

王 岚 主编

TEM-4客观试题效度研究

陈晓扣 著



 复旦大学出版社



TEM-4客观试题效度研究

陈晓扣 著

图书在版编目(CIP)数据

TEM-4 客观试题效度研究 / 陈晓扣著. —上海: 复旦大学出版社, 2009. 6

ISBN 978-7-309-06494-0

I. T… II. 陈… III. 英语—高等学校—水平考试—试题—研究 IV. H310.42

中国版本图书馆 CIP 数据核字 (2009) 第 018223 号

TEM-4 客观试题效度研究

陈晓扣 著

出版发行 **复旦大学出版社** 上海市国权路 579 号 邮编: 200433

86-21-65642857(门市零售)

86-21-65100562(团体订购) 86-21-65109143(外埠邮购)

fupnet@fudanpress.com <http://www.fudanpress.com>

责任编辑 于文雍

出品人 贺圣遂

印	刷	句容市排印厂
开	本	890×1240 1/32
印	张	9.375
字	数	261 千
版	次	2009 年 6 月第一版第一次印刷

书	号	ISBN 978-7-309-06494-0 / H·1282
定	价	28.00 元

如有印装质量问题, 请向复旦大学出版社发行部调换。

版权所有 侵权必究

解放军外国语学院英语博士文库

主编

王 岚

编委

(以姓氏笔画为序)

王 岚	孙建中	孙致礼	朱永生	严辰松
何其莘	吴 冰	张世平	李公昭	李绍山
李经纬	杨惠中	陆谷孙	姚乃强	姚云竹
徐烈炯	梅仁毅	黄国文	黄曾阳	程 工
熊学亮	潘永樑			

序

在陈晓扣同志的博士学位论文《TEM-4 客观试题效度研究——整体检验法》即将付梓之际，写上几句，既抒愉悦之情，亦表祝贺之意。

从 1993 年开始，陈晓扣作为我指导的第一个硕士研究生，就对语言测试产生了浓厚兴趣，其后兴趣愈浓，用心愈专，成果连连，学术影响也日渐显现。此篇论著正是陈晓扣多年不懈努力的结果，也是其语言测试研究长期积累的结果。作为导师，对于陈晓扣的成长和进步，我感到由衷的高兴和欣慰。

这里，先就语言测试说上几句，权作一孔之见。首先，在过去的几十年里，语言测试的地位发生了巨大变化。语言测试，尤其是外语测试，是随着外语教学而出现的，其应用性比较突出，但现在它越来越多地被作为研究手段用于有关语言学习和使用等重大问题的研究，在一定程度上，它已经开始具备解释力，同时它在应用语言学中也从实用性的地位上升到了理论性的地位，从边缘地位上升到了主导地位，其内部一致性大大增强了，产生了可以称之为“学术性语言测试”或用于研究目的的测试。其次，对于语言测试的研究也发生了明显的变化。一是从单纯测试实践经验的总结尤其是对命题技巧的探讨转向对理论层面的关注；二是从对测试结果的分析转向对测试

过程(尤其是受试心理过程)的分析;三是从孤立地研究受试在测试中的表现转向关注其与非测试环境中的表现之间的关系;四是在研究方法上从定量分析转向定量与定性分析相结合。

讲以上这些,并非要对语言测试研究的现状进行全面、严谨的学术评价,而是作为读者阅读和评价陈晓扣同志论文意义与价值的一个大体参照。简言之,我以为,陈晓扣同志在前人研究的基础之上,将对效度的理论研究向前推进了一大步,拓宽了研究视角,创新了研究方法,首创了反映效度整体观的效度检验程序和模型,为今后的效度研究提供了理论基础和方法借鉴。当然,恰恰由于该研究的创新性,不少方面尚处于探索之中,还有待作者及其他研究者在今后的研究实践中加以改进和完善。

另外需要说明的是,陈晓扣同志的研究是以 TEM-4 为例。TEM 是目前国内几个主要大型考试之一,经过多年的探索和实践,其科学性、权威性和影响力已被学界和社会所广泛认可,选择 TEM-4 作为研例,无疑会提高该研究成果的推广价值;此外,我本人作为外指委委员也多年参与 TEM 的相关工作,对该项考试较为熟悉。但无论以何种考试为例,都无意对该考试本身作出评判;最重要的不是针对该考试所得出的具体结论(无论正确与否),而是该研究所体现的理论见解、所采用的研究方法、所得出的带有普遍意义的结论。

最后,我想说,陈晓扣同志虽然在语言测试研究上辛勤耕耘,成绩斐然,但我国的语言测试研究依然任重道远,希望他把此专著的出版作为新的起点,不断取得新的成就,做出新的贡献。

是为序。

李绍山
2009 年 4 月

前 言

效度是语言测试涉及的一个根本问题。自效度概念于 20 世纪 30 年代提出以来,其研究的重点经历了最初的效标定义、50 年代至 80 年代中期类型的划分,发展到当今的整体观:“效度是指证据和理论在多大程度上支持考试使用所需的分数解释。”(American Psychological Association, 1999:9)但是纵观国内外大量的效度研究,我们发现两大问题:(1)在理论层面上,影响分数解释的“构念”无关和“构念”“污染”、“构念”代表性不足、内容代表性程度以及各种证据间关系等一系列问题都没有得到解决;(2)在实证层面上,绝大多数效度研究依然采用单一证据收集方法(主要是定量研究中的相关法),无法体现效度整体观的本质,研究结果难以令人信服。因此,从理论层面上创建一个科学的效度研究模型,解决上述效度研究面临的问题,体现效度整体观的本质,并指导效度实证研究是当务之急。

TEM-4 自 1990 年在我国实施以来,影响不断扩大,考生人数逐年增加,到 2008 年达到数十万人,成为我国一项重要的大规模考试。由于考生能否通过 TEM-4 直接决定他们是否有资格参加 TEM-8,而 TEM-8 证书又为社会所普遍看重,因此可以说 TEM-4 也是一项高风险的考试。经过多年的摸索和实践,TEM-4 日臻完善,具备了较高的科学性和权威性,为检验和促进我国英语专业教学质量做出

了重要贡献。随着 TEM-4 从 2005 年起执行新的《高等院校英语专业教学大纲》(以下简称《教学大纲》),该考试的题型和难度都发生了一定的变化(主要体现在客观试题方面),但迄今为止,尚无人对新 TEM-4 进行过认真和系统的效度研究。另一方面,客观试题本身的效度长期饱受争议,其根本原因就是客观试题分数只是对考试行为的一种数字概括,本身并不能直接反映考生对所考知识、技能或能力的掌握与否。因此,运用最新的效度观对新 TEM-4 客观试题进行整体的效度检验和科学论证就显得极其重要和迫在眉睫。

基于上述考虑,本书在前人的研究基础上,首先在理论层面上构建一个效度整体检验模型,然后以该模型为指南并结合其他语言学理论来检验 TEM-4 客观试题的效度。

就笔者自身而言,真诚地希望本书提出的效度整体检验模型能够在理论上有功于拓宽我国语言测试效度研究的视野,为我国大型考试的效度研究提供切实可行的理论基础;在研究应用上,笔者也希望本书对于如何从多角度、多层次对我国的各种考试进行科学的效度实证研究提供很好的借鉴。

在本书付梓之际,我由衷地感谢我的恩师李绍山教授,正是在他的悉心指导、热情鼓励和不断教诲下,我从一个语言测试的门外汉渐渐地对语言测试这门学科产生了浓厚的兴趣并取得了一定的成果。可以说,导师渊博的学识、严谨的治学态度、细致的工作作风、孜孜不倦和实事求是的科研精神令我终身难忘。

此外,坦诚而言,笔者撰写此书时本着实事求是的态度,无意对 TEM 作出任何评判,诚如恩师在序中所言:“……无论以何种考试为例,都无意对该考试本身作出评判;最重要的不是针对该考试所得出的具体结论(无论正确与否),而是该研究所体现的理论见解、所采用的研究方法、所得出的带有普遍意义的结论。”同时,由于笔者水平有

限,书中会有不妥之处,也希望得到专家和同行们的指正。

最后感谢我的母校解放军外国语学院和复旦大学出版社对本书的出版所给予的多方面的帮助。

2009年4月于洛阳

目 录

第一章 绪论	1
1.1 研究目的和意义	3
1.2 研究方法	3
1.3 本书结构	4
第二章 语言测试发展概述	5
2.1 科学前语言测试	5
2.2 心理计量—结构主义语言测试	6
2.2.1 心理计量学	6
2.2.2 分离式测试	11
2.3 综合性测试	12
2.4 交际语言测试	15
2.4.1 Hymes 的交际能力模式	15
2.4.2 Canale 和 Swain 的交际能力模式	16
2.4.3 Bachman 的交际语言能力模式	17
2.4.4 Bachman 的测试方法层面模式	23
2.5 小结	28

第三章 效度研究	30
3.1 效度定义及演变	30
3.2 Messick 效度检验程序和方法	33
3.2.1 逻辑分析	34
3.2.2 证据收集	35
第四章 TEM-4 客观试题效度研究模型	49
4.1 国外语言测试界效度实证研究现状	49
4.2 国内语言测试界效度实证研究现状	50
4.3 目前效度研究面临的问题及解决方法	52
4.4 TEM-4 客观试题效度研究模型	54
第五章 TEM-4 客观试题效度研究的目的、 问题和方法	56
5.1 研究目的	56
5.2 研究的问题	58
5.3 研究步骤和方法	58
5.4 研究对象及实验过程	59
5.5 预期答题过程判断和内容专家判断	60
5.6 统计及分析	60
第六章 TEM-4 听力理解测试效度研究	61
6.1 听力理解原理	61
6.2 影响听力理解的因素	63
6.2.1 听者自身因素	63
6.2.2 外部因素	64
6.3 听力分技能	67
6.3.1 阶段法	67
6.3.2 认知技能法	68

6.3.3 交际分类法	68
6.4 听力理解“构念”定义法	70
6.4.1 基于能力的听力“构念”定义法	71
6.4.2 基于任务的听力“构念”定义法	71
6.4.3 能力—任务交互定义法	72
6.4.4 小结	72
6.5 TEM-4 听力理解测试内容相关证据	72
6.5.1 数据统计结果	73
6.5.2 数据统计结果分析和讨论	77
6.6 TEM-4 听力理解测试“构念”相关证据	85
6.6.1 TEM-4 听力理解测试的“构念”及预期答题过程	85
6.6.2 听力考试应试技巧	87
6.6.3 数据统计及分析	88
6.7 T 检验	106
6.8 小结	108
第七章 TEM-4 完型填空测试效度研究	110
7.1 完型填空测试原理	110
7.2 影响完型填空测试的因素	111
7.3 TEM-4 完型填空测试内容相关证据	112
7.3.1 数据统计结果	113
7.3.2 数据统计结果分析和讨论	115
7.4 TEM-4 完型填空测试“构念”相关证据	118
7.4.1 TEM-4 完型填空测试所考查的“构念”及 预期答题行为	119
7.4.2 数据统计和分析	120
7.5 T 检验	131

7.6 小结	133
第八章 TEM-4 语法、词汇测试效度研究	135
8.1 语法、词汇多项选择填空测试原理	137
8.1.1 语法“构念”定义	137
8.1.2 词汇“构念”定义	138
8.1.3 影响语法、词汇多项选择填空测试的因素	139
8.2 TEM-4 语法、词汇测试的内容相关证据	140
8.2.1 数据统计结果	141
8.2.2 数据统计结果分析和讨论	144
8.3 TEM-4 语法、词汇测试“构念”相关证据	147
8.3.1 TEM-4 语法、词汇测试所考查的“构念”及 预期答题过程	147
8.3.2 数据统计和分析	148
8.4 T 检验	160
8.5 小结	161
第九章 TEM-4 阅读理解测试效度研究	164
9.1 阅读理解原理	164
9.1.1 自下而上模式	165
9.1.2 自上而下模式	165
9.1.3 交互作用模式	166
9.1.4 图式理论	166
9.1.5 阅读本质	168
9.2 阅读成分技能	169
9.3 影响阅读理解的因素	172
9.3.1 读者自身因素	172
9.3.2 篇章因素	173

9.3.3 题目因素	175
9.4 TEM-4 阅读理解测试内容相关证据	176
9.4.1 数据统计结果	177
9.4.2 数据统计结果分析和讨论	180
9.5 TEM-4 阅读理解测试“构念”相关证据	184
9.5.1 TEM-4 阅读理解测试“构念”及预期答题过程	184
9.5.2 应试技巧	186
9.5.3 数据统计和分析	186
9.6 T 检验	198
9.7 小结	199
第十章 讨论	201
10.1 “构念”理论与效度	203
10.2 教学大纲或语言使用与效度	203
10.3 考试方法层面理论与效度	204
10.4 小结	208
第十一章 总结	210
11.1 主要研究发现	210
11.2 对效度研究的启示	212
11.3 本书研究的不足与建议	212
附录一 TEM-4 听力理解测试内容相关证据问卷调查表 ..	214
附录二 TEM-4 完型填空测试内容相关证据问卷调查表 ..	222
附录三 TEM-4 语法、词汇测试内容相关证据问卷调查表	226
附录四 TEM-4 阅读理解测试内容相关证据问卷调查表 ..	231
附录五 TEM-4 客观试题真题(2005)	235
参考文献	263

第一章

绪 论

随着我国与国际社会在文化、经济、教育、科技等诸多领域交流与合作的进一步深入,我国对各种外语人才的需求日益增大,各种衡量外语学习者外语水平的考试也应运而生。考试与教学相辅相成,密不可分。一方面,考试用于检验和评估教学质量;另一方面,考试又对教学产生反拨作用。从严格意思上来说,任何考试,特别是高风险考试,都要经受质量的检验,因为考试的高质量将直接对教学产生积极的促进作用,反之就会误导教学,影响教学质量的提高。效度和信度是检验考试质量的两大主要指标,其中效度又尤其重要,因为它是语言测试的基本出发点。

自效度概念于 19 世纪 30 年代提出以来,效度研究一直备受语言测试学和统计学等学科的关注。随着语言学理论的不断拓展,语言测试学相应地逐渐修改效度定义,使其日臻完善。针对效度定义的不断变化,统计学也不断探索和研究新的统计工具,为效度的定量研究提供便利。总之,多门学科都以自己独特的方式对效度研究做出了贡献,为我们理解效度的复杂性提供了多方位的视角。当前最新的效度研究结果表明,效度是一个整体概念,效度研究不仅需要证据收集,而且也要理论检验。因此,要想对考试分数作出恰当、有意义的解释,我们不但需要联系教学大纲或语言使用、“构念”理论、考

试方法理论,更要将它们融为一体。

客观试题以其覆盖面宽、抽样量大和信度高而受到人们的青睐,是当今绝大多数语言测试普遍采用的题型之一。但同时,客观试题的效度也一直备受人们的质疑,其根本原因就是考试使用者无法洞悉客观试题的分数到底是不是考生的语言知识或语言技能的真正体现,即人们对猜测因素到底在多大程度上帮助考生做对题一直心存怀疑。尽管有些语言测试学家(如 Heaton, 1988)认为考生单凭猜测很难做对四选一或五选一多项选择题,但是这种观点是建立在客观试题经过精心和科学设计的基础上的。换言之,一旦客观试题的设计不够严谨,不够科学,出现“构念”“污染”或无关等问题,客观试题的效度就会面临很大的挑战。因此,对客观试题的效度进行检验就显得极其重要。

TEM-4 自 1990 年在我国实施以来,已成为我国一项重要的英语考试,为检验和促进我国英语专业教学质量做出了重要贡献。随着新的《教学大纲》的颁布,以及 TEM-4 难度和题型的变化,对新的 TEM-4 进行科学的效度检验,特别是对 TEM-4 客观试题(占整个试卷分值的比例约 65%)进行科学的论证显得极其重要。

但是目前的效度研究无论在理论上还是在实证上尚有许多地方不尽如人意:首先,在理论层面,影响分数解释的“构念”“污染”、“构念”无关、“构念”代表性不足、内容代表性以及各种证据间关系等一系列问题都没有得到妥善解决;其次,在实证研究层面,目前的绝大多数效度检验依旧直接利用卷面分数,采用定量分析的方法来获取“构念”证据,忽略了“构念”理论、考试方法层面理论在效度研究中的作用;最后,目前国内对效度的研究,无论是理论上还是实证上,都十分欠缺,一些大型的英语考试更是没有经过效度检验。鉴于上述考虑,本书从效度最新定义的内涵和外延角度出发,同时借鉴前人的研

研究成果,首先系统地论证如何进行科学的效度检验并以此为基础创立客观试题效度研究的理论模型,然后以该模型为指导对 TEM-4 客观试题进行效度检验。

1.1 研究目的和意义

本书的目的主要体现在以下几个方面:

(1) 从效度最新定义出发,系统地分析效度的内涵和外延,建立客观试题效度检验的模型,使读者更好地认识和理解如何解决客观试题中的“构念”“污染”、“构念”无关、“构念”代表性不足和内容代表性程度等诸多问题,为语言测试的效度研究构建切实可行的研究模式;

(2) 运用客观试题效度检验模型来检验 TEM-4 客观试题的效度,为我国大型语言测试的效度研究提供具体可行的实例,也为丰富、发展和完善效度理论提供实证依据;

(3) 最近十年来,我国语言测试的效度研究已迈出了可喜的一步,取得了不菲成绩,发表和出版了一些效度研究的文章和专著,但主要的研究方法是定量分析,定性分析方法少,将受试答题过程分析法和专家判断融为一体对 TEM-4 客观试题进行效度研究更是空白。本书将弥补这一方面的不足,为我国的各种语言测试提出一些建设性建议。

1.2 研究方法

本书采用定性分析为主、定量分析为辅的方法。首先通过 TEM-4 内容相关证据问卷调查、个别访谈等方法,获取 TEM-4 考试