

多项能力倾向测验的初步编制

王进礼^{1,2}, 龚耀先¹

(1. 中南大学湘雅二医院医学心理研究中心, 湖南 长沙 410011; 2. 武警医学院, 天津 300162)

【摘要】 目的: 把长 - 鞍团体智力测验发展成为多项能力倾向测验以供职业选拔之用。方法: 根据工作分析与借鉴范例来建构 MAT。随机整群在武警湖南总队取 991 人实测。对其中 45 人相隔 23 天重测, 进行条目及信效度分析。结果: 各分测验难度在 0.38 ~ 0.60 之间, 总体难度是 0.50, 具有有效区分度 (0.20 以上) 的条目占 85 % 以上。重测相关系数为 0.79, 总分 系数为 0.80。效度检验结果符合心理测量学结果。结论: 多项能力倾向测验的建构合理, 达到了编制目的。

【关键词】 能力倾向测验; 条目分析; 武警部队

中图分类号: G449.5

文献标识码: A

文章编号: 1005 - 3611(2004)02 - 0116 - 05

Constructing a Multidimensional Aptitude Test (MAT)

WANG Jin - li, GONG Yao - xian

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

【Abstract】 **Objective:** A Multidimensional Aptitude Test (MAT) was developed to provide an objective measure of general aptitude for personnel - selection and placement. **Methods:** We formed the structure of MAT through job analysis. A total of 991 Armed Police Force men were tested, 45 of whom were re - tested after 23 days. **Results:** The difficulty of sub - tests ranged from 0.38 to 0.60, while the average difficulty of the whole scale was 0.50. The discrimination of 85 % items was over 0.2. Reliability and Validity meets the criteria of psychometrics. **Conclusion:** The reliability and validity of MAT were supported in the present study.

【Key words】 Aptitude test; Item analysis; Armed police force

能力倾向 (aptitude, 或称性向) 是指一个人经过一定的训练或置于适当的环境下完成某项任务的可能性和水平^[1,2]。个人的能力倾向有特点, 人们之间的能力倾向有差异, 能力倾向测验是测量这些特点和差异的手段之一。能力倾向测验是从智力测验的演进中产生和发展的^[3]。智力测验和能力倾向测验都是测量心理能力。一般认为, 智力测验侧重测量一般心理能力; 而能力倾向测验侧重测量某些特殊能力^[4]。测量这些特殊能力的测验又有测量某单一能力和测量多重 (多项) 能力之分。国外的能力倾向测验较多, 国内尚很少, 见到的有刘锦山参照美国教育与工业测验服务中心编制的《职业能力安置量表》编制的《BEC 职业能力测验》^[5], 方俐洛等修订的美国《一般能力倾向成套测验 — GATB》(删除用器械的分测验)^[6]。我们试图以龚耀先等编制的《长 - 鞍团体智力测验》^[7]为基础, 增加一些分测验, 将其发展成为一个能测量一般智力和几种能力倾向的成套测验即多项能力倾向成套测验 (Multipolar Aptitude Test, MAT)。以期此测验对甄别就业或入伍人员的能力倾向有所帮助。本文是这一工作的初步报告。

1 方法

1.1 编制预测题本

本量表的编制借鉴国内外成功的能力倾向测验模式, 结合采用从实际应用出发的策略^[8,9], 条目是以长 - 鞍团体智力测验为基础, 又构思了一些新的测验条目构成的。其中 8 个分测验采用了长 - 鞍团体智力测验 (甲式) 的条目, 又重新增编了其它 9 个分测验, 形成试测本, 适宜团体实施。

1.2 编制实测题本

用试测本在某省武警总队某支队 250 人中进行试测, 对试测结果作了条目的难度和区分度分析, 淘汰了不合要求的条目, 按难度将条目排序, 去掉了与效标相关相对偏低的一个注意测验和一个文字校对测验, 然后编成由 15 个分测验构成的实测题本。同时也确定了每个测验的时限 (表 1)。

1.3 实测样本

实测样本采取随机整群取样方法在某武警总队抽取, 全部取样工作由一人主试。取样工作在 2002 年 10 月至 12 月完成。共测查 991 人, 获得有效试卷 974 份, 全部为男性, 年龄: 平均 21 ± 2.55 岁, 最小 17 岁, 最大 39 岁; 衔级: 列兵 277 人, 上等兵 306

人,士官 348 人,干部 43 人;民族:汉族 894 人,其它民族 80 人;兵源:农村 602 人,城镇 372 人;教育程度:大专以上 37 人,高中或中专 610 人,初中 319 人,8 人不详;职业:司机 95 人,文书 33 人,公务员 18 人,警卫值勤类 714 人,其它 114 人。

表 1 MAT 的内容、项目数和时限

分量表	分测验	项目数	时限
抽象思维	常识	23	3
	文字分类	24	2
	图案分类	21	2
	数字接龙	23	6
	图案接龙	18	6
空间能力	折叠 *	14	4.5
	拼配 *	16	3.5
	补缺	16	3
	文字校对 *	56	5
注意和速度	数字校对	58	4.5
	图形选择 *	36	2
	编码	296	5
	打点速度 *	528	2
动作稳定	左手描线 *	1	4
	右手描线 *	1	4
合计		1146	56.5

注: *系新加测验,余系沿用《长一鞍团体智力测验》

1.4 测验记分方法

测验均为 0、1 记分,即答对记 1 分,答错记 0 分。各分测验的原始分都转化为标准 Z 分数。量表总分是所有分量表标准分之和。

测验 14 及 15 两个描画线条的分测验,采用了特殊的记分方法,首先用扫描仪将被试描画的线条与标准线条一起扫进电脑,然后记录每一对应的描画的点和标准的点,以像素为单位计算它们的差数,即描画点对标准点的偏离数,再计算出所有这些差数的平均值作为测验分数。

2 结 果

2.1 测验基本结果

各分测验原始均分依次为常识 10.79、文字分类 13.72、图案分类 11.75、数字接龙 14.44、图案接龙 10.54、折叠 6.12、拼配 6.95、补缺 6.26、打点速度 289.18、文字校对 38.69、图形选择 43.03、数字校对 57.23、编码 151.89、左手描线 0.63、右手描线 0.51。

2.2 条目分析

知觉注意测验的 5 个分测验属于纯速度测验,描画线条测验整个测验都是同一的手部动作,两者都无法作出有意义的项目分析,故省略。对其它分测验的条目进行了难度和区分度分析。

2.2.1 难度分析 本研究按常规做法将被试在测验条目上的通过率作为难度的指标,并将其分为 0.10、0.11~0.29、0.30~0.69、0.70~0.89、和 0.90 五个等级,把各分测验的各个条目的难度及百分数都划入相应等级之中。结果显示,各分测验难度在 0.38~0.60 之间,平均难度是 0.50。见表 2。

表 2 各分测验的难度等级分布

难度等级	常识	文字分类	图案分类	数字接龙	图案接龙	折叠	拼配	补缺
0.10	4(17)	2(8)	3(14)	2(8)	0(0)	1(7)	3(19)	4(25)
0.11~0.29	4(17)	3(1)	3(14)	2(8)	3(25)	4(29)	3(19)	3(19)
0.30~0.69	9(39)	10(42)	9(43)	9(39)	6(40)	7(50)	6(38)	7(44)
0.70~0.89	5(22)	7(29)	5(24)	10(43)	9(50)	2(14)	4(25)	2(13)
0.90	1(4)	2(8)	1(5)	0(0)	0(0)	0(0)	0(0)	0(0)
平均难度	0.46	0.57	0.55	0.60	0.57	0.43	0.43	0.38

注:括号内是人数百分数,下同。

表 3 各分测验的鉴别指数等级分布

鉴别指数等级	常识	文字分类	图案分类	数字接龙	图案接龙	折叠	拼配	补缺	总数
<0.20	4(17)	4(17)	3(14)	1(4)	1(6)	4(29)	3(19)	4(25)	24(15)
0.20~0.29	3(13)	4(17)	3(14)	6(26)	1(6)	3(21)	3(19)	0(0)	23(15)
0.30~0.39	2(9)	7(29)	6(29)	2(9)	2(11)	1(7)	3(19)	3(19)	26(17)
0.40	14(61)	9(38)	9(43)	14(61)	14(78)	6(43)	7(44)	9(56)	82(53)
D 均值	0.43	0.41	0.35	0.49	0.51	0.36	0.36	0.35	

2.2.2 区分度分析 本研究做了两方面的区分度的考察,首先是分析鉴别指数(表 3)。另一种是计

算每个条目与其所属的分测验总分的相关(表 4)。表 3 显示,85 % 的条目的鉴别指数在 0.2 以上,各个

分测验的平均鉴别指数在 0.35 ~ 0.51 之间。表 4 显示:除补缺分测验条目与总分相关低于 0.2 的占 15 % 外,其余各分测验条目与总分的相关高于 0.2 的占 95 % 以上,各分测验条目与总分的相关均值在 0.31 ~ 0.51 之间。

表 4 各分测验条目与总分的相关分布 (r 值)

r 等级	常识	文字分类	图案分类	数字接龙	图案接龙	折叠	拼配	补缺
< 0.20	0(0)	0(0)	1(5)	0(0)	1(6)	2(14)	1(6)	4(25)
0.20 ~ 0.29	1(4)	2(8)	4(19)	2(9)	1(6)	3(21)	3(19)	0(0)
0.30 ~ 0.39	9(39)	5(21)	4(19)	2(9)	1(6)	4(28)	5(31)	6(38)
0.40	13(57)	16(67)	12(57)	19(83)	15(83)	5(36)	7(44)	6(38)
D 均值	0.42	0.43	0.31	0.51	0.47	0.33	0.35	0.33

表 5 测验的重测相关

	初测	重测	相关值 r
常识	9.27 \pm 3.54	12.87 \pm 3.65	0.61 **
文字分类	13.62 \pm 3.56	15.42 \pm 3.90	0.61 **
图案分类	10.78 \pm 3.52	12.98 \pm 3.02	0.59 **
数字接龙	13.44 \pm 5.04	16.28 \pm 4.70	0.67 **
图案接龙	10.91 \pm 3.48	11.57 \pm 2.84	0.48 **
折叠	5.76 \pm 2.19	6.71 \pm 2.04	0.59 **
拼配	6.65 \pm 2.30	7.82 \pm 1.87	0.47 **
补缺	6.00 \pm 2.60	7.02 \pm 2.41	0.54 **
打点速度	282.52 \pm 36.15	313.45 \pm 34.98	0.78 **
文字校对	32.48 \pm 7.70	39.70 \pm 9.99	0.54 **
图形选择	38.98 \pm 7.40	48.89 \pm 9.44	0.60 **
数字校对	53.30 \pm 14.24	59.07 \pm 13.50	0.86 **
编码	143.18 \pm 27.68	158.43 \pm 40.26	0.63 **
左手描线	1.20 \pm 0.19	1.21 \pm 0.21	0.88 **
右手描线	0.88 \pm 0.44	0.88 \pm 0.48	0.85 **
抽象思维	50.40 \pm 23.9	58.58 \pm 28.20	0.80 **
空间能力	15.70 \pm 8.56	19.40 \pm 7.94	0.80 **
注意和速度	407.76 \pm 50.76	456.15 \pm 61.32	0.73 **
动作稳定	2.08 \pm 0.53	2.10 \pm 0.58	0.88 **
总分	485.32 \pm 65.43	547.31 \pm 74.31	0.79 **

注: * $P < 0.05$, ** $P < 0.01$, 下同。

2.3.2 Cronbach' 系数 各分量表的系数分别是注意和速度 0.75, 抽象思维 0.72, 空间能力 0.53, 动作稳定 0.75, 总分 0.80。

2.4 效度检验

2.4.1 结构效度 内部相关分析:先考察了各分测验分量表之间及它们与总分的相关。从各分测验与总分之间的相关看:左手描线、右手描线和打点速度分测验与总分的相关低于或等于 0.4, 其余分测验在 0.55 ~ 0.64 之间。各分量表与总分的相关:双手描线为 0.289, 其余三个分量表在 0.694 ~ 0.807 之间。各个分量表与它们所包含的分测验的相关明显高于与其它分测验的相关。在四个分量表中, 抽象思维、空

2.3 信度检验

2.3.1 重测相关 对 2 个中队 45 名被试相隔 23 天后重测。各分测验的重测相关系数在 0.47 ~ 0.88 之间 ($P < 0.01$), 四个分量表的重测相关系数在 0.73 ~ 0.88 之间, 总分为 0.79 (表 5)。

间能力、注意和速度三个分量表相关在 0.402 ~ 0.472 之间, 双手描线分量表与这三个分量表相关在 0.1 以下。

探索性因素分析:采用主成分分析与 Equamax 相等最大值转轴法进行因素分析, 得到 4 个特征值大于 1 的因子, 分别是:因子 1 抽象思维因子; 因子 2 注意和速度因子; 因子 3 空间能力因子; 因子 4 动作稳定因子。四个因子所含的分测验与前面各分量表所含分测验完全一致。这四个因子解释了总方差的 54.28 %。

验证性因素分析:使用 Amos 统计软件, 根据验证性因素分析的原理, 用测量的数据拟合 MAT 的结构模型, 结果显示, 拟合指数 CFI、NNFI 均在 0.95 以上, RMSEA 均小于 0.05, CMIN/DF 小于 2。

2.4.2 效标效度 群体区分效度:文书、公务员、司机三种不同群体与总体的比较, 总分分别为 3.72 \pm 6.87, 3.08 \pm 3.00, - 2.29 \pm 6.88, 差异非常显著。

对不同学历被试的测验分数比较:不同学历的群体除描画线条外的所有测验分数差异都显著。年龄与 MAT 各分测验、分量表及测验总分的相关:年龄与常识分测验和抽象思维分量表相关虽然显著但极微弱, 其它的均不显著且相关系数极低。年龄与数字校对、图形选择、打点速度、图案分类、折叠、拼配、补缺、描画线条分测验, 注意和速度、空间能力、动作稳定分量表存在微弱负相关。MAT 与训练成绩的相关:武警部队最基本的三项训练有擒拿技巧、军事体能与队列动作, 我们用六个中队 136 人的三项训练成绩与测验做相关分析, 结果显示:两总分间有显著相关:0.22。MAT 与韦氏成人智力测验的相关:用韦氏成人智力测验四合一简式与 MAT 做相关分析^[10], 结果两总分间相关系数 0.64。

3 讨 论

MAT 是出于实践的需要,把龚耀先等的长 - 鞍团体智力测验发展成为多项能力倾向测验,以满足武警部队及相关团体人员选拔与安置等方面急需而编制的。下面讨论 MAT 的内容问题、难度、区分度、信度和效度问题。

3.1 MAT 内容问题

编制 MAT 是供团体测验使用的,许多团体都需要动作能力倾向的测量,动作能力倾向的测量一般都要使用器械,可是团体测验的施测方式又不便于使用器械。为了能够不使用器械,只用简单的纸笔形式测量到某些动作能力,本测验除了测验思维、注意等方面的内容外,还参照原属神经心理测验的动作精确性测验^[11]编制了适合电脑扫描的描画线条测验,用来测量手的动作的稳定性和准确性。在儿童心理测验也有类似测验使用^[12,13]。原先的记分方法是用地图尺人工测量被试所描的线条与标准线条偏离的长度,方法比较粗糙,只能量出特别大的偏离,对正常人的微小差别测量不出来,而且主试手持着地图尺跟踪被试的轨迹时还产生自己的手的动作偏离的误差,这就更不能鉴别正常人的双手动作稳定性和精确性了。我们首次尝试使用现代科技手段改善记分方法,把作为标准用的线条改造成适合微机处理的图形,用扫描仪扫描被试所描画的轨迹,以像素为单位记录被试偏离的平均数和标准差,这样就能把正常被试手的动作的微小偏离识别出来。而恰好是这种微小偏离的差异暴露出个体动作稳定性和准确性方面的区别,这种区别对于团体的工作任务有重要意义,这种方法的改进使得描画线条测验的应用能够拓展到职业测验领域;更重要的是这种拓展能够用纸笔测验部分地代替某些器械测验,使进行心理运动某些动作的大规模团体测量成为可能。有意义的是,效度研究显示描画线条与训练成绩和领导评定分数相关显著,这证实了该设计的功效。

文字校对测验。我们使用生造错别字或死字来做文字校对测验的测验材料,以排除学习记忆的作用,考察被试实际具有的对方块汉字的知觉辨析能力。在设计文字校对测验时,最初我们通过查阅汉字频率词典检索了一组高频常见字和一组低频字做测验材料,后来发现,高频常见错别字辨析是小学、中学语文的重要学习内容,所以我们认为如果使用它们做校对材料,测查出的更多是以前学习和记忆的结果,这样这个测验就变成了成就测验,这显然是

不合适的。几经考虑,我们决定采用几乎谁都没见过的变形汉字或低频字做材料,保证了对所有被试的文化公平,测查出被试对方块汉字的迅速而准确的知觉能力。为了保证这个测验的质量,曾先后五易其稿,最后保留常见错别字校对和不常见错别字校对两种材料做预测验,预测验结果显示在职业区分方面不常见错别字优于常见错别字,这样才最后确定使用不常见错别字的文字校对测验。

3.2 条目分析

各测验的平均难度在 0.43 到 0.57 之间,难度等级在 0.30 ~ 0.69 的条目占 38% ~ 50%,难度等级过高的条目略多,有 19 个,占 12.25%。

从鉴别指数看,85%的条目都具有有效区分作用,各个分测验的平均鉴别指数在 0.35 ~ 0.51 之间;从各分测验条目与总分的相关看,具有有效区分效度的条目除了补缺(76%)和折叠(85%)偏低外,其余在 94%以上,各分测验条目与总分的相关均值在 0.31 ~ 0.51 之间。考察结果说明本测验的大多数条目具有有效区分效度。

3.3 信度

除知觉注意分量表重测信度为 0.73,其余在 0.79 ~ 0.88 之间,表明 MAT 有较好的跨时距稳定性。除空间能力分量表的系数较低,为 0.53,其余都在 0.72 ~ 0.80 之间。这些信度指标与同类测验的相应指标水平基本一致^[14,15]。

3.4 效度

从分测验、分量表之间及与总分的相关分析结果来分析,除描画线条外,其余各分测验与总分均有中度或以上的相关。一个显著的结果就是 MAT 在分量表之间、分量表与总分之间出现了相关大小不均衡的情况,即:抽象思维,空间能力,注意和速度三个分量表中度相关,且均与总分有较高的相关,抽象思维,空间能力,注意和速度三个分量表与动作稳定分量表相关很低,且后者与总分相关很低,可以说明不同分测验测量了不同的能力倾向。

从因素分析来看,MAT 的注意和速度、抽象思维、空间能力、动作稳定性 4 个因子较好地解释了 15 个分测验,同经典的 GATB 结构一致。

用验证性因素分析的方法,用数据拟合 MAT 的结构模型,各拟合指数均达到统计测量学的要求^[16],也表明 MAT 的结构模型得到了很好的验证,MAT 具有理想的结构效度。

本研究结果显示,总分、各分量表分及大部分分测验分数都对文书、分务员和司机等 3 种基本武警

职业群体有显著区分作用。不同学历的被试在测验分数上呈现显著差异,学历高的被试的测验分数显著高于学历较低的被试,这与其他研究结果一致^[17]。其原因可能是互为因果的:能力较强的个体比较可能升入大专以上学校,另一方面,个体受了大专以上教育学业能力得到更好的发展。

测验分数与训练成绩的相关显示,数字校对、打点速度、左手描线、图案接龙分测验,抽象思维、双手描线、注意和速度分量表以及总分都达到了显著水平,虽然相关指数并不很高,但也达到同类测验的水平,(能力倾向测验预测效度多在 0.20 与 0.30 之间,很少有超过 0.30 的^[4,9])能够证实 MAT 的效度。

MAT 与韦氏成人智力测验的相关表明,韦氏总分与 MAT 各分量表有从 0.537 ~ 0.667 的相关,各个分测验与 MAT 中相应的分测验或分量表都有中等水平的显著相关,这说明两套测验都测量了一般能力,但测量的能力倾向水平有所不同。

参 考 文 献

- 1 Brown HE. Principles of Educational and Psychological Testing (3rd ed). Hold, Tinehart & Winston, New York, 1983
- 2 Aroll JB. The aptitude - achievement: The Cas of Foreign Language Aptitude and Proficiency. In Green DR. The Aptitude - Achievement Distinction: Proceedings of the Second CTB/McGraw - Hill Conference on Issues in Educational Measurement. Monterey, Calif: CBT/McGraw - Hill, Inc., 1974
- 3 Anastasi A. Psychological Testing (6th ed). Macmillan, New York, Publishing Co. Ink, 1990. 476
- 4 龚耀先. 心理评估. 北京:高等教育出版社, 2003. 137 - 148
- 5 刘锦山. 人才素质测评. 北京:中国工商联合出版社, 2001. 311 - 382
- 6 方俐洛,凌文轩,刘大维,等. 一般能力倾向测验中国城市版的建构及常模的建立. 见:中国测验学会主编. 华文社会的心理测验. 台北:心理出版社, 1994. 351 - 367
- 7 龚耀先. 长 - 鞍团体智力测验手册, 湖南医科大学, 长沙, 1997
- 8 Paterson DG, Elliott RM, Anderson LD, et al. Minnesota Mechanical Ability Tests. Minneapolis: University of Minnesota Press, 1930
- 9 Aiken LR. Test and Examination. New York, 1998, 140 - 144; 154 - 175; 166 - 168
- 10 龚耀先,林传鼎,张厚粲,等. 中国修订韦氏智力量表简式手册. 湖南医学院, 1986. 9 - 16
- 11 Ayres AJ. Southern California Sensory Integration Tests. Western Psychological Services, USA, 1972
- 12 Blau TH. The Psychological Examination of the Child, New York, 1991. 136 - 139
- 13 Henderson SE, Sugden DA. Movement assessment battery for children. 24: 448 - 460
- 14 James V, Mitchell JR. The Ninth Mental Measurements Yearbook. The Bureau of Mental Measurements of The University of Nebraska - Lincoln, 1985. 137 - 138; 563 - 565
- 15 陈明终,许胜哲,吴清山. 我国心理教育测验汇编. 高雄复文书局, 1988
- 16 孟庆茂,侯杰泰. 协方差结构模型与多层线性模型原理及应用. 北京师范大学心理计量与统计分析教研室, 2001. 31 - 50
- 17 洪 炜,龚耀先. 一般行政能力倾向测验的建构及信度、效度研究. 中国临床心理学, 2000, 8(1): 1 - 6

(收稿日期: 2004 - 01 - 15)

《中国临床心理学杂志》2005 年征订启事

《中国临床心理学杂志》是经国家科委批准由中国心理卫生协会主办的中央级学术性刊物,国内外公开发行。本刊为中国科技论文统计分析源期刊、中国医学核心期刊。主要内容涉及心理治疗与咨询、心理和教育测量、神经心理、健康心理、病理心理、病人心理、儿童青少年学习和行为问题等。主要面向读者为临床心理学工作者、心理学工作者、医务人员、儿童保健、教育、司法鉴定、学校卫生、人事管理、人才开发以及其他特殊职业(体育、航空、军事等)工作人员等。

本刊为季刊,大 16 开本。从 2004 年起再次扩版,由原来的每期 80 页增至每期 110 页。每期定价 12.0 元,全年 48.0 元,全国各地邮局均可订阅,邮发代号 42 - 122。当地邮局订阅不便或因故漏订者,可与本刊编辑部联系补订,每本另收 3 元挂号邮寄费。

地 址:长沙市人民中路 86 号中南大学湘雅二医院《中国临床心理学杂志》编辑部

邮政编码:410011 电话:(0731)5524222 - 2662 E-mail: zlcjy@chinajournal.net.cn