



# 复旦大学英语水平考试 的开发与效度研究

Development and Validation of the Fudan English Test

范劲松 季佩英◎著

 复旦大学出版社

## 图书在版编目(CIP)数据

复旦大学英语水平考试的开发与效度研究/范劲松,季佩英著. —上海:复旦大学出版社,2019.12  
ISBN 978-7-309-14801-5

I. ①复… II. ①范…②季… III. ①大学英语水平考试-研究-中国 IV. ①H310.421

中国版本图书馆CIP数据核字(2019)第288378号

### 复旦大学英语水平考试的开发与效度研究

范劲松 季佩英 著  
责任编辑/朱莉芝

复旦大学出版社有限公司出版发行  
上海市国权路579号 邮编:200433  
网址:fupnet@fudanpress.com <http://www.fudanpress.com>  
门市零售:86-21-65642857 团体订购:86-21-65118853  
外埠邮购:86-21-65109143  
当纳利(上海)信息技术有限公司

开本 890 × 1240 1/32 印张 6.625 字数 158 千  
2019年12月第1版第1次印刷

ISBN 978-7-309-14801-5/H · 2957  
定价:48.00元

---

如有印装质量问题,请向复旦大学出版社有限公司发行部调换。  
版权所有 侵权必究

# 序

校本英语测试是由高校依据本校的大学英语课程要求开发和实施的学业测试。测试对象是本校非英语专业的学生；测试目的是检验学生是否达到了本校大学英语课程的要求。与大规模、标准化的英语考试相比，校本英语测试有其独特的优点。由于校本测试为各高校自主研发，因此其测试内容和方法与其所在高校的教学大纲等联系紧密，而且也能够尽量充分地考虑到本校学生的特点；校本英语考试规模相对较小，因此在考试的设计上具有更大的灵活性。鉴于以上这些特点，校本英语测试对大学英语教学能够产生更为积极的反拨效应。但是，校本英语测试也面临一些问题。众所周知，语言测试是一个跨学科领域，测试的开发、实施和效度研究需要从业者具备一定的专业素养。校本英语测试一般由各高校的大学英语教学部门负责研发，但是大学英语教学部门不一定具有经过语言测试专业培训的教师，因此在试题质量控制、审题、拼题、预测、评分、数据处理、成绩报告等环节上可能有一定的困难，也可能存在一些不足之处。同时，测试体系的质量评估和效度研究则更需要具有语言测试专业背景的研究者。

复旦大学是我国较早开发和实施校本英语测试的高校之一。复旦大学的大学英语教学从课程设置到师资力量等方面在我国均处于领先地位。难能可贵的是，作为一项校本英语测试，复旦大学英语水

平考试(Fudan English Test,简称FET)在测试开发、实施、评分和数据分析等方面参照了语言测试领域被广泛认可的行业标准和行为规范,如《国际语言测试协会良好测试行为规范》《教育与心理测试标准》等,同时也参照了我国大规模、标准化考试的实践,严格控制测试的每一个步骤,确保试题的质量和考试的信度、效度等。

自2011年FET项目正式启动以来,本书的两位作者多年来奋战在校本测试的第一线,指导并参与了FET的设计、开发、实施、评分、数据处理和效度研究等各个环节的工作,积累了丰富的语言测试实践经验。在本书中,两位作者详细报告了FET的效度研究工作。FET的效度研究采用了目前语言测试研究领域被广泛认可的效度理论框架——解释性效度论据理论。本书一共报告了五项FET效度研究,对应解释性效度论据框架中的不同推断。这些效度研究采用多种研究方法,从不同的角度对考试的效度进行论证,从而为FET的效度提供了大量证据,也从中发现了考试在将来需要不断完善和改进之处。

作为我国校本英语测试项目的代表性成果之一,FET项目不仅为评价大学英语学习提供了有重要参考价值的理论和实践经验,也同时为语言测试领域的人才培养提供了重要契机。FET的核心团队由大学英语教学部的十几位英语教师组成,其中包括语言教学和测评方面具有丰富经验的教师。团队成员在FET项目中紧密合作,确保考试开发和实施的顺利进行。近几年,FET团队获得了包括“复旦大学教学成果特等奖”“上海市教学成果一等奖”等荣誉。可以想象,这些荣誉的背后是辛苦的付出。一个语言测试项目的运转需要众多人员的参与和多方面的支持。FET的组织和运行模式对我国其他高校开展校本英语测试具有重要的借鉴与启示作用。

期待复旦校本英语测试能够从测试理念、测试内容和方法等方面不断改革和创新,建立更加完善的大学生英语能力综合评价体系,

开展以学习为导向的评价实践,使校本英语测试更好地服务于大学英语教学,更有力地推动大学英语教学改革,同时也更好地发挥复旦校本英语测试的领航作用。

金 艳

上海交通大学外国语学院教授

全国大学英语四、六级考试委员会主任

2019年7月

# 前 言

复旦大学英语水平考试(Fudan English Test,简称FET)是复旦大学外文学院大学英语教学部语言测试中心自主研发的校本英语水平考试,是复旦大学多元化、多层次外语教学体系的重要组成部分。该项目自2010年实施以来,已经成功举行九次,参加FET的考生人数逾两万。相关研究结果显示,FET的实施不仅为测量大学生的综合英语应用能力提供了客观、有效的工具,而且也有力地推动了大学英语教学改革,对大学英语教学发挥了积极、正面的反拨效应。FET的开发与效度研究模式也为我国其他高校开展校本英语测试提供了重要参考,具有一定的示范效应。

FET的开发与实施是多个团队密切配合、共同努力的结果。大学英语教学部语言测试中心与复旦大学教务处在FET的开发与实施过程中发挥了关键性作用。其中,语言测试中心主要负责FET的命题、审题、拼题、试测、评分和数据分析等工作;复旦大学教务处主要负责组织考试的报名、考场安排、监考、成绩发布和查询等工作。语言测试中心内部又分为四个团队,分别为听力与写作团队、阅读团队、口语团队和数据分析团队。团队成员均为大学英语教学部的老师。各位老师承担繁重的教学与科研任务的同时,在FET的开发和研究中投入了大量的时间和精力。FET的每一套考题都经过命题、审题、磨题等步骤,确保试题质量优良。此外,大学英语教学部的

所有老师每年均参与 FET 的阅卷工作。FET 的考生人数每年约为 3 000-4 000 人,有些年份的考生人数超过 5 000。每份 FET 的写作试卷和口语音频均需要两位老师进行评阅,工作量之大可想而知!因此,FET 的顺利实施凝聚了多个团队和众多老师的心血,是集体智慧的结晶。

本书中报告的内容为语言测试中心近些年对 FET 开展效度研究所取得的一些阶段性成果。从 FET 项目启动开始,我们一直秉承考试开发与效度研究相辅相成的理念。一方面,考试开发与实施为效度研究提供大量的数据;另一方面,效度研究的相关结果能够为进一步完善考试质量提供重要依据。在 FET 的效度研究过程中,我们对目前学术界比较认可的效度理论框架进行详细的回顾和梳理,最终确定采用查普尔等人(Chapelle et al.,2008)提出的解释性效度论据框架(interpretative validity argument framework)对 FET 的效度开展研究。该理论框架基于近些年教育与心理测试领域效度理论研究的最新成果,其重要优势是能够将不同的效度研究证据进行组合形成效度论据,从而发现考试和效度研究中的不足之处,为进一步完善考试和效度研究工作提供重要的指导作用。该框架已经成功应用于托福网考的效度研究。

在该理论框架的指导下,本书共报告了 FET 的五项研究,分别对应解释性效度论据框架中的不同推理链条。本书中所报告的研究,有些为未出版的内部研究报告,有些已经在国内外的语言测试会议上(如 Language Testing Research Colloquium,简称 LTRC)作为会议论文进行宣读,有些成果已经公开发表在国际或国内的学术期刊上,如 *The Asia-Pacific Education Researcher*, *University of Sydney Papers in TESOL*,《外语教学理论与实践》《中国外语》《外语测试与教学》等。本书的主要特点可以概括为:1) 聚焦于校本英语测试的开发、实施与效度研究;2) 采用最新的效度理论框架指导效度研究工作;

3) 采用多种定量和定性研究方法对 FET 效度的不同方面展开研究。书中的研究成果对于将来进一步完善 FET 具有重要的启示作用。

在 FET 的开发与实施过程中,大学英语教学部语言测试中心得到国内外众多语言测试专家的指导。例如,上海交通大学的金艳教授、上海外国语大学的邹申教授等长期关注 FET 项目并为 FET 的开发与研究提供宝贵意见;国际著名语言测试专家、澳大利亚墨尔本大学的蒂姆·麦克纳马拉(Tim McNamara)教授和乌特·诺奇(Ute Knoch)副教授曾应邀担任语言测试中心的顾问,为 FET 的效度研究提供指导。在本书付梓之际,我们谨向各位关注和指导 FET 开发与研究的专家们表达我们深切的谢意;同时,我们也希望各位专家、学者和国内外的语言教学和测试同行继续支持、关注 FET 和语言测试中心的其他工作。在本书的写作过程中,复旦大学外文学院的张晓艺博士仔细审读了本书第一章到第五章的内容并提出了修改意见;中国海洋大学的陈颖副教授审阅了本书的最后两章并提出了修改意见;西安外国语大学的王丽副教授帮忙整理了本书的参考文献。我们在此对这三位老师表达诚挚的谢意。

范劲松(澳大利亚墨尔本大学语言测试研究中心)

季佩英(复旦大学外文学院大学英语教学部)

2019年6月

# 目 录

第一章 绪论 .....	1
1.1 研究背景 .....	1
1.2 FET 概况 .....	4
1.2.1 FET 的内容与形式 .....	4
1.2.2 FET 的考试开发:以阅读部分为例 .....	7
1.3 效度理论框架 .....	12
1.3.1 效度理论的发展 .....	12
1.3.2 解释性效度论据框架 .....	14
1.4 本书结构 .....	20
第二章 大学生英语能力需求分析 .....	22
2.1 研究设计 .....	22
2.1.1 被试 .....	22
2.1.2 研究工具 .....	24
2.1.3 数据收集 .....	25
2.1.4 数据分析 .....	25
2.2 研究结果 .....	25
2.2.1 问卷信度统计 .....	25
2.2.2 听力 .....	26
2.2.3 阅读 .....	32

2.2.4	写作	37
2.2.5	口语	42
2.3	讨论	47
2.3.1	听力	48
2.3.2	阅读	49
2.3.3	写作	51
2.3.4	口语	52
2.4	本章小结	55
<b>第三章</b>	<b>FET 的内部因子结构探析</b>	<b>57</b>
3.1	文献回顾	59
3.1.1	语言考试的因子结构分析	59
3.1.2	多样本 CFA	61
3.2	研究设计	62
3.2.1	研究样本	62
3.2.2	数据收集	62
3.2.3	模型设定	63
3.2.4	数据分析	65
3.3	研究结果	69
3.3.1	正态检验	69
3.3.2	模型评估	70
3.3.3	因子结构的不变性分析	74
3.4	讨论	76
3.5	本章小结	79
<b>第四章</b>	<b>使用学生自评探讨 FET 的效度</b>	<b>80</b>
4.1	文献回顾	81

4.1.1	语言研究中的自我评估 .....	81
4.1.2	自我评估在考试效度研究中的应用 .....	83
4.1.3	拉施模型在自评量表效度研究中的应用 .....	84
4.2	研究设计 .....	86
4.2.1	研究问题 .....	86
4.2.2	自评量表的开发 .....	86
4.2.3	被试 .....	87
4.2.4	数据收集 .....	87
4.2.5	数据分析 .....	88
4.3	研究结果 .....	92
4.3.1	变量图 .....	92
4.3.2	信度与分隔系数 .....	93
4.3.3	描述语分析 .....	94
4.3.4	量表类别分析 .....	96
4.3.5	维度分析 .....	97
4.3.6	DIF 分析 .....	99
4.3.7	考试和自评量表的内部因子结构分析 .....	100
4.3.8	结构回归分析 .....	102
4.4	讨论 .....	104
4.5	本章小结 .....	107
<b>第五章</b>	<b>学生对 FET 的态度研究 .....</b>	<b>109</b>
5.1	文献回顾 .....	110
5.1.1	态度的定义 .....	110
5.1.2	关于态度的实证研究 .....	112
5.2	研究设计 .....	114
5.2.1	研究问题 .....	114

5.2.2	被试	114
5.2.3	研究工具	114
5.2.4	数据收集	116
5.2.5	数据分析	117
5.3	研究结果	118
5.3.1	因子分析	118
5.3.2	学生对 FET 的态度	120
5.3.3	学生态度与个人特征之间的关系	124
5.3.4	学生态度与考试成绩之间的关系	126
5.4	讨论	128
5.5	本章小结	131
<b>第六章</b>	<b>FET 对学生英语学习的反拨效应研究</b>	<b>133</b>
6.1	文献回顾	134
6.1.1	语言测试的反拨效应	134
6.1.2	反拨效应实证研究	137
6.2	研究设计	140
6.2.1	设计方案	140
6.2.2	研究问题	140
6.2.3	被试	141
6.2.4	研究工具	141
6.2.5	数据收集	142
6.2.6	数据分析	143
6.3	研究结果	144
6.3.1	因子分析	144
6.3.2	学生对 FET 的看法	146
6.3.3	性别和英语水平对学生看法的影响	148

6.3.4	FET 对学生英语学习的影响 .....	150
6.3.5	性别、英语水平和 FET 的反拨效应 .....	152
6.4	讨论 .....	154
6.5	本章小结 .....	158
<b>第七章</b>	<b>结语</b> .....	<b>159</b>
7.1	主要发现 .....	160
7.2	FET 效度论据构建 .....	163
7.3	研究启示 .....	166
7.4	本章小结 .....	168
<b>参考书目</b>	.....	<b>170</b>
<b>附录 1</b>	<b>学生英语能力需求调查问卷</b> .....	<b>186</b>
<b>附录 2</b>	<b>考生对 FET 态度调查问卷</b> .....	<b>192</b>

# 第一章 绪 论

## 1.1 研究背景

在我国,全国大学英语四、六级考试(College English Test,简称 CET)是测量非英语专业大学生英语学习和水平的重要工具,其信度、效度已经得到社会的广泛认可(如 Jin, 2010; 杨惠中、Weir, 1998; 金艳、程李颖, 2013),对我国的大学英语教学发挥了重要的促进作用,产生了积极的社会效应(如辜向东, 2007; 辜向东、彭莹莹, 2010; 金艳, 2006)。但是由于我国高校数量众多,且分布在全国各地,学校的发展存在一定的不平衡性,因此 CET 作为全国性的英语能力考试,难以满足不同等级的院校和学生对英语考试的需求。此外, CET 考试规模庞大,目前每年的参考人数约为 1 800 万,是世界上规模最大的单科考试。如此庞大的规模在一定程度上制约了考试的设计(Jin, 2011, 2014)。例如,由于考生人数众多, CET 很难采用大量需要人工评阅的主观性考题,而且也很难在全部考生中推行口语考试。鉴于 CET 目前存在的一些不足之处,有些高校在使用 CET 的同时,也开始自主研发校本英语水平考试,作为 CET 有益和必要的补充(如清华大学测试小组, 2012; 中国人民大学 TOPE 项目组, 2013)。

复旦大学自 2010 年起启动复旦大学英语水平考试(Fudan English Test,简称 FET)项目,并于 2011 年 12 月开始在校内实施 FET。截至 2017 年 12 月, FET 已经成功举行七次,考生人数共计逾

两万。FET 的考试对象为复旦大学非英语专业的本科生。考试每年 12 月举行一次。根据复旦大学教务处制定的相关政策, FET 相当于一门英语必修课, 占两个学分, 学生在 FET 中所取得的等级将被转为相应的绩点成绩。同时, 学生在四年本科学习中至少需参加并通过一次 FET。学生如果对 FET 的成绩不满意, 可以重复参加考试, 成绩以最后一次为准。学生需要和教务处确定成绩, 且成绩一旦确定, 就不能再次参加考试。如果学生在毕业之前无法通过 FET, 那么他们将无法按时获得复旦大学的学士学位。学生的绩点对于读研、留学或就业等均会产生一定的影响, 而是否能够如期获得复旦大学的学士学位对学生的影响更甚。因此 FET 是一项高风险的语言考试 (high-stakes test, 如 Bachman & Palmer, 1996)。考试的高风险性对于考试本身的质量提出了更高的要求。

FET 的开发和实施是一项系统的团队工程。图 1.1 展示了 FET 的工作机制流程。该图显示, FET 的开发以《复旦大学人才培养目标》中对学生英语能力的要求为指导, 并参照国际应用语言学界广泛认可的语言能力理论模型 (如 Bachman, 1990; Bachman & Palmer, 1996; Canale & Swain, 1981; Chalhoub-Deville, 1997)。考试包括听、说、读、写四大组成部分, 每部分各占总成绩的 25%。在 FET 开发和实施的过程中, 复旦大学教务处和大学英语教学部相互配合, 共同确保考试的质量与顺利实施。如图所示, 复旦大学教务处主要负责制定相关的考试政策、组织考试的报名、为考试的顺利实施提供设备方面的保障等; 同时, 教务处也负责考场的安排、考试的实施和考试成绩的发布等方面的工作。大学英语教学部主要负责考试的命题、审题、拼题、印刷、考务等工作; 同时, 也负责监考人员的培训、配合教务处完成考试的实施、组织教师阅卷、对考试的数据进行分析并且组织考试的相关效度研究工作等。图 1.1 同时显示, FET 与学生的课程学习之间密切相关。一方面, FET 的实施为测量学生的课程学习效

果提供了测试手段,考试也能对学生的课程学习产生积极、正面的反拨作用;另一方面,每次 FET 的数据分析结果可以为教务管理部门制定相关的教育政策提供重要参考,而且也可以用于大学英语的课程设置和规划(如季佩英等,2016)。FET 兼具水平测试与成就测试的特征。一方面,FET 为学生的英语水平提供科学、有效的测量;另一方面,FET 的考试内容和形式有效体现了《复旦大学人才培养目标》和复旦大学英语教学大纲对学生英语能力的要求,能够检验学生的英语学习成果。

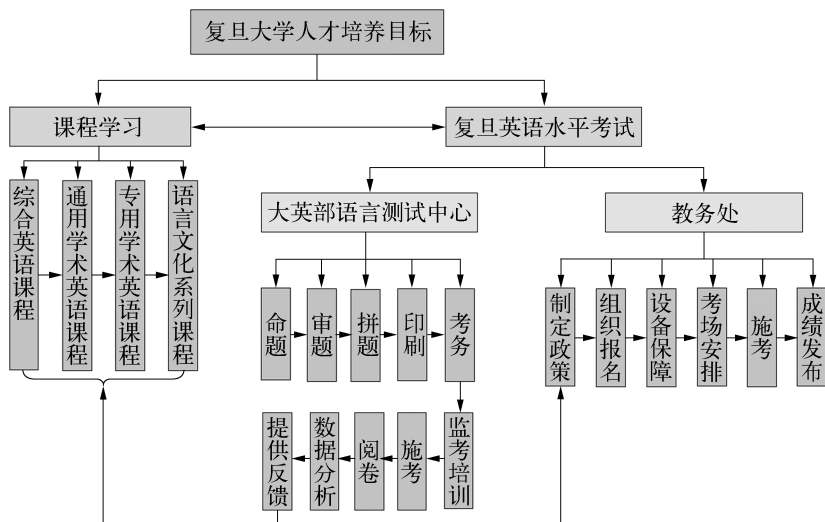


图 1.1 FET 工作机制流程图

在语言测试研究领域,研究人员普遍认为考试的开发和效度研究应相辅相成,考试开发的过程同时也是考试效度研究的过程(如 Alderson et al.,1995; Bachman & Palmer,1996)。韦尔(Weir,2005)认为,考试效度研究应包括前期(*a priori*)和后期(*a posteriori*)两个组成部分。其中,前期效度研究主要体现在考试的开发过程,而后期效度

研究主要体现在考试正式实施以后。在语言测试研究领域具有重要影响的《教育与心理测试标准》明确指出,考试的效度研究是一个连续不断的过程,贯穿于考试开发、实施和使用(AERA et al., 2014)。在 FET 的开发和实施过程中,我们采用了这一观点。从 FET 开始酝酿到考试的开发与正式实施,我们始终强调效度研究在确保考试质量方面的重要作用。在本书中,我们采用目前语言测评领域最新的效度研究框架——解释性效度论据框架(interpretive validity argument framework)(如 Bachman & Palmer, 2010; Chapelle et al., 2008; Kane, 1992, 2013)——探讨 FET 的效度研究。我们将选取近些年 FET 的一些代表性效度研究成果来探讨 FET 效度论据的构建以及未来 FET 的发展方向,以期对我国的其他语言考试,尤其是校本英语考试的开发与效度研究提供参考和启示。

## 1.2 FET 概况

### 1.2.1 FET 的内容与形式

FET 的内容和形式请参见表 1.1。FET 采用目前国际应用语言学界关于语言能力普遍认可的观点,即语言能力既包括整体语言能力,也包括该能力之下的分项能力(如 In'nami & Koizumi, 2011; In'nami et al., 2016; Sawaki et al., 2009)。一般认为,分项能力包括听、说、读、写能力。国际上著名的语言水平考试如雅思、托福等均采用该观点。FET 包括听、说、读、写四大组成部分,每部分各占总成绩的 25%。表 1.1 显示, FET 的笔试部分包括听力、写作和阅读三个部分。听力部分包括复合式听写、听对话后完成选择题和听学术讲座后完成选择题三个部分。写作包括两个任务,即议论文写作(essay writing)和应用文写作(practical writing)。FET 的作文评分目前采用