

英语测试信度与效度的多维分析

罗明姝

(河南大学,河南 开封 475001)

摘 要: 信度与效度是评价英语测试的两个重要依据。文章回顾了语言测试史及信度与效度的发展,并对大学英语测试题型的信度与效度进行了多维分析。指出,只有主客观题型在一定比例上的兼顾才能保证考试的可靠性与准确性。

关键词: 信度;效度;客观性试题;主观性试题

中图分类号: H310.4

文献标识码: A

文章编号: 1009 - 1750(2006)06 - 0107 - 02

一、引言

信度和效度是大学英语学业成绩测试应具备的两个质量特征。信度涉及测试结果的稳定性和可靠性,即测试是否对学生的语言水平提供可靠的度量。效度则涉及测试的有效性,即一套训练题所考的是否就是设计人想要考的内容。作为衡量学生英语水平的一种尺度,考试本身必须可靠、有效。然而,大学英语测试中占有较大比例的多项选择题是否能客观地衡量学生的外语水平?而对于近年来分数比例日益增大的主观性试题的信度与效度又是怎样?这两类试题的优劣何在?这正是本文所要研究的内容。

二、语言测试史的回顾及信度与效度的发展

测试理论的发展大致经历了三个阶段,第一阶段被称为“前科学测试阶段”(prescientific period)。半个世纪以前,传统的语言学家把语言作为一种孤立的现象加以研究。在这一基础上产生了语法、翻译法,它以语法分析、翻译和短文写作为主体。测试内容严重偏重文法,忽视口语。此时的测试毫无信度可言,虽有一定效度,但因其测试形式和评判标准的极大主观随意性而大大减弱。随后语言测试进入了第二阶段,即心理、结构测试阶段(psychometric - structuralist period)。结构主义语言学家们把语言看成是一个形式体系,是语音、语法和词汇的总和。此时的测试取消了作文和翻译,而以多项选择和人机对话的方式取而代之。罗伯特·拉多(Robert Lado)在《语言测试》一书中强调采用分离性题型,把语言化整为零,分门别类地对不同的语言成分(语言、词汇、语法)和技能(听、说、读、写)逐项测试。此时的测试内容由于考察的语言点覆盖面广,且以多项选择为代表的客观试题能够确保评分的一致性和客观性,受人为主观因素影响较小,因而信度较高。随着语言学家的不断探索,语言测试进入了第三阶段,即社会语言测试阶段(sociolin-

guisticperiod)。从 20 世纪 60 年代起,在乔姆斯基(Chomsky)的转换生成语法的理论基础上产生了综合测试法,即把语言化零为整,对学生的综合语言能力进行测验,如阅读理解、完形填空、阅读改错、会话,等等。但后来的社会语言学家海姆斯(Hymes)发表了《论交际功能》,强调了语言的社会功能,在这一理论影响下产生了交际教学法和交际测试法,因此,语言交际功能的培养是语言教学的首要目标。同时衡量语言交际能力也成为语言测试的首要任务。此时的语言测试由第二阶段的一味追求测试的信度转向信度与效度的综合考虑,并趋向更高的信度和效度。

三、现行大学英语测试题型信度与效度的多维分析

语言水平测试是一项很复杂的活动,因此,我们必须从多层面、多角度对它进行考查。听、说、读、写部分在试题中的主观题和客观题分值比例对考试的可靠性及准确性有很大影响。结构主义测试专家称多项选择题、判断正误题以及只有一个标准答案的题为“客观性试题”,因为这些题的评分不受阅卷人主观因素影响。称作文、缩写、翻译、简答等类题为“主观性试题”,因为这些试题或多或少要靠评分者的主观判断和印象来打分。这是根据它们各自的评分方法而不是根据试题本身的性质而命名的。

英语多项选择题通过测试学生所掌握的语音、词汇、语法等知识,推断学生生成句子和理解句子乃至语篇的能力。由于多项选择题考点层次较低,因此在有限的测试时间内,多项选择题通常题量较大,语言点覆盖面较广泛,这使阅卷人能够更精确地推断出考生语言知识的掌握程度,因此,多项选择题的信度较高。此外,多项选择题衡量标准明确,评分标准较客观,对学生的语言水平提供了科学准确的度量,从而进一步提高了测试的信度。

然而,多项选择题虽然信度较高,但与主观题型相比,其效度并不高。影响多项选择题效度的主要原因有两种:

收稿日期:2006 - 10 - 09

作者简介:罗明姝(1970 -),女,河南开封人,讲师,研究方向为英美文学。

一是多选题孤立地对构成语言的元素进行测试,忽视了这些元素在实际语境中的相互作用功能。而语言的本质不仅仅是各元素处于孤立状态时呈现出的特性,更重要的是其在语言整体体系中相互作用时体现出的功能。因此,多选题考察的只是学生对语言形式的识别能力,而对其语言知识的运用和产出能力无从判断,对其语言知识与语境特征相结合时的创新能力无从判断。

与客观性试题截然相反的是主观性试题,如近年来考试中出现的听写、简答、翻译、作文、口试等题型。一篇试卷中主观性试题越多,其效度就越高。主观题型不允许猜测因素的存在。例如,一篇英语作文的优劣取决于考生英语词汇、语法、句法、写作、语篇等综合知识的运用。一个生词不会写,一个句型没有把握都很难靠猜测来完成。

翻译与作文是综合运用语言的能力测试项。对写作能力的检测实际上是对语言基本功的全面检测。中国自古以来就十分重视这两项测试。在《中国近代学制史料》中,创立于 1861 年的京师同文馆,就开始用“汉文”、“洋文”互译的方法对学生进行测试。翻译、写作能力的测试在当今英语测试中仍显示着其不可替代的作用。

简答题重在考查学生对于所给材料的理解能力,对于提出问题的分析能力以及回答问题时的语言组织能力。

听写测试涉及多种综合技能的练习,包括听力理解、听力记忆幅度、拼写、识别音段和熟悉语言的语法和词汇模式,等等。

英语口语是增强学生口头交际能力的一项重要手段。传统的英语测试由于不设口试,给实际教学带来了一定的负面影响。教师为了提高通过率,学生为了过关,都不同程度地忽视口语的训练和提高,使得一些毕业生的英语口语的实际应用能力远不能适应实际工作的需要。因此,将口语导入交际测试,使测试对教学有良好的反拨作用,促使学生语言知识和语言交际能力的共同提高成为大学英语教学的当务之急。

值得欣慰的是,大学英语四、六级考试命题委员会就考试的信度和效度采取了一系列的改革,并决定从 2006 年下半年开始,全国大学英语四、六级考试全面启用新大纲、新

题型。新旧两种题型在评分体制、考察手段、考察程序和考察重点方面有明显的区别。新四级考试把听力部分的分数比重从 20% 上升到 35%,综合讲,新四级考试提高了听的能力和读的速度且更加侧重主观题的考查,主观题所占比重达到 40% 以上。考生要写的部分更多了,比如:听力部分的复合式听写,快速阅读的最后三个题目,改错题,等等。成绩测试中主观题型的合理采用,势必对大学英语教学产生巨大的作用,并极大地提高大学生实际应用英语语言的能力。

四、结 语

综上所述,客观性试题的信度虽高,但效度偏低;主观性试题的效度虽高,但信度偏低。一套试题中兼有客观题与主观题,且主客观题的比例适中,才能保证考试的可靠性与准确性,增强测试的目的性,从而提高测试的质量。

参考文献:

- [1] Anderson, J. Charles Language Test Construction and Evaluation [M]. 北京:外语教学与研究出版社, 2000.
- [2] Hughes, A. Testing for Language Teachers [M]. 北京:外语教学与研究出版社, 2000.
- [3] 陈坚林. 英语教学组织与管理 [M]. 北京:外语教学与研究出版社, 2000.
- [4] 井升华. 我国大学英语教学费时低效的原因 [J]. 外语教学与研究, 1999, (3): 21 - 23.
- [5] 刘润清, 韩宝成. 语言测试和它的方法(修订版) [M]. 北京:外语教学与研究出版社, 2000.
- [6] 牛 强. 现行高校英语测试中的问题 [J]. 外语教学与研究, 2001, (2): 140 - 143.
- [7] 杨惠中. 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, 1998.
- [8] 朱有献. 中国近代学制史料 [M]. 上海:华东师范大学出版社, 1983.
- [9] 邹红英. 大学英语测试新思路:论交际测试的介入 [J]. 福建外语, 1997, (1): 40 - 43.

责任编辑:孙咏梅

The Multi - analysis of the Reliability and Validity of the English Tests

LUO Ming-shu

Abstract: Reliability and validity, are two essential bases in evaluating English tests. In this essay, the author looks back on the history of language tests and development of reliability and validity and analyses the reliability and validity of different components in one test. The author points out that only a test with subjective parts and objective parts in reasonable proportion is reliable and precise.

Key words: reliability; validity; objective parts; subjective parts