间的相关见附图(方框内数字为项目的序号, 小数为项目负荷和相关系数), 四个因子间有一定的相关, 其中 P 量表与 L 量表、P 量表与 N 量表间的相关相对较高, 各项目在相应因子都有一定的负荷, 个别项目负荷较低。由于 P 量表与 L 量表、P 量表与 N 量表间的相关相对较高, 所以这些分量表间的项目有些交错现象, 这在探索性因素分析中已得到证实, 这种现象是可以理解的, 因为人格本身是一个整体, 各特质维度间必然存在一定的相关, 有些项目可能测到多个维度的特质, 用不同样本所做的因素分析时项目可能出现错位现象。

参 考 文 献

- 1 Eysenck HJ. Manual of Eysenck Personality Questionnaire. London: Hodder & Stoughton, 1975
- 2 Barrett PT, Kline P. Personality factors in Eysenck Personality Questionnaire. *Personality and Individual Differences*, 1980, 1: 318- 333
- 3 Barrett PT, Kline P. Radial parcel factor analysis. *Personality and Individual Differences*, 1980, 2: 311- 318
- 4 Eysenck HJ. Biological dimensions of personality. In L. A. Pervin (Ed), *Handbook of Personality: Theory and Research*, New York: Guilford, 1990, 244- 276
- 5 Eysenck HJ. Dimensions of personality: 16, 5 or 3? Criteria for taxonomic paradigm. *Personality and Individual Differences*, 1991, 12: 773- 790
- 6 龚耀先. 修订艾森克个性问卷手册, 长沙: 湖南医学院, 1981
(收稿日期: 2003- 09- 06)

(上接第 5 页)

- 3 Anastasi A. *Psychological testing* (6th ed). New York: Macmillan, 1990
- 4 Kevin RM, Davidshofer CO. *Psychological testing*. Prentice- Hall, Inc, 1994
- 5 Cronbach LJ. *Essentials of psychological testing* (5th ed). New York: Harper & Row, Publishers, 1990
- 6 Hopkins KD, Stanley JC, Hopkins BR. *Educational and psychological measurement and evaluation* (7th ed). New Jersey: Englewood Cliffs, 1990
- 7 戴海崎, 张 锋, 陈雪枫. *教育心理测量*. 广州: 暨南大学出版社, 1999
- 8 James V, Mitchell JR. *The ninth mental measurements yearbook. The Buros Institute of Mental Measurements of The University of Nebraska- Lincoln*, 1985
- 9 中国测验学会. *华文社会的心理测验*. 台北: 心理出版社, 1994
- 10 陈明终, 许胜哲, 吴清山. *我国心理教育测验汇编*. 高雄: 高雄复文书局, 1988
- 11 张月娟, 龚耀先. 学业能力倾向测验综述. *心理发展与教育*, 2002, 18(1): 92- 95
- 12 林传鼎, 张厚粲, 周 谦, 等. *少年儿童学习能力测验的编制*. 北京师范学院教育研究所, 1979
- 13 龚耀先. *长- 鞍团体智力测验手册*. 湖南医科大学, 1997
- 14 莫 雷. 中小学生学习语文阅读能力结构的发展特点. *心理学报*, 1992, 24(4): 346- 353
- 15 罗照盛, 张厚粲. 中小学生学习语文阅读理解能力结构及其发展特点研究. *心理科学*, 2001, 24(6): 654- 656
- 16 陈任泽. 数学学习能力的因素分析. *心理学报*, 1997, 2: 172- 176
- 17 Aiken LR. *Psychological testing and assessment* (sixth ed.). Allyn and Bacon, Inc., 1988
- 18 漆书青, 戴海崎, 丁树良. *现代教育与心理测量学原理*. 南昌: 江西教育出版社, 1998
- 19 林崇德. *学习与发展*. 北京: 北京师范大学出版社, 1999
(收稿日期: 2003- 03- 28)