

· 心理学研究 ·

心理学研究生入学统考四年试卷 内容与能力结构分析*

戴步云, 罗照盛

(江西师范大学 心理学院, 南昌 330027)

摘要: 我国的心理学专业研究生入学考试实行全国统考已经四年了。四年试卷质量如何, 值得关注。通过分析四年试卷所测内容的稳定性及其与大纲的一致性、所测能力结构的稳定性与合理性, 可对四年试卷的内容效度做出评价。分析表明: 考题的内容结构基本符合大纲的要求, 且相对稳定; 所考查的能力结构有明显的学科差异, 年度间也不稳定, 尚有需要改进之处。

关键词: 心理学; 研究生入学统考; 内容结构; 能力结构

中图分类号: B84-069

文献标识码: A

文章编号: 1004-8332(2011)01-0103-04

一、引言

从2007年起, 我国的心理学专业研究生入学考试实行全国统一考试(以下简称“统考”), 绝大多数具有心理学硕士点的学校都参加统考。心理学统考的专业考试为一份试卷, 由教育部考试中心统一命题, 在每次考试前, 教育部考试中心都会发布“全国硕士研究生入学统一考试心理学专业基础综合考试大纲”(以下简称“大纲”)。大纲规定:“全国硕士研究生入学心理学专业基础综合考试是为高等院校和科研院所招收心理学学科的硕士研究生而设置的具有选拔性质的统一入学考试科目。其目的是科学、公平、有效地测试考生掌握心理学学科大学本科阶段专业基础知识、基本理论、基本方法的水平和分析问题、解决问题的能力。”^[1] 这既是命题的方向和依据, 也是考生进行专业复习的向导。

截至2010年, 全国心理学统考已经实行四年。统考试卷质量如何, 测量目标和试卷结构是否稳定、是否符合大纲要求, 试卷是否具有较宽的知识内容覆盖面, 是否测量了科学合理的能力结构, 这些问题都值得关注。本文拟通过分析四年试卷所测内容结构的稳定性及其与大纲的一致性、所测能力结构的稳定性与合理性, 从内容效度的角度对四年试卷的质量做一个评价。^[2-4]

二、研究方法

研究材料: 2007年至2010年的统考试卷各1份, 2007年至2010年的大纲各1份。

研究过程: 对试卷所测的内容结构和能力结构进行考查, 既分析试卷是否符合大纲和学科要求, 又比较四年试卷结构的发展与变化。

每个试题所测内容与能力的认定工作采用的是定性分析方法。由数名心理学研究生彼此独立地将4份试卷的所有题目逐一与大纲相比较, 评定每道题目所考查的内容领域

和所考查的能力。评定中, 如果认为有一道题涉及到了两个科目, 那么就认为这两个科目各考了半道题; 以该题所测最高能力层次为它所测能力。有7名来自7所不同高校的研究生参与对试卷的评定。

在评定者们将评定结果提交给笔者后, 如果评定者们对某道题的评定发生意见分歧, 则笔者采纳占人数优势的意见。对于个别难以界定的题目, 则由一位心理测量学教授为这些题做最终评定。

三、结果与分析

1 四年试卷所测内容结构分析

心理学统考试卷的内容, 按层次可以分为考试科目、一级内容、二级内容和三级内容。考查2007年至2010年的四个大纲版本, 34项一级内容从未有过实质性的改动; 100多项二级内容和400多项三级内容之中有少量增删, 但实质性的改动不超过5%, 因此, 可以认为: 四个版本的大纲同质。

(1) 四年试卷的内容结构变化及与大纲之比较

在考试科目上, 大纲规定“试卷内容结构: 普通心理学, 约100分; 发展与教育心理学, 约70分; 实验心理学, 约60分; 心理统计与测量, 约70分。”^[1] 根据试题所考内容与能力评定结果, 统计出四年试卷在各科目上的占分, 如表1所示。

表1 四年试卷各科目内容占分变化及大纲要求

科目	2007年	2008年	2009年	2010年	大纲要求
普心	100	102	100	100	约100
发展教育	70	70	70	70	约70
实验	59	45	60	60	约60
统计测量	71	83	70	70	约70

从表1可以看出, 2007、2009、2010年试卷的各科目占分

* 收稿日期: 2010-09-09 修回日期: 2010-11-08

作者简介: 戴步云, 男, 上海人, 江西师范大学心理学院2008级硕士研究生; 罗照盛, 男, 江西吉安人, 江西师范大学心理学院教授, 硕士生导师, 心理学博士, 均主要从事测验编制与人员测评研究。

分布都与大纲的要求一致。2008年的普心和发展教育部分与大纲要求一致,但实验部分的分数比大纲要求少15分,统计测量部分的分数比大纲要求多13分,这可能与统计和实验的内容难于界定有关。总体来说,除了2008年的实验和统计测量,这几份试卷在各科目分数分布上与大纲的要求还是基本一致的,命题的科目内容结构相对比较稳定。

进一步统计出各年试卷一级内容所占的分数,发现:大纲所规定的34个一级内容有33个在4份试卷中都有所涉及,其中有25个一级内容每年都考,另有3个一级内容只在3份试卷中出现过,4个一级内容只在2份试卷中出现过,1个一级内容只在1份试卷中出现过。2007年至2010年的命题,分别涉及了3Q、3Q、2Q、2Q 29个一级内容,所考查范围的广度比较一致。可以说,考试的内容基本上有效地覆盖了大纲的内容。但是,发展教育心理学中的“童年期心理发展”这个一级内容从未单独命题(4份试卷中有个别选择题涉及到了这方面的知识,但主要考查的是其他内容领域)。普心和

表2 四年试卷科目与题型的双向分布表

科目	单选题		多选题		简答题		综合题		总分
	分数	%	分数	%	分数	%	分数	%	
普心	164	40.8%	48	11.9%	70	17.4%	120	29.9%	402
发展教育	96	34.3%	24	8.6%	40	14.3%	120	42.9%	280
实验	110	49.1%	24	10.7%	30	13.4%	60	26.8%	224
统计测量	150	51.0%	24	8.2%	60	20.4%	60	20.4%	294
合计	520	43.3%	120	10.0%	200	16.7%	360	30.0%	1200

大纲规定每年试卷单选题、多选题、简答题、综合题各占13Q、3Q、5Q、90分。我们用卡方拟合度检验来检验不同科目所使用的题型分布是否与大纲基本一致。结果如下:

普心 $\chi^2 = 2.25$ 实验 $\chi^2 = 4.05$ 都没有达到差异显著,说明普心和实验这两科的题型分布与大纲分布基本一致;而发展教育 $\chi^2 = 22.24 (P < 0.01)$ 、统计测量 $\chi^2 = 16.49 (P < 0.01)$,说明这两科的题型分布与大纲要求不太一致。根据表2可见发展教育科目单选题占分比例较小,综合题占分比例较大。统计测量单选题和简答题在各科目中占分比例都是最大,而综合题占分比例最小。作为应用性较强的统计测量用综合题考查的比例最小,我们认为是不太合理的。但值得一提的是,2010年试卷中,统计测量的综合题达到了30分,而此前三年试卷的统计测量的综合题总共只有30分,这可能说明命题人已经意识到了这个问题。

进一步将3份试卷的一级内容按题型进行统计。汇总

表3 四年试卷合计所测能力占分比变化

能力	2007年		2008年		2009年		2010年		四年合计	
	分数	%	分数	%	分数	%	分数	%	分数	%
记忆	120.5	40.2%	178.0	59.3%	117.5	39.2%	128.5	42.8%	544.5	45.4%
理解	115.0	38.3%	80.0	26.7%	118.5	39.5%	145.5	48.5%	459.0	38.3%
应用	64.5	21.5%	42.0	14.0%	64.0	21.3%	26.0	8.7%	196.5	16.4%

从表3中可以看出,4份试卷整体上考记忆的题目的分值占了总分值的45.4%,已经接近一半;考理解的题目的分值占38.3%;而考应用的题目的分值仅占16.4%,还不到考

发展教育部分,有些知识点占的分值随年份变化而有较大变化,这是因为命题组在各知识点上轮流出题^①所致,合情合理。除去来自大题的因素,各知识点的分数分布从总体上是比较稳定的。但值得注意的是,2010年试卷里的“心理测量的基本理论”部分所占的分值明显少于前三年的试卷,只有8分,而前三年的试卷分别有21分、25分和20分。从该知识点在心理学中的重要性来看,这有些不合理。

(2) 试卷的内容结构与题型结构关系分析

考试所用的不同题型有不同的功能,题型使用是否得当也是试卷质量的重要影响因素。从内容角度来说,有的内容则更适合使用这类题型,有的内容更适合使用那类题型,并不能一概而论。但是,题型使用也会受命题者的使用习惯和个人偏好影响,造成题型单一或所用题型不能很好测量所测的内容。根据试题所测内容的评定,我们制作了四年试卷科目内容与题型的双向分布表,如表2所示。

后发现:“心理学概述”、“学习与心理发展”、“实验心理学概述”、“描述统计”等十几个一级内容只出过选择题,而“推断统计”、“心理测量的基本理论”、“心理实验的变量与设计”这三个一级内容每年必出分值较大题。这反映出不同的一级内容在实际研究中的重要性不同。

2 四年试卷所测能力结构分析

根据大纲的规定,我们将考试所考查的能力分为“记忆”、“理解”、“应用”三个层次。参照布鲁姆的认知领域类型的划分方法,^[5]本文在评定试题所测能力时将以上三个名词的定义如下:记忆,再认或回忆事实或专门的知识;理解,解释、转换、概括或释义信息;应用,用信息和已有知识来解决实际工作、生活中的问题。

(1) 四年试卷所测能力结构概况

根据试题所测能力评定结果,统计出下表:

记忆题的一半,这不能不说是命题设计中的一个偏差。作为研究生招生考试,理应重点考查学生的理解和应用能力,但统考试卷却用大量试题考查记忆。我们认为,这是一种不正

① 本文把简答题和综合题统称为大题,而单项选择题和多项选择题统称为选择题。

确的导向。心理学是一门以实验研究为基础的科学,要想使招进来深造的研究生未来在心理学研究上有所建树,在考试中应该加强对理解和应用能力的考查。否则,可能会导致录取的研究生的科研实践能力不足,影响心理学的未来发展。不过,考试大纲未明确规定试卷所考查的能力应如何分布,也是一大失误。

根据表 3 做卡方同质性检验,分析四年试卷所测能力的分布是否一致。计算结果: $\chi^2 = 57.77(P < 0.01)$, $df = 6$ 说明 4 年试卷所测能力的分布是不一致的。分年看,2007 年和 2009 年的试卷所测能力分布大致相当;相比之下,2008 年试卷对记忆能力的考查更加偏重,几乎达到了总分的 60%,更加忽视对理解能力和应用能力的考查。在 2008 年

表 4 四年试卷不同科目所测能力比较

科目	记忆能力分		理解能力分		应用能力分		总分
	分数	%	分数	%	分数	%	
普心	247.5	61.6%	133.5	33.2%	21.0	5.2%	402
发展教育	137.5	49.1%	97.5	34.8%	45.0	16.1%	280
实验	43.5	19.4%	117.5	52.5%	63.0	28.1%	224
统计测量	116.0	39.5%	110.5	37.6%	67.5	23.0%	294
合计	544.5	45.4%	459.0	38.3%	196.5	16.4%	1200

用卡方同质性检验来探究不同科目所考查能力结构的一致性。 $\chi^2 = 132.21(P < 0.01)$, $df = 6$ 这说明对不同科目所考查的能力结构是有显著差异的。由表 4 可知,实验心理学所考查的能力结构以考查理解 and 应用较多,对记忆的考查较少,这应该是较为理想的。而普心对记忆的考查所占比例很大,对应用的考查所占比例很小。发展教育和统计测量都

表 5 各科所测能力逐年统计表

科目	所测能力	2007年	2008年	2009年	2010年
普心	记忆能力分	61.5	90	48	48
	理解能力分	37.5	12	37	47
	应用能力分	1	0	15	5
	合计	100	102	100	100
发展教育	记忆能力分	27	37	30.5	43
	理解能力分	28	18	24.5	27
	应用能力分	15	15	15	0
	合计	70	70	70	70
实验	记忆能力分	4	9	6	24.5
	理解能力分	33.5	28.5	23	32.5
	应用能力分	21.5	7.5	31	3
	合计	59	45	60	60
统计测量	记忆能力分	28	42	33	13
	理解能力分	16	21.5	34	39
	应用能力分	27	19.5	3	18
	合计	71	83	70	70

分科看:普心这科,考记忆能力的分值变化最大,2007 年 61.5 分,2008 年 90 分,几乎增加了 50%,2009 年和 2010 年又大幅度减少为 48 分,占比不仅远低于 2008 年,而且也比 2007 年减少 13.5 分。这说明命题人已经意识到对记忆的考查比例不能过大。值得关注的是:2007 年和 2008 年的普心题几乎没有考查应用能力;而在 2009 和 2010 年的试

统考之后不久,一种“心理学研究生很容易考,只要你能背”的观点被不少考生所认同,这或许与这个不良导向有关。2009 年的试卷纠正了过于偏重考查记忆能力的倾向,重新回到 2007 年的能力分布上来,随后“只要能背就能考上心理学研究生”之说有所减少。到了 2010 年,考理解能力的分值达到历史最高,但考记忆能力的分值与 2007 年、2009 年相比不降反升,而考应用能力的分值却达到了历史最低,这说明整个试卷所考查的能力水平又一次降低了,这又是一种倒退。

(2) 试卷各科目四年合计所测能力比较

进行不同科目之间所考查能力分布的比较,如表 4 所示:

是对记忆的考查多于对理解的考查,对理解的考查多于对应用的考查。

(3) 试卷各科目四年所测能力纵向比较

下面对各科目的命题趋势做纵向分析,制作各科所测能力逐年统计表,如表 5 所示:

卷中,考应用能力的分达到 15 分和 5 分,说明普心已经开始注重考查应用能力。这对于原本不够灵活的普心题的命题是有益的尝试,可以作为今后命题普心题的方向。

发展教育这科,在前三年试卷中,应用能力题的考查分比较稳定,而记忆能力分 2008 年陡增 10 分,理解能力分减了 10 分,2009 年又各自回归 2007 年结构。但到了 2010 年,

考查记忆能力的分值超过了前三年,考查应用能力的分值却从前三年的每年 15分突然降到了 0分,这显然是不合理的。

实验这科,在前三年试卷中,所测记忆能力占分比例相对较小,所测理解能力分呈逐年减少趋势。但 2010年的试卷又一次打破了这个趋势,所测记忆能力分值突然大幅提高,而所测应用能力分值达到历史最低值。作为最贴近实际研究的实验心理学,对应用能力的考查却大幅减少,对记忆能力的考查反而大幅提高,这又是不正确的导向。

统计测量这科,前三年试卷中记忆能力题的占分比例总在高位以较大幅度波动,而 2010年有大幅下降;考理解能力的分值在四年来呈逐年增大趋势。这是合理且正确的,因为统计测量本来就需要灵活掌握。但四年的试卷中,对应用能力的考查分值却是大幅波动的,这说明命题尚未达到稳定。

总体看,不同科目四年试卷所考查的能力分布常有大幅度反复。这说明在能力考查上,各科命题者之间、各年度命题者之间认识都还未达到一致,导致总不能形成稳定的能力测试结构,当然这还是与教育部大纲没有明确规定考试的能力结构有关。2008年试卷对记忆能力考查过多,2010年试卷对应用能力考查过少,这都不太合理。相比之下,2007年和 2009年的试卷要更合理一些。

四、结论和建议

通过对四年的心理学研究生入学统考试卷的分析,发现:

1 在试卷的内容结构方面,四年试卷的科目分布、各科的一级内容分布基本上都符合大纲要求;

2 在试卷的能力结构方面,四年试卷记忆能力的占分比例都偏高,相对削弱了对理解和应用能力的测量,这不利于入学研究生的进一步培养和提高;

3 统考试卷所测能力结构在年度之间和科目之间都有较大差异,年度之间还常有较大反复,试卷的能力结构还很不稳定。

建议教育部颁发新大纲时应该进一步明确试卷的能力结构,并且引导研究生考试向更高层次能力的考核发展。

参考文献:

- [1] 教育部考试中心. 全国硕士研究生入学统一考试心理学专业基础综合考试大纲(2010年版)[S]. 北京: 高等教育出版社, 2009.
- [2] Kaplan Saccuzzo Psychological Testing[M]. 赵国祥等,译. 西安: 陕西师范大学出版社, 2005
- [3] 安娜斯塔西,厄比纳. Psychological Testing[M]. 缪小春等,译. 杭州: 浙江教育出版社, 2001
- [4] 吴根洲. 高考效度研究[M]. 武汉: 华中师范大学出版社, 2008.
- [5] 雷新勇. 大规模教育考试: 命题与评价[M]. 上海: 华东师范大学出版社, 2006

责任编辑: 吴旻

An Analysis of the Content Structure and the Ability Structure in Four Years of Postgraduate Psychology Entrance Examination

DAI Bu-yun, LUO Zhao-sheng

(College of Psychology, Jiangxi Normal University, Nanchang 330077, China)

Abstract Postgraduate Psychology Entrance Examination (PPEE) had been executed in China for 4 years, and the quality of these examination papers were worthy of paying attention. The aim of this article is to evaluate the content validities and construct validities of these papers, and their consistencies with the outline. Results showed that the content structure was basically agreed with the requirements of the outline, and was relatively stable. However, the to-be-tested ability structure was unstable in four years and had significant differences between different branch subjects within psychology. Suggestions were given to improve the quality of further PPEE papers.

Key words psychology; postgraduate entrance examination; content structure; ability structure