

ENGLISH LANGUAGE TESTING

—SOME THEORETICAL
AND PRACTICAL
CONSIDERATIONS

英语 语言 测试

——理论与操作

邹 申 主编



上海外语教育出版社

英语语言测试

理论与操作

邹 申 主编



上海外语教育出版社

H311

英语语言测试
——理论与操作
邹 申 主编

上海外语教育出版社出版发行

(上海外国语大学内)

江苏昆山亭林印刷总厂印装

新华书店上海发行所经销

开本 850×1168 1/32 8.5 印张 5 插页 211 千字

1998 年 7 月第 1 版 1998 年 7 月第 1 次印刷

印数: 1 000 册

ISBN 7-81046-399-3

G·676 定价: 12.00 元

本版图书如有印装质量问题,可向承印(订)厂调换。

序

《英语语言测试—理论与操作》是由上海外国语大学英语学院中青年教师为主体编写的论文集。论文撰写者的平均年龄为 33 岁。他(她)们均在国内外接受过正规的测试理论培训或进修,并多年来参与全国高等院校英语专业学生考试(TEST FOR ENGLISH MAJORS - TEM4&8)的命题、审题、预测、阅卷和分析等工作。这些同志既掌握测试理论知识,又积累了实践经验,在完善和稳定考试质量方面起到了很大的作用。受国家教委高等学校外语专业教学指导委员会英语组的委托,全体同志在 1993 - 1996 年间参与了 TEM 考试效度研究项目。该项目采用定性和定量的手段,对现有 TEM 考试进行了多层次、多方面的评估。在此基础上撰写的 TEST FOR ENGLISH MAJORS(TEM)VALIDATION STUDY(上海外语教育出版社,1997)得到了有关部门与专家的一致好评,并获得 1996 年度上海市教委教学成果一等奖。

本论文集是 TEST FOR ENGLISH MAJORS(TEM)VALIDATION STUDY 的续集。论文集共收集论文 11 篇,术语介绍 1 篇。文章是同志们在多年科学实践和调研的基础上,总结考试相关方面的经验而写成。文章内容涉及听、说、读、写技能的评估,间接数据的使用,命题方式,出题技巧,考试分析,常用术语解释等方面。

论文集的作者们以应用语言学和测试学的理论、原则、方法为

指导,结合近几年亲身经历的 TEM4 /8 考试实践,对 TEM 考试效度研究作了全面而系统的科学总结,对 TEM 考试所涉及到的或今后会涉及到的阅读、写作、听力、语法、口语等方面的测试,对 TEM 考试的题型、题量、出题技巧以及对考试的评估等均作了精辟的分析。论文集文章论述的主题明确,理论依据充分,引用的数据亦令人信服。其中每篇论文均围绕重点展开,主题明确,并各有特色。同时,论文集做到理论与实际相结合。在进行理论分析的同时,又注重测试的实用性,许多例子与操作方法具有较高的实用价值。因此,论文集无论在测试理论方面,还是在测试实践方面,对测试工作者和教师都有极高的参考价值。论文集的出版对推动英语语言测试科学化有着积极的意义。

上海外国语大学 校长

全国高等学校外语专业教学指导委员会 主任

戴炜栋 教授

1998 年 1 月

目 录

1. TEM 考试效度研究项目总结报告	邹 申	·1·
2. 从阅读能力到阅读测试	冯慎宇	·22·
3. 英语专业学生考试(TEM)中写作能力的 评估	邹 申	·60·
4. 科学公正地测试学生的口语能力	黄素华	·90·
5. 语用测试之听写及 TEM4 考试	张艳莉	·115·
6. 间接数据在考试评估中的运用 —— 问卷调查实例分析	邹 申	·131·
7. 多项选择题出题技巧及应用	黄素华	·157·
8. 一种值得推荐的出题方法 —— “头脑构图法”(MINDMAPPING)	冯慎宇	·180·
9. TEM4 和 TEM8 中的听力理解	黄素华	·191·
10. 1995 年全国高校英语专业四级统测 —— 分析与建议	陈汉生	·214·
11. 专业语法教学与英语专业学生四、八级 考试的反拨作用	何琮周	·230·
12. 语言测试术语及其运用	黄素华	·239·

TEM 考试效度研究项目总结报告

邹 申

1993 年 9 月在国家教委有关部门和英国文化委员会的支持下, 高等学校英语专业学生考试(以下简称 TEM4 /8)效度研究项目正式成立。项目研究的目的是对 TEM 考试进行全面的效度研究。其具体过程是通过科学的测量途径及方法, 对考试进行多方位的评估, 并在此基础上提出对考试的改进意见, 以进一步提高考试的效度与信度, 使考试结果更具有合理性、科学性。项目研究历时三年, 于 1996 年 6 月完成。本文是对项目内容及实施过程的总结。

一、TEM 考试效度研究目的

1. TEM 考试介绍

a. 考试目的与性质

TEM4 /8 分别从 1990 年与 1991 年开始实施。考试的目的是为了检查《高等院校英语专业基础阶段教学

大纲》与《高等院校英语专业高年级教学大纲》(以下简称教学大纲)的执行情况。从这个意义上说,TEM4 /8 是属于尺度参照性考试,因为这两种考试均以教学大纲的标准作为测量的依据,以大纲的教学内容作为测试内容的基础,而同时又有别于一般的成绩考试,不以某教材或某课程为准绳,而是参照了应用语言学理论和测试学的原则和方法。从这点上讲,TEM4 /8 又兼具水平测试的特征,当然,严格地说,TEM4 /8 属于尺度参照性水平测试的范畴。

b. 考试内容

鉴于 TEM 考试目的是为了检查各校对教学大纲的执行情况,故测试的内容力求反映大纲的要求。然而,由于两大纲在不同阶段的教学过程中实施不同的标准,TEM4 与 TEM8 在测试的侧重点上也就有所不同。

TEM4 主要测试听力、阅读及写作三大语言技能,及语言基本知识,以评估学生的语言综合技能。根据这个精神,TEM4 试卷中既有考核学生单项语言技能的项目,如听力理解、语法及词汇等;也有考核语言综合技能的项目,如听写、完形填空等。

TEM8 的测试内容基本上与 TEM4 相一致。所不同的是,TEM8 更注重学生语言运用的能力评估,如听力理解中有记笔记后填空的项目、改错等。

虽然口语能力的培养在两大纲中都占有重要的地位,但由于具体实施上的问题尚有待于进一步探讨、解决,因而目前 TEM 考试暂不包括口试内容。

c. 题型与比重

TEM 的题型采用国内外较常见的模式,即客观题与主观题相结合。两种题型的结合使用,主要是考虑其在提高效率与信度方面的互补性。

在 TEM4 考试中客观题占总分的 65%,主观题占 35%。而在 TEM8 考试中客观题则占总分的 40%,而主观题占总分的 60%。

TEM 考试效度研究项目总结报告

TEM4 与 TEM8 题型比重上安排的不同旨在贴近教学大纲内对两个阶段教学重点的区别。基础阶段的教学大纲侧重语言基本知识及单项语言技能,同时又注意培养学生的语言综合技能及交际能力。参照基础阶段大纲的基本精神,TEM4 考试中有相当部分的客观题,是测试学生的语言知识及单项语言技能。高年级教学大纲中对学生语言水平的要求从语法、技能的掌握转到语言的运用。为体现高年级教学大纲的这个根本教学要求,TEM8 考试中主观题多于客观题,着重测试学生的语言运用能力。

d. 考生概述

TEM4 /8 考生人数每年都在增长。下列两表显示历年来 TEM 考生及学校的数目(到 1997 年为止)。

TEM4 考试

年	学生数	学校数
1992	8554	155
1993	9582	163
1994	11070	173
1995	13675	193
1996	15776	212
1997	19106	247

TEM8 考试

年	学生数	学校数
1992	4613	86
1993	4459	89
1994	5163	103
1995	6325	117
1996	7377	129
1997	9456	169

考生遍布全国各地:有沿海发达地区,也有内陆及边疆地区。考生来自五类院校的英语以及相关专业:综合性大学、理工科大学、外语院校、师范院校与其他院校。其中有全国重点院校;有地方院校;有建校历史长的学校,也有新近成立的学校。考生主要为汉族,也有我国其他少数民族,如维吾尔族、哈萨克族等。

e. 试卷结构(1990—1995 年)

为了进一步完善考试以提高考试的科学性和合理性,TEM 考试自开考以来在题型、比重、项目等方面进行了一系列的修改。1990 至 1995 年间考试的变化情况如下:

TEM4 考试

听力理解 由最初的四个部分改为三个部分,比重由 20% 降至 15%。削减出的 5% 增加在听写项目上,提高了听写在试卷中的比例,以突出其重要性。听力理解项目的题目增加题干,以便与 TEM8 考试听力部分的题目形式保持一致。

听写 该项目的比重由 10% 增至 15%,以便更好地体现教学大纲的要求。

校对改错 该项目从 1993 年起取消。

写作 该项目的题目由一个增至两个,比重从 15% 增至 20%,以便更有效地衡量考生的写作能力。

语法与词汇 该项目由原先的两个单独项目合并而成,比重从 10% 增至 15%。

阅读理解 该项目分为两个测试不同阅读技能的部分,以便有效地考核学生仔细阅读与快速阅读的能力。从 1995 年起快速阅读部分的题目放在阅读材料之前,以便较为真实地体现实际快速阅读环境。

TEM8 考试

TEM 考试效度研究项目总结报告

听力理解 该项目由四个部分减为三个部分,以减少测试内容上的重复,比重也相应地由 20% 降至 15%。

填空 1994 年该项目的时由原先的 15 分钟增至 20 分钟,以降低难度。同时,填空接在 PART TWO 校对改错后,而以前该项目是在 PART THREE 阅读理解之后。

校对改错 该项目的题型由客观题改为主观题,比重也从 15 分增至 20 分。

快速阅读 这部分的比重从 10% 降至 5%,与 TEM4 试卷一样,该部分的题目现印在阅读材料之前。

翻译 该项目的比重由 15% 增至 20%,以突出翻译能力测试的重要性。时间也由 45 分钟增至 60 分钟。考生通常只需翻译所给材料的划线部分,其余部分为翻译提供上下文。

写作 该项目的比重从 15% 增至 20%,以体现写作能力测试的重要性,时间也从 45 分钟增至 60 分钟。

2. 效度研究

a. 考试现状

自 TEM4 /8 考试实施以来,其影响在全国各校不断扩大。各校把考试作为他们检查教与学水平的一种方式,作为衡量本校学生英语能力的一把尺子。总而言之,各校对考试的反映是积极、肯定的。

然而,作为一种在全国范围内实施的大规模考试,TEM 在当时(至 1993 年)尚处在起步阶段,在许多方面还不成熟,不完善。主要表现在 TEM 考试尚未达到国内外大型考试的标准:理想的效度与信度。比如,当时未对 TEM 的考试效度进行深层次、多方面的调查研究;在保持和提高考试的信度方面尚未找出有效的统计方法。在主观题阅卷方面存在的问题是:1. 评分标准不够稳定,且有的部分未客观地、科学地反映考生的实际水平;2. 对评

阅人未有足够、有效的培训。总之,考试命题人员和设计人员的测试理论水平和知识以及评阅人员的水平有待于进一步提高。

要使 TEM 考试向前发展,真正成为国内有影响、有权威性的考试,成为衡量英语专业学生英语水平有效的、科学的手段,进一步提高 TEM 考试的科学性与合理性成了当务之急。

b. 项目建议及启动

TEM 考试至 1993 年已实施数年,其效果到底怎样?是否评估了所预期评估的语言技能及语言能力?是否经得起公众的检验?第一线的老师对考试的反映如何?考生的意见又是如何?这些都是亟需解答的问题,也是进行效度研究的初衷。用 HENNING (1987)的话来说,考试效度指的是某一考试或其任何一个组成部分是否恰当地测量了预期测量的内容。换句话说,也就是考试的内容、手段是否与考试的宗旨相一致。

全国高等院校外语专业教学指导委员会英语组在 1992 年的年会上,提出了要对 TEM 考试进行评估。英语组希望,对考试的一系列评估会进一步改善 TEM 作为全国性考试的形象,使之具有作为测试英语专业学生语言水平考试的权威性。英语组这一建议得到了国家教委有关部门的采纳。在教委有关部门的推动下,英国文化委员会与 TEM 考试中心所在单位上海外国语大学进行了接触。英方对建议的实施表示有兴趣。同年,英国文化委员会委托当时在上海交通大学 CET 中心进行工作访问的 C.J. WEIR 博士来 TEM 考试中心具体商谈考试效度研究项目的立项事宜。双方经过反复商讨后写出项目立项建议书以及项目研究任务书。项目建议书及任务书于同年年底送交国家教委与英国文化委员会。1993 年 4 月,项目建议书获得了批准,成为英国文化委员会资助的 ELT 项目。项目预定于 1993 年 9 月开始,1996 年 6 月结束。

TEM 考试效度研究项目总结报告

c. 项目组成员

项目组中方成员为上海外国语大学英语学院教师。本文作者为项目负责人。在项目开展的过程中共有 10 位教师参与了项目的工作及培训。

项目组英方成员为一名项目常驻专家(1993—1995 年)及两名外聘顾问。外聘顾问在项目进行期间每年来测试中心工作两至三周。

在此期间,中方共有 7 位成员赴英国里丁大学(READING UNIVERSITY)接受 10 周的测试理论培训,一位成员接受为期一年的理论进修。

d. 项目目标及性质

项目的远期目标是:提高英语专业基础阶段与高年级阶段的教学水平。

项目的近期目标是:1)提高 TEM 考试的效度,更好地反映教学大纲中提出的教学目标;2)改进 TEM 考试设计,更有效地测量考生的语言技能及知识水平;3)改进 TEM 考试设计,对英语教学起到正面后效作用;以及 4)培训中方测试专业人员从事大规模考试研究的科研能力。

项目具有终结性和进行性双重性质。项目具有终结性质是基于 TEM4 /8 考试到 1993 年为止已分别实施了三年,所以有必要对前一阶段的工作进行总结与评估,使涉及考试的有关人员(如学校、教育管理部门、教师及学生等)了解考试的质量及特点。同时,考试又具有进行性质,这是因为效度研究不仅仅着眼于考试的现状,还注重考试的发展和未来。在总结和评估考试的基础上,项目将提出改进考试的具体建议,并对考试作一些修改,使考试能更有效地测出 TEM4、TEM8 学生的语言水平。

e. 项目研究进度

根据 TEM 考试的实际情况,项目组经过讨论决定将效度研究工作分三个阶段进行。

第一阶段(1993.9—1994.9)主要工作:对照教学大纲,回顾总结历年考试,找出其成功与不足之处;撰写新的考试规范,并参照考试规范编写新试题,力求做到新编试题(在项目过程中曾被称为 PROTOTYPE TESTS)更具客观性、代表性和科学性。

第二阶段(1994.9—1995.9)主要工作:完成写作、翻译、听写的评阅标准,与 TEM 考试配套;制订评阅人员培训方案和命题人员培训方案的初步计划;完成平行卷的命题;实施 PROTOTYPE TESTS、平行卷和用于共时效度研究的 TEEP 试卷;在 TEM4 和 TEM8 两个层次内,对教师、学生、监考人员进行大范围的问卷调查;对所得的数据进行计算机分析,并写出初步的分析报告。

第三阶段(1995.9—1996.6)主要工作:修正和研究计算机统计分析数据,并在此基础上提出对 TEM 考试的改进意见和建议,撰写出项目终结报告。

二、TEM 考试效度研究内容

考试效度研究是论证某一考试是否具有预期效度的过程。考试效度包括三个方面(ALDERSON, et al 1995):内部效度,外部效度和构想效度。内部效度主要研究考试的可见内容及其作用;外部效度研究考生成绩与其他外部测量能力的方法之间的关系;构想效度包含上述两项效度,主要是为了证实考试在理论上的可行性。效度研究方式通常有定量与定性两大种类。定量即通过统计手段对考试的效度进行量化分析,通过所获数据来论证考试的效度。定性即通过问卷调查、专家评估等途径,通过收集各方面的看法与意见来论证考试的效度。

TEM 考试效度研究内容分为上述三大方面。TEM 考试内部效度研究主要是指 TEM 考试内容效度及 TEM 考试信度的论证。TEM 考试外部效度研究包括 TEM 考试共时效度和预测效度。

TEM 考试效度研究项目总结报告

TEM 考试的构想效度研究,主要从统计的角度论证 TEM 考试及其各个项目测试的有效性。本章以下部分将分别叙述这三方面研究的内容及过程。

1. 内部效度

a. 内容效度

内容效度指考试的内容是否具有代表性。在考试效度研究中,内容效度的评估是一个必不可少的环节。考试是否测试了预期测试的内容,选择的内容样本是否具有代表性,这些都是衡量一个考试是否具有内容效度的标准。比如,口语测试若通过笔试的形式进行,或者旨在测试考生全面语言水平的考试只评估了考生的阅读能力等,这类考试就缺乏内容效度。如果考试缺乏内容效度,其他效度以及考试的信度就无从谈起。TEM 考试效度研究的一项重要内容就是考试的内容效度。为了保持和提高 TEM 考试的内容效度,项目组以 1994 年出版的英语专业四、八级考试大纲(上海外语教育出版社)为基础,制订了 TEM4 /8 考试规范各一套。考试规范是一份详细文件,明确规定了考试的目的、目标、形式及内容等项目。在考试规范的制订过程中项目组着重考虑以下几个因素:第一,考试规范必须与教学大纲的要求保持一致;第二,考试规范必须突出测试学生的整体语言能力和交际能力;第三,考试规范中试题形式必须是有效、可靠和可行的。

TEM 考试规范详细规定了 TEM 的测试目的、种类、评分方法、被试技能和考试具体操作方法,为考试设计、命题和评估提供了可参照的依据,同时为评估以后 TEM 考试的内容效度提供了一套明确的标准。

考试规范制订工作完成后尚须检验其科学性、合理性与可操作性。鉴于此目的,项目组下一步的工作便是按照考试规范的要求编写新的四、八级试题各一套(当时这两套试题被称为 PROTO-

TYPE TESTS)。为了使这两套题目成为以后试题的样板,试题编写人员在编写过程中,必须尽可能全面地反映考试规范的基本原则与要求。

试题编写完后要经过使用才能了解其统计性能,以证明参照依据——考试规范的可靠性、科学性、合理性和可操作性。试题的使用涉及到选用样本的规模。从统计学的角度来看,样本愈大,所获数据就愈能反映以后真实考生群体的特征,也就愈具有代表性。经过反复权衡,项目组决定将这两套试题作为 1995 年正式考试用题。这样做对以后的研究工作有两方面的益处:第一,样本选择上不存在是否具有代表性的问题,因为使用试卷的考生人数即为 1995 年参加考试的人数;第二,试卷使用后可对 TEM 考试进行全面的统计分析。采用实际考生群体将使所产生的数据能较真实地反映 TEM 考试的各类功能,从而有助于 TEM 考试的评估。

b. 统计效度

TEM 考试内部效度研究的另一目的是建立 TEM 的统计效度。具体地说,就是借助统计学上的一些原理,制订出适用于 TEM 考试特点的统计分析方式,以提高考试的统计性能,从而提高考试的信度。所谓信度,就是指分数的一致性、可靠性,这也是直接关系到考试是否有效的一个重要因素。一个具有较高效度的考试必须同时具有信度,不然便没有效度,因为任何一个考试最终是通过分数来表述其结果的。如果分数不可信,那么我们就无法说考试有效地评估了学生的语言技能或能力。在这方面,项目组主要做了以下工作:

(1) 制订科学的、可行的统计方式。1995 年 TEM 试卷的统计分析被视为建立科学的、可行的统计方式的起点。在英方统计专家的协助下,项目组对两套试卷进行了三个层次的分析: ITEM /COMPONENT /TOTAL TEST LEVEL。在工作的进展过程中,项目组一边从统计角度了解 TEM 考试的特征以及性质,

TEM 考试效度研究项目总结报告

一边总结出一套适应于 TEM 考试实际情况的统计操作方法。目前考试采用的统计方法是以当时的项目研究统计手段为基础。在三个层次上的统计分析使项目组既了解每道题的性能,项目与项目或项目与总分之间的关系,也了解整个考试的情况。这样,项目组掌握了每题具体详细的数据的同时,也掌握了考试的全面数据。这些数据给考试的修改提供了科学依据。例如,TEM4 /8 考试题量是否可以减少、减少多少等的决定都是在考试统计分析基础上作出的。目前无论在考前的试测中还是在考后的分析中我们都借助于统计方法,这样既保证了考试题目的质量,也便于检查。实践证明,这是提高考试信度的有效方法之一。

(2) 完善 TEM 考试命题及评阅的操作过程和规范操作程序。这方面开展的工作主要涉及两个方面:命题及评分。

i. 命题

考试是否具有信度,题目的本身是关键。为此,项目组下了一番功夫使命题过程逐步科学化。主要的做法如下:

a) 注意挖掘和培养命题人员;与他们保持经常的联系;每年定期定量向他们征集题目。命题人员的队伍不是一成不变的,根据命题的质量适当做一些调整。在条件许可的情况下组织命题人员培训以及提供命题质量的信息反馈。

b) 征集来的各类题目由专人筛选、修改或提供意见。在第一轮修改的基础上集体讨论题目并作必要的调整。

c) 在拼卷的过程中严格参照教学大纲以及考试大纲的要求,做到不超纲,不偏离规定的测试范围。同时,我们也注意所选用的文章、材料与体裁的搭配,尽量做到不“撞车”,力争较大的题材、体裁覆盖面。此外,题材的选用尽量向各专业具有共性的类别靠拢。

d) 将题目放在一定的范围内试测,随后根据试测的数据对题目作最后的取舍决定。

e) 将试卷的修改稿提交专家审定,最后定稿。