外教社
博学文库

英语录音口试评分过程研究

A Study of the Process of Assessing
Tape-Mediated EFL Speaking Test Performance

王海贞 著

上海外语教育出版社
外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS
www.sflep.com

外教社
博学文库

英语录音口试评分过程研究

A Study of the Process of Assessing
Tape-Mediated EFL Speaking Test Performance

王海贞 著

上海外语教育出版社
外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS

图书在版编目（CIP）数据

英语录音口试评分过程研究/王海贞著. —上海：上海外语教育出版社，2009
（外教社博学文库）
ISBN 978-7-5446-1390-3

Ⅰ.英… Ⅱ.王… Ⅲ.英语-口试-评分-研究
Ⅳ.H319.9

中国版本图书馆CIP数据核字（2009）第074218号

出版发行：上海外语教育出版社
（上海外国语大学内） 邮编：200083
电　　话：021-65425300（总机）
电子邮箱：bookinfo@sflep.com.cn
网　　址：http://www.sflep.com.cn http://www.sflep.com
责任编辑：张亚东

印　　刷：上海叶大印务发展有限公司
经　　销：新华书店上海发行所
开　　本：890×1240 1/32 印张 8.625 字数 257 千字
版　　次：2009 年 12 月第 1 版 2009 年 12 月第 1 次印刷
印　　数：2 100 册

书　　号：ISBN 978-7-5446-1390-3 / G · 0559
定　　价：24.00 元

博学文库
编委会成员

（按姓氏笔划为序）

出版说明

上海外语教育出版社始终坚持“服务外语教育、传播先进文化、推广学术成果、促进人才培养”的经营理念，凭借自身的专业优势和创新精神，多年来已推出各类学术图书600余种，为中国的外语教学和研究做出了积极的贡献。

为展示学术研究的最新动态和成果，并为广大优秀的博士人才提供广阔的学术交流的平台，上海外语教育出版社隆重推出“外教社博学文库”。该文库遴选国内的优秀博士论文，遵循严格的“专家推荐、匿名评审、好中选优”的筛选流程，内容涵盖语言学、文学、翻译和教学法研究等各个领域。该文库为开放系列，理论创新性强、材料科学翔实、论述周密严谨、文字简洁流畅，其问世必将为国内外广大读者在相关的外语学习和研究领域提供又一宝贵的学术资源。

上海外语教育出版社

前言

本书研究英语录音口试评分员构建评分标准以及做出评分决策的过程，旨在探讨评分员的作用以及评分过程的本质。

评分员的作用和评分过程的本质是评分理论的两个核心问题。长期以来外语测试界未对这两个问题进行过公开争辩，但现有文献直接或间接反映出人们的不同观点。早期研究关注评分标准的制定和验证，以及评分员因素与评分成绩的相关性，将评分员看作运用评分标准的工具，将评分过程看作按部就班的线性过程。近年来有些研究采用有声思维的方法探索英语写作的评分过程，发现评分员不但主动构建文本意义而且做出评分决策，整个评分决策过程呈现动态循环的特征。然而，这些发现分散在不同研究中，缺乏完整性和系统性，对评分员的作用和评分过程的本质这两个问题尚不能形成定论。再者，现有实证研究集中在英语作文评分过程领域，缺乏对英语录音口试评分过程的探讨，缺乏对评分员评估学习者英语口语能力时的决策行为研究。

针对这一研究空白，本研究采用评分员口头报告的研究方法追踪评分员评估英语录音口试样本时的思维过程，着重研究评分员的评分行为和评分过程，旨在揭示英语录音口试评分过程的本质和评分员的作用。本研究具体围绕两大研究问题：（1）评分员如何构建评分标准？（2）评分员如何作出评分决策？

本研究的研究对象为 24 名大学英语专业教师，分别来自全国 11 所

高校，均具有硕士或硕士以上学历，专业英语教学经验 1—21 年不等，其中 11 名教师是评分新手，另有 5 名教师具有 3 年以上的全国英语专业四级口试评分经验。本研究选用 2005 年全国英语专业四级口试五盒录音样本，研究对象经培训后，对照分析性评分量表和具体评分细则进行评分。评分任务为即席讲话，评分项目包括讲话内容、语音语调和语法与词汇三个分项。在听取录音样本评分的同时，研究对象采用即时回溯、受激回忆等方法口头汇报评分时的思维活动，并在评分全部结束后接受访谈。本研究对研究对象的口头汇报和访谈全程录音，每位研究对象的录音长度为 44—86 分钟不等。完成全部语料收集和文字转写后，本研究对语料进行了精细的定性分析，旨在通过翔实的第一手资料描述录音口试评分员的评分过程和评分行为，分析评分行为的差异性和规律性，并探讨产生这些差异性和规律性的原因。

本研究结果归纳如下：

首先，评分员对评分标准的阐释和应用存在一致性和差异性。一方面，几乎所有评分员都考虑了评分量表里的标准（即内容切题、丰富、有条理；语言流利、准确、自然）；只有六名评分员对词汇量未加判断，因为他们认为不能依据该标准判断考生成绩或认定该标准不好操作。另一方面，所有评分员都评判了故事的完整性，大约 50%—60%的评分员考虑了故事的新颖性、句法复杂性和自我修正，这些标准在评分量表中没有规定。另外，不同的评分员对相同的评分标准理解不一致，例如在评判流利性时，除了评分量表里明确规定的停顿特征以外，评分员还考虑了其他 12 种不同特征，如持续说话的能力、语速、犹豫和重复等。再者，评分员在评判时参照考试标准和个人评判标准不断构建和再构建评分标准，当考生的语言特征与等级描述语不相匹配或当考试标准难以操作时，评分员倾向于使用个人评判标准。

其次，评分员在决定分数时不仅考虑了三个评分项目（即内容、语音语调和语法词汇）的不同侧重点而且关注了每个评分项目的不同侧重点。一方面，在内容、语音语调和语法词汇三个评分项目中，16 名评分员（67%）认为内容最重要，3 名评分员侧重语音语调，另 5 人不考虑项目间的侧重。另一方面，大多数评分员（> 83%）在评判语法和语音语调时倾

向于找错误，在评判词汇时倾向于找“亮点”加分，在评判内容时倾向于综合考虑是否切题、合乎逻辑，故事是否完整，并对内容丰富、新颖或生动的叙述加分。再者，在评判三个项目时评分员采用循环评分方法，对一个分项形成判断后转而重点关注下一个项目并不断循环；在评判每个项目时评分员先形成分数假设，再找证据不断验证或修订假设，呈现假设—验证/修订的循环过程。

最后，在以上研究的基础上，本研究尝试性地提出一个录音口试评分过程理论模型。该模型不仅描绘了评分员循环构建评分标准和分数决策的动态过程，而且明确了评分员在评分中的中心作用。

本研究成果既具有较强的理论价值，又具有方法意义和实践意义。首先，本研究首次明确提出评分员在评分过程中的中心作用，提出以评分员为主体的动态评估录音口试的理论模型，丰富和发展了现有的语言测试理论。其次，本研究对即时回溯、受激回忆、访谈等多种质化研究方法的有机结合进行了有益尝试，拓展了二语口语研究以及语言测试研究的研究方法和研究手段。另外，了解英语录音口试评分的实际过程以及评分员的作用，对选拔评分员，改进评分员培训，提高全国英语专业四、八级口试的评分质量具有重要指导意义。

此外，必须指出本研究的样本较小，属探索性研究，目的是为英语录音口试评分过程研究提供一个新的理论假设。未来的研究可以采用更大的样本，进一步验证或修订这一理论模型；也可以进一步探讨该模型是否广泛适用于大规模语言行为测试的评估过程。

本书是作者的博士论文，于2007年5月在南京大学完成。论文从选题到设计，从初稿到定稿，自始至终都得到了南京大学外国语学院多位教授和多位学长的帮助、同学的鼓励以及全国英语专业四、八级口试中心的支持，在此表示衷心的感谢。

首先感谢我的导师文秋芳教授。她不仅引领我进入二语习得研究和语言测试研究的学术殿堂，而且以她渊博的专业知识、睿智的思维和严谨的学术态度不断地引导我、鼓励我。她对学术研究的热情和孜孜以求的钻研精神使我受益终生。

感谢南京大学丁言仁教授。他语言功底深厚，为人率真质朴，热心慈善，是一位值得我敬佩的教授。同样感谢南京大学王海啸教授、陈新仁教授、王文宇博士、周丹丹博士以及南京师范大学马广惠教授和对外经济贸易大学王立非教授，他们耐心地对我论文的初稿和修改稿提供了宝贵的、具有建设性的意见。

我还要向那些以不同方式向我提供帮助的同学、同事和朋友们致以诚挚的谢意。感谢与我朝夕相处的南京大学应用语言学专业的博士和博士生同学们，他们无私地与我分享学术信息和研究经验，并在我陷入苦恼和困惑时提供强有力的精神支持和鼓励；特别感谢王宇博士和周卫京博士，她们分享和见证了我研究的全过程，并始终如一地支持和鼓励我；感谢英国兰开斯特大学肖忠华博士，他无偿地为我寄送了宝贵的研究资料；感谢南京大学Don Snow教授和苏州大学外籍教师Jennifer Littlejoin女士，他们分别为本书的初稿和定稿进行了语言校对和润色；感谢全国英语专业四、八级口试中心和南京大学英语口语研究所的所有老师，他们不仅提供了相关课题的研究经费，而且在资料和设备等方面提供了便利；感谢参加本研究的24名高校英语专业教师，他们有的是我的同事或朋友，有的则仅是一面之交，但都热心、积极地配合我完成数据收集；感谢苏州大学外国语学院给我提供了全脱产的学习机会，使我全身心地投入本课题研究。

王海贞

2008年12月于苏州

Contents

List of Abbreviations

ACTFL	American Council on the Teaching of Foreign Languages
BEC	Business English Certificates
CET-SET	College English Test — Spoken English Test
EFL	English as a Foreign Language
FCE	Cambridge First Certificate in English Speaking Test
FSI	Foreign Service Institute
IELTS	International English Language Testing System
L1	Primary Language(s)
L2	Second or Non-Primary Language(s)
NNS	Non-Native English Speakers
NS	Native English Speakers
OPI	Oral Proficiency Interview
PETS	Public English Tests
SOPI	Simulated Oral Proficiency Interview
TEM4-Oral	Graded Test for English Majors — Band Four Oral Test
TEM8-Oral	Graded Test for English Majors — Band Eight Oral Test
TOEFL	Teaching of English as a Foreign Language

List of Tables and Figures

INTRODUCTION

This book reports an exploratory study which investigated the process of assessing the Graded Test for English Majors — Band Four Oral Test, a nationwide tape-mediated English speaking test for College English majors in China. The verbal reports and interview data of 24 raters were analyzed, with a view to better understanding how the raters assessed tape-mediated EFL speaking performance. Specifically speaking, this study investigated how the raters applied the rating criteria to tape-mediated performances and how they made their judgments and decisions in the scoring process. Being qualitative and exploratory in nature, this study seeks to generate hypotheses and stimulate further research in this field.

0.1 Motivations

This study was motivated by the present researcher's personal interest in language assessment, especially in the assessment of the Graded Test for English Majors — Band Four Oral Test (hereinafter TEM4-Oral). Specifically, this study was motivated by (1) queries about different ways of applying