

信效度研究在语言测试中的运用

——浅析大学英语考试试卷

孙迪辉 程启翔

(国防科技大学 人文与社会科学学院, 湖南 长沙 410074)

[摘要] 本文从语言测试的角度简要地分析了九院04级学员第四学期期末考试英语试卷的信度、效度, 为以后设计更为科学、合理的测试试卷以促进、提高英语测试水平提供了一定的参考。

[关键词] 语言测试; 信度; 效度

[中图分类号] G642.474 [文献标识码] A [文章编号] 1672-8874 (2007) 02-0042-02

一、试卷概况

大学英语考试是一项全校性的教学考试, 其目的在于推动大学英语教学大纲的贯彻执行, 对我校大学生的实际英语能力进行客观、准确的测量, 为提高我校大学英语课程的教学质量服务。考虑到在我校2004级1至7院及9院学员教学中采用的是两种不同的教材, 所以这次期末考试

除听力部分相同外, 其余题型均不相同。本文分析采用的是针对9院指挥合训类学员的命题试卷, 这套试卷的性质基本上是学业成绩考试 (achievement test)。

据统计, 06年6月共有2017名2004级9院学员参加了这次英语期末考试。本文打算对这次的期末考试试卷做一些初步的分析, 旨在促进、提高英语测试水平及更合理地命题。现将06年英语试题概况列表如下:

Paper	Part	Item NO	Name	Type	S/O	NO of Items	Weighting	Duration (mins)
Paper One	I	1- 20	LC	M/C	O	20	20%	20
	II	21- 40	RC	M/C	O	20	40%	35
	III	41- 70	VS	M/C	O	30	15%	20
	IV	71- 90	CL	M/C	O	20	10%	15
Paper Two	V	91	WR	Writing	S	1	15%	30
Total						91	100%	120

Table 1: 2006年春九院04级学员第四学期期末考试试卷组成

LC= listening comprehension; RC= reading comprehension; VS= vocabulary and structure; CL= cloze; WR= writing; M/C = multiple choice, O= objective test, S= subjective test

(一) 信度 (Reliability)

所谓信度, 是指测试结果的可稳定性或可靠性的程度, 也就是看测量的结果是否真实客观地反映了考生的实际水平。信度是反映测试是否受到非测试因素的影响, 反映测试客观性和可靠性的指标。影响测试信度的因素很多, 主要涉及到试题和评分两个方面。就试题本身而言, 其信度主要取决于测试范围与题量; 而评分的信度则取决于评分标准的客观性与准确性。

先看看测试范围。这次试题的覆盖面较广, 对学生各个方面的能力都进行了考核: 听力部分考核了数字、地点、人物关系等; 阅读部分包括主旨题, 词义题, 推断题, 是非判断题, 逻辑细节题, 态度题等; 词汇与结构部分主要考察了动词短语的用法, 介词短语及其他词语搭配, 同义、

近义词辨析等; 完形填空部分重点考察了非谓语动词、副词、介词以及虚拟语气等; 写作部分则让学生们联系实际状况, 与目前形势相结合, 以所学所掌握的知识各抒己见。

从题量来看, 这次的考试试卷共有5大题91小题, 考试时间则定为2个小时。相对于大学二年级英语学习者的语言水平来说, 这次考试的题量与时间安排是比较合理的。

从评分标准来看, 如Table 1所示, 这次考试试卷的组成中, 客观性试题占总分的85%, 主观性试题只占了15%, 就整份试卷来说, 评分很大程度上兼具了客观性与准确性。

所以, 从以上三个方面来看, 这次测试试卷的信度是比较高的。同时, 笔者还运用了计算机来分析这次测试的可靠性, 即信度。把所有考生的成绩输入计算机, 选用 α 系数来计算一致性信度系数 (因为这次测试中包括有非0-1

* [收稿日期] 2006-08-28

[作者简介] 孙迪辉 (1972), 女, 湖南邵阳人, 硕士, 国防科技大学讲师。

记分的题目), 得出结果如下:

Alpha= . 7034

Standardized item alpha= . 7001

在全国大学英语考试中, 一致性信度 α 系数一直稳定地位于 0. 80 以上; 而从我们的试卷的信度系数可以看到, 这次测试与大规模的大学英语四六级标准化考试还存在着一定的距离, 试卷的信度系数还有待提高。

(二) 效度 (Validity)

所谓效度, 是指测试结果的准确性和有效性的程度, 也就是看测试是否达到了预期的目的。效度高体现在需要测量的东西应该全部包含在测试的内容里。效度研究主要涉及到内容效度 (content validity)、编制效度 (construct validity)、实证效度 (empirical validity) 和表面效度 (face validity), 下面就从这四个方面对本次试卷进行研究检验。

A) 内容效度指试题是否测试了为达到该考试的目的而应测试的内容, 而测试内容则应该与教学大纲、教学内容保持一致。对照教学所用的教材——清华大学出版社出版的新英语教程阅读第三版, 发现测试的语法项目与书中的语法内容基本保持一致, 而词汇内容与教科书之间存在着一定的差距, 有些单词或词组没有在教科书词汇表中出现过。但是, 并不能就此认为本试卷的内容效度不高, 因为在一份英语考试试卷当中, 可以允许出现少量不包含在教科书词汇表中但属于“领会性掌握”范围内的非冷僻词。总观全卷, 可发现本试卷对说和译的能力无反映, 故在这两方面无有效性; 但在听、读、写方面有具体要求, 对这几个方面的考核存在相当高的效度。

B) 编制效度是指从教学理论上分析试卷是否测试了该种理论应测试的内容, 也称作观念效度, 也有译为结构效度的。本试卷共有五大项考题, 91 小题, 总计 100 分。其中 90 道小题, 占总分 85% 的题目为多项选择题, 属客观题; 1 小题, 占总分 15% 的为写作题, 属主观题。从题型来分析, 以结构主义语言学为理论根据将语言分成语音、词汇、语法, 设计出离散的题目, 通过多项选择题的具体

方式将语言技能分解成听、读等项目, 并运用主观题的方式测试了写的能力; 以心理语言学为理论基础对多种语言技能和语言项目进行了综合性的测试, 如完形填空、作文等题型。

C) 实证效度的含义是测试结果应该与考生的平时成绩或者其他标准测试的结果相吻合。如 Table 2 中 04 级入校以来第一至第四学期期末考试及格率与平均成绩所示, 第四学期总体来说与往年测试结果相吻合。

考试年度	及格率	平均成绩
2004 年秋季学期	96. 9%	74. 41
2005 年春季学期	95. 0%	71. 44
2005 年秋季学期	97. 4%	71. 27
2006 年春季学期	96. 2%	71. 06

Table 2

D) 表面效度指的是试卷印刷清楚, 基本上无错误。这次试卷无论从印刷质量、布局排版, 还是从答案选项 (除第三大题第 70 小题中出现了一个小的错误外)、篇幅长短上都基本做到了这一点, 并且试卷与答题卡、试卷二也就是主观题答卷分开, 便于改题, 也便于机器阅卷。

从以上四方面对这次考试试卷进行考察, 如果以新教学大纲为依据, 此次试卷的效度尚存在一些差距, 但针对当前对二年级学员英语学习合格标准而言, 这次试卷的效度良好。

三、改进试卷的建议

1) 试卷格式有待调整

将参加考试的 2017 名学员成绩按照听理解 (LC)、阅读理解 (RC)、词汇与结构 (VS)、完形填空 (CL) 及写作 (WR) 五项分开, 与总分 (TOTAL) 一起输入计算机计算各部分之间相关系数, 结果列表如下:

Pearson Correlation

	TOTAL	LC	RC	VS	CL	WR
TOTAL	1. 000	. 534* *	. 833* *	. 577* *	. 476* *	. 262* *
LC	. 534* *	1. 000	. 200* *	. 211* *	. 223* *	. 010
RC	. 833* *	. 200* *	1. 000	. 280* *	. 232* *	. 049*
VS	. 577* *	. 211* *	. 280* *	1. 000	. 207* *	. 108* *
CL	. 476* *	. 223* *	. 232* *	. 207* *	1. 000	. 001
WR	. 262* *	. 010	. 049*	. 108* *	. 001	1. 000

* * Correlation is significant at the 0. 01 level (2- tailed).

* Correlation is significant at the 0. 05 level (2- tailed).

Table 3: 九院 04 级学员第四学期期末考试试卷内部相关系数

就语言测试而言, 上述各部分相关系数说明此次考试的试卷格式还有待调整。各部分之间相关系数稍偏低, 说明试卷能够分别测试出考生各种不同的语言能力, 但语言能力的综合性测试强度不够。

2) 个别选择题的备选答案的设计不够科学合理

通过对题目的难易度及区分度指数分析, 可以看出个别选择题的备选答案的设计不够科学合理。随机抽取一个学员队, 将成绩在前 27% 者划为好学员组, (下转第 73 页)

个环节中需要掌握的方法和注意的问题, 使学生能够较为完整地完成本专业领域的一次研究活动, 帮助学生掌握知识和培养技能。

五、结束语

研究生专业课程建设是一个系统性的工作, 需要考虑方方面面的影响因素, 本文只是对研究生专业课程建设中一些共性问题的初浅思考。对于不同的研究专业和不同的学生主体, 专业课程建设应该具体问题, 具体分析, 只要坚持研究生专业课程为培养科研人才和创新人才的最终目标不变, 就能够形成一门优秀的研究生专业课程。

(上接第 43 页) 成绩在后 27% 者划为差学生组, 从两组中分别抽取 10 名学生。从他们的答卷中分析第二大题第 40 小题难易度指数 (F. V.) 及区分度 (D. I.)。通过计算可以看出, 这道试题设计还是比较合理, 难易度指数为 55%, 说明题目有一定难度; 区分度为 0.50, 其中 80% 的好学生都选择了正确答案, 只有 30% 的差学生选对了, 本题可以起到检测学生对本篇文章阅读能力的作用, 但是其中的 B 选项, 无论好学生还是差学生都没有人选择, 这说明 B 选项仍需改进, 以起到干扰项之作用。

第三大题第 54 小题的备选答案中除正确答案 consciousness 外, 其他三个 consensus, consent, conscience 都不曾包含在所用教材词汇表中, 对考生选择正确答案造成了一定的困难。而在第三大题第 50 小题中, 备选答案 scream, exclamation, whistle 对于正确答案 roar 的干扰作用不大, 这样学生做这道题时就能很轻易地选出正确答案。

3) 可考虑增加口试

从 1999 年起, 大学英语四六级考试开始实施口语考试。它标志着这项考试进入了一个相对完善的新阶段: 可以对考生的听、说、读、写能力进行全面的鉴定。在 04 级第四学期期末考试中还尚未采用口语考试手段, 仅根据学生平时上课课堂活动的参与, 课后作业的完成等给出一个平时成绩以作参考。这种情况已经在 05 级学员中得到解决, 在 05 级期末考试中已经采用了口语考核方式, 对学生的说的能力作了一定评价。

4) 可适当增加主观题所占比例

我校大学英语考试借鉴了全国大学英语四级考试题型, 分为固定型和可替换型两类。其中听力部分的对话题、阅读理解部分、词汇与结构部分、写作部分属于固定型, 而听力部分的短文与复合式听写, 完形填空与简短回答或英汉互译均可互相选择替换。这次考试题型的选择与上个学期略有不同, 主要体现在听力部分, 将复合式听写改成了短文。从有利的方面来讲, 比较全面地考核了学生听力技巧的掌握, 但同时也使得客观题的比例大幅增加, 主观题的数量则减至 15%, 有可能由于客观题过多而产生一定的副作用。

[参考文献]

- [1] 李明兰. 我国硕士研究生课程设置的现状、问题及改进 [J]. 煤炭高等教育, 2006, (2).
- [2] 谢安邦. 构建合理的研究生教育课程体系 [J]. 高等教育研究, 2003, (5).
- [3] 包迪鸿等. 完善我国研究生课程教学的思考与研究 [J]. 黑龙江高教研究, 2005, (6).
- [4] 罗尧成. 我国研究生教育课程体系研究 [D]. 上海: 华东师范大学, 2005.
- [5] 应义斌等. 研究生创新能力培养的探索 [J]. 高教探索, 2003, (2).

(责任编辑: 赵惠君)

5) 可考虑把作文考试与其他部分考试时间分开

从 1990 年起, 这种做法已经在四六级考试中使用。为了保证作文考试的时间, 试卷分成两部分。客观题部分为试卷一, 作文题部分为试卷二。到达规定时间, 收掉试卷一, 考生必须把最后 30 分钟时间用于写作文。这主要是针对有些学生只靠客观选择题拿分, 不做作文的做法, 以达到促使考生们重视作文, 也就是写作能力培养的目的。

四、分析试卷的必要性

语言测试作为一门独立的学科, 有自己的研究领域和研究方法。它是伴随着语言教学出现的, 没有语言教学也就无所谓语言测试。语言教学是第一性的, 语言测试为语言教学服务。语言教学的任务是养成实际运用英语的能力, 而语言测试的目的则是提供一种科学的测试工具, 通过对学生语言运用的抽样对学生的语言能力进行客观的、准确的、公正的评价, 反映教学的长处和短处, 促进教学的改进。因此, 为了设计出高度科学化的试卷, 考后的统计与分析是完全必要的。只有这样, 才能依据统计分析结果尽可能找出可能存在的问题和改进方法, 保持测试的准确度, 同时还可以采用各种新题型、新方法, 防止某些题型的老化, 以便不断提高测试的信度与效度。

[参考文献]

- [1] Alderson, J. Charles and Caroline Clapham, Dianne Wall. Language Test Construction and Evaluation [M]. 北京: 外语教学与研究出版社, 2000.
- [2] 大学英语编写组. 英语测试研究文集 [C]. 北京: 北京航空学院出版社, 1986.
- [3] 李绍山. 语言研究中的统计学 [M]. 西安: 西安交通大学出版社, 2001.
- [4] 杨惠中, C. Weir. 大学英语四六级考试效度研究 [M]. 上海: 上海外语教育出版社, 1999.

(责任编辑: 范玉芳)