

Analysis on Reliability and Validity from Proofreading in TEM8: Based on Bachman's Communicative Language Testing Model

Yanhong Zheng¹ Wenlin Wu² Zhihua Qiu³ Yiqing Chen⁴

1.Xiamen University Tan Kah Kee College, Zhangzhou, Fujian, 363105, China

2.Zhangzhou Jingcheng Middle School, Zhangzhou, Fujian, 363000, China

3.Changting No.1 Middle School, Longyan, Fujian, 366300, China

4.Fuzhou No.2 Middle School, Fuzhou, Fujian, 350001, China

Abstract

In accordance with Bachman's model of communicative language testing, this paper aims to examine the reliability and validity of proofreading in TEM8 (Test for English Majors-VIII in China), by the contrast of the test between 2014 and 2015, so as to present some enlightenment for the test of proofreading, English proficiency and teaching. The results showed that: ① The correct rate of the test in 2015 was higher than that in 2014; ② The reliability of proofreading from 2015 was much higher than 2014, but both of them did not have desirable reliability; ③ Proofreading from 2014 and 2015 contained intermediate face validity and construct validity together with high content validity, whereas the latter had higher validity than the former except content validity.

Keywords

communicative language testing; TEM8; proofreading; validity; reliability

巴赫曼交际语言测试理论下英语专八考试之改错的信效度分析

郑艳红¹ 吴文琳² 丘智华³ 陈怡清⁴

1. 厦门大学嘉庚学院, 中国·福建 漳州 363105

2. 漳州靖城中学, 中国·福建 漳州 363000

3. 长汀第一中学, 中国·福建 龙岩 366300

4. 福州第二中学, 中国·福建 福州 350001

摘要

基于巴赫曼交际语言测试理论, 论文选取了2014年和2015年英语专业八级考试的短文改错真题, 对比二者在信度和效度的差异, 试图为该类型试题和考试提供有意义的启示, 同时反哺教学。研究表明: ①2015年的答题正确率高于2014年; ②与2014年相比, 2015年的信度更高, 但二者的信度都不够理想; ③两年改错均具有中等程度的表面和结构效度及较高级别的内容效度, 除了内容效度, 2015年的效度都稍高于2014年。

关键词

交际语言测试; 英语专八; 改错; 效度; 信度

1 引言

测试或考试, 是日常教学过程中不可或缺的一部分。它与教学相辅相成, 相互促进, 旨在帮助教学者更及时有效地了解学习者的学习情况。因此, 如何使语言测试更好促进语

言教学是一个值得探究的话题。对语言测试者而言, 无论测试的内容是什么, 他们更关注测试想要检验受试者的哪些语言能力以及如何有效实现这一目的^[1]。

英语专业八级考试, 简称专八, 是面向英语专业学生的一项重量级测试, 旨在检测其英语综合水平。改错是专八的题型之一, 主要考查考生的语法能力。根据本研究的问卷调查结果, 超一半认为改错是专八最难的题型之一, 所以专门

【作者简介】郑艳红(1992-), 女, 中国福建漳平人, 研究生学历, 助教, 从事心理语言学、语言测试研究。

分析改错有着一定的现实基础和意义。然而,经文献查阅发现,虽然以往分析专八的研究并不鲜见,但是较少能从改错角度分析,且较少谈及其在信度和效度的情况,因此并不能解决一些现实问题,如改错的考点设计及其对应试者和出题者的影响。基于此,有关专八尤其是改错部分的研究亟待丰富。本研究试图基于巴赫曼的交际语言测试理论,通过实验分析专八改错的信效度情况,并基于研究结果给予语言测试相应的启示。

2 理论与研究回溯

2.1 交际语言测试理论

在交际语言理论问世之前,语言测试先后受到了结构主义、功能主义和社会主义的影响。然而,越来越多人发现其局限性,故交际语言理论应运而生。海姆斯^[2]是该理论的首创者,他将乔姆斯基的语言能力内涵扩宽至四个方面。随后,卡耐尔和斯温^[3]又对交际语言能力进行新的解释却未能说清四者间的关系。10年以后,巴赫曼再次进行修缮,其理论思想在语言测试史上被视为一个里程碑^[4]。他认为,语言能力首先应当包含语法规则知识以及懂得语言是如何被应用于特定交际目标之中^[5]。同时,语言的使用被视为一个动态过程,在此之中所有语言能力相互作用和影响。此外,交际语言能力作为该理论的核心,主要包括语言能力、策略能力和心理生理机制。其中,他对前面两种能力给出了十分详细的描述。

2.2 信度与效度

效度和信度是语言测试的基本构成要素。信度指测试分数多程度是可靠的,而效度则关注是何种能力促成了这个可靠结果^[5]。信度包含不少因素,例如测试方法、受试者属性等,而且还会受到测试任务、操作过程和评分的影响^[6]。比起信度,效度更关注从测试分数中推算受试者的真实能力,也更重视受试者在测试过程的表现。效度有许多组成部分,但只有一些值得研究,如表面、内容和结构效度。故本研究仅考察以下两点:

第一,表面效度主要评估测试任务是否反映受试者的能力,内容效度则考察测试任务与考试大纲要求有多大程度的一致性,因此其高低取决于测试本身的设计^[6]。

第二,结构效度作为效度的基本成分,关注的是受试者在测试的表现多大程度符合测试者基于理论的预测。因此,结构效度能回答一项测试真正检验了什么的问题。

2.3 专八改错

自专八考试施行以来,改错就是其不可或缺的一部分。该考试大纲曾明确规定考生应当“能使用语言知识辨别文中错误,提供相应的正确解决方案”。改错,类似于完形填空,属于一种检验单一技能的测试,是单一语言能力假说的产物^[7]。该假说认为语言的使用过程是多种语言技能协调与合作的结果。事实上,改错不仅要求受试者掌握语法和词汇知识,也要求他们理解语篇结构和意义。这一说法的证据主要来自刘建达和杨满珍^[8]的研究,他们证实了改错的目标远远超出语法等基础知识。相反,改错更侧重评估受试者对单词、短语、句子和语篇的理解,而这四个方面实际上源自层次分析法的原理。基于此,论文将改错考点分为四个方面,即词汇、短语、句子和语篇。词汇包括单词的拼写、意义和词法等;短语包括词组、介词短语等常见固定表达;句子指句子内部结构、意义及句法等;语篇指句子间的意义、逻辑关系和修辞等。根据这一分类方法,发现近十年改错最重视句子,其次是短语和语篇,最后是词汇(如表1所示)。

表1 2010年—2019年改错的考点分类情况

年份	词汇	短语	句子	语篇
2010	2	2	3	3
2011	4(最高)	3	1(最低)	2
2012	2	4(最高)	3	1(最低)
2013	4(最高)	0(最低)	4	2
2014	3	3	2	2
2015	2	2	4	2
2016	2	1	3	4(最高)
2017	1(最低)	2	4	3
2018	1(最低)	2	5(最高)	2
2019	1(最低)	4(最高)	3	2
总计	22	23	32	23

3 实验

3.1 研究问题

(1) 就正确率而言, 哪一种考点得分最高, 哪一种考点得分最低?

(2) 2015 年和 2014 年的改错题型中, 哪一年的信度和效度更高?

3.2 研究方法

3.2.1 被试

某大学 48 名一年级英语专业研究生自愿参与本实验。其中, 女 30 人, 男 18 人, 平均年龄 24 岁。他们都曾参加过专八考试, 水平相当, 且对 2014 年和 2015 年的考试真题不熟悉, 因此适合作为本研究对象。

3.2.2 研究材料

研究包含两种实验材料: 调查问卷以及 2014 年和 2015 年专八改错真题。问卷旨在调查被试的背景信息, 而选取 2014 年和 2015 年改错的原因在于两点。一是参与实验的真题不能过于贴近被试参加专八的时间, 且年份也不宜久远; 二是这两年真题的考点分布比较均匀(无最高或最低情况), 因此适合进行对比。

3.2.3 研究过程

首先, 随机选取 100 名英专学生参加问卷调查。其次, 基于调查结果, 筛选出适合作为本实验的被试, 接着要求他们在 30min 内完成测试。测试结束以后进行评分。最后, 采用 SPSS 进行数据分析。

4 数据分析

4.1 描述性数据

表 2 2014 年和 2015 年改错在各考点的正确率分布情况

	句子	语篇	短语	词汇	平均正确率	平均值
2014	68.5%	32.5%	18.0%	38.3%	39.3%	3.72
2015	47.8%	43.5%	13.0%	55.5%	40.0%	4.15
平均正确率	58.2%	38.0%	15.5%	46.9%	39.7%	

根据表 2 的数据结果, 可知被试在句子考点的正确率最高(58.2%), 在短语考点的正确率最低(15.5%), 说明被试者不擅长短语类考点, 或者对短语的考察较不敏感。此外, 2015 年的平均值为 4.15, 其正确率高于 2014 年(平均值为 3.72)。

4.2 信度

表 3 呈现了 2014 年和 2015 年改错测试的信度情况: 2014 年为 $0.526 > 0.5$, 而 2015 年为 $0.658 > 0.5$, 显然后者高于前者, 说明 2015 年内部题项之间更为一致。然而, 根据吴明隆^[9]对信度的解释, 2014 年的信度不够理想, 但其构念是可接受的, 而 2015 年的信度构念较佳且其结果是可接受的。

表 3 2014 年和 2015 年的改错信度情况

年份	克隆巴赫系数	基于标准化项目的克隆巴赫系数	题项数
2014	.516	.537	10
2015	.658	.652	10

表 4 2014 年改错错误率

考点	句子		语篇		短语		词汇			
题号	4	9	1	8	2	3	6	5	7	10
具体考点	从句	语态	逻辑	逻辑	动词搭配	固定词组	介词短语	冠词	名词	词义
错误率	20%	43%	15%	35%	61%	85%	100%	35%	74%	76%

表 5 2015 年改错错误率

考点	句子				语篇		短语		词汇	
题号	1	4	6	10	2	5	8	9	3	7
具体考点	从句	从句	从句	语态	逻辑	逻辑	介词短语	固定搭配	词性	副词
错误率	9%	59%	52%	89%	39%	74%	83%	91%	63%	26%

4.3 效度

4.3.1 表面效度

问卷调查结果显示,超过65%的被试认为改错能帮助检验语法和词汇水平,但也有部分人认为改错对语篇意义和逻辑推理的考察不够充分,尤其是2014年。从这点上看,这两年改错具有中等程度的表面效度,但2015年稍高于2014年。

4.3.2 内容效度

如表1所示,与其他年份相比,2014年和2015年改错考点的总体分布较好,且2014年的分布比例更优。再结合上文提到的考试大纲对改错的要求,这两年改错均考察了语言知识,还涉及了逻辑关系。综上,这两年改错均具有较高的内容效度,其中2014年较优于2015年。

4.3.3 结构效度

如前所述,改错主要检验考生的语法知识储量,发现并正确解决问题。由表4和表5可知,它们均考察了考生的短语知识积累、句子的语义理解和语法知识、语篇的逻辑推理以及词汇的词性转换和意义理解等。以上考察内容相对均匀,与考试大纲要求大体吻合,但部分测试题项可能隐藏了与构念无关的因素。以2015年改错第五题为例(错误率为74%),该题要求考生能判断句子的基本结构特征,同时通过分析句子间关系找出错误。根据巴赫曼对语言能力的定义,该题同时检验了语法和文本能力,与理念相悖(非单一技能)。总的来看,两年改错的结构效度中等。但与2015年相比,2014年的结构效度较低。

5 结语

本研究基于巴赫曼交际语言测试理论框架和问卷调查结果,通过实验对比2014年和2015年改错在信度和效度的差异,有了一些实质性发现,具体包括以下三种:

第一,信度方面,2015年的信度高于2014年,但这两年的信度系数均不够高,尽管其结果是可接受的。

第二,效度方面,两年改错都具有中等程度的表面效度

和结构效度,以及较高等度的内容效度,其中2015年的结构效度和表面效度高于2014年,但内容效度低于2014年。

第三,2015年的正确率高于2014年。

此外,本研究也存在一些不足,如研究样本量不够大以及纳入测试的真题不过多等。但本研究采用了层次分析法的分类方法帮助理清了改错的考点分布情况,有着一定的现实意义。同时,对改错的效度和信度分析以及对问卷结果的研究应用,这些事实证据兴许都能给予测试者一些启示,如题目设计更重视语篇理解和逻辑推理,以更好衡量考生的语言能力,而不是单一的语言知识储量。

参考文献

- [1] 韩宝成. Lyle F. Bachman 的语言测试理论模式 [J]. 外语教学与研究, 1995(01):55-60.
- [2] Hymes, D. H. On Communicative Competence [A]. In J. B. Pride, J. Holmes (Eds.), *Sociolinguistics: Selected Readings* [C]. Harmondsworth: Penguin.
- [3] Canale, M., Swain, M. Theoretical bases of communicative approaches to second language teaching and testing [J]. *Applied Linguistics*, 1980,1(01):1-47.
- [4] 薛荣. 交际语言测试:理论模式与评估标准 [J]. 外语教学, 2008,29(03):68-71.
- [5] Bachman, L. F. *Fundamental considerations in language testing* [M]. Oxford: Oxford University Press, 1990.
- [6] 李筱菊. 语言测试科学与艺术 [M]. 长沙:湖南教育出版社,1997.
- [7] 孙成岗. 现代语言测试与试卷分析 [J]. 解放军外国语学院学报, 2000,23(04):82-86.
- [8] 刘建达, 杨满珍. 成段改错试题考了什么? [J]. 现代外语, 2001(02):170-180.
- [9] 吴明隆. 问卷统计分析实务——SPSS 操作与应用 [M]. 重庆:重庆大学出版社,2010.