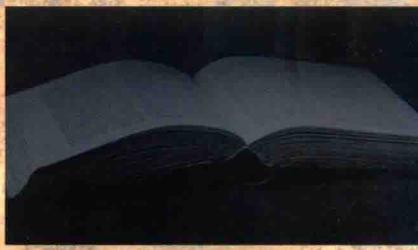
外教社
博学文库

大规模英语考试作文评分信度与网上阅卷实证研究

Marking Reliability and Online Marking of Compositions Produced on a Large-scale English Examination
—An Empirical Study



王跃武 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

大规模英语考试作文评分信度与网上阅卷实证研究

Marking Reliability and Online Marking of Compositions Produced on a Large-scale English Examination
—An Empirical Study

王跃武 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

大规模英语考试作文评分信度与网上联机阅卷实证研究/王跃武著.
—上海：上海外语教育出版社，2015
（外教社博学文库）
ISBN 978-7-5446-3984-2

Ⅰ.①大… Ⅱ.①王… Ⅲ.①计算机系统-评分-应用-英语-写作-考试-研究
Ⅳ.①H315-39

中国版本图书馆CIP数据核字（2015）第083783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：蒋浚浚

印　　刷：上海信老印刷厂
开　　本：890×1240　1/32　印张 12.25　字数 354 千字
版　　次：2015 年 7 月第 1 版　2015 年 7 月第 1 次印刷

书　　号：ISBN 978-7-5446-3984-2 / G • 1258
定　　价：37.00 元

博学文库
编委会成员

（按姓氏笔画为序）

姓　名	学　校
王守仁	南京大学
王腊宝	苏州大学
王　蔷	北京师范大学
文秋芳	北京外国语大学
石　坚	四川大学
冯庆华	上海外国语大学
吕　俊	南京师范大学
庄智象	上海外国语大学
刘世生	清华大学
杨惠中	上海交通大学
何刚强	复旦大学
何兆熊	上海外国语大学
何莲珍	浙江大学
张绍杰	东北师范大学
陈建平	广东外语外贸大学
胡文仲	北京外国语大学
秦秀白	华南理工大学
贾玉新	哈尔滨工业大学
黄国文	中山大学
黄源深	上海对外贸易学院
程朝翔	北京大学
虞建华	上海外国语大学
潘文国	华东师范大学
戴炜栋	上海外国语大学

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作出了积极的贡献。

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

谨以此书献给我敬爱的母亲

序言

我国是教育考试大国，数量众多的大规模考试动辄涉及数十万甚至上百万考生，而且考试结果往往决定考生一生命运，成为真正意义上的高风险考试。由于考生数量庞大，目前已有一些大规模考试采用网上阅卷，以求提高阅卷效率，但是在阅卷信度控制方面还有不少尚待改进的空间，因此本博士论文对大规模考试的网上阅卷信度控制进行深入的理论探讨与实证研究，其研究成果在当今仍具有迫切的现实意义。

网上作文阅卷的评分信度，主要涉及语言测试理论与实践。

对于大规模考试来说，尤其是跨地区考试中的作文阅卷，传统的会议阅卷（网下阅卷）方式，很难实现严格意义上的试卷随机分发、对阅卷员的监控和对阅卷过程的质量控制也难以做到与阅卷同步进行。对于网下阅卷中存在的不足，作者认为实施网上阅卷可以得到改善。实施网上阅卷的关键除了技术问题外，主要是提高网上作文阅卷的评分效度和信度，提高实时质量控制的可操作性和效率。

作者从研究作文阅卷的心理过程入手，对网上阅卷过程的各个环节进行了探索，提出了网上阅卷的模型，开发了相应的计算机网上阅卷实验系统，并开展了实证研究。实证研究内容涉及与作文网上联机阅卷有关的阅卷员心理决策过程和作文评分过程中的认知作用问题。实证研究发现，由于网上阅卷可以做到确保作文卷分配的充分随机性、可以加强阅卷过程的质量监控。网上阅卷方式，在其他条件相同的情况下，

作文评分信度、阅卷员之间的评分一致性都高于网下阅卷，阅卷员内部一致性也很高。实证研究结果证明网上阅卷有利于提高阅卷信度和阅卷效率。

但是，作者正确地指出，网上阅卷与常规的会议阅卷（网下阅卷）比较，只是阅卷方式的变化，网上阅卷本身并不能提高阅卷信度。保证网上作文阅卷的信度还是要依靠阅卷过程各个环节的质量控制，包括制定严格的作文题评分原则及具体的评分标准、准确选择评分样卷、严格训练评分员、阅卷过程中作文卷的随机分配、滤除作文评分过程中的随机误差等，只有这些措施得到严格实施，才有可能确保大规模考试中作文评分的质量，保证考试的信度和效度，对于考试结果可能对学生未来产生重大影响的大规模考试来说这一点尤为重要。

本项研究提出的基本思路和方法可供大规模考试使用，对我国语言测试具有实践意义。

是为序。

杨惠中

2013年2月10日

前言

大规模语言考试的作文评分信度是保证整个考试的效度与公平公正的重要条件。本书基于十年前完成的博士论文。研究选取大学英语四、六级考试（以下简称CET）作文阅卷作为个案，对其评分信度特别是网上阅卷信度进行实证研究。

在本研究完成前，CET作文一直在多个阅卷点以集中会议（纸笔）方式进行批阅。CET考试委员会和各阅卷点制定并贯彻了一整套有力措施来保证作文评分的信度。这些措施包括：制订严格的作文评分原则及具体而明确的评分标准，准确选择参照卷并用其来校准阅卷员对评分标准的掌握尺度，对阅卷员进行严格的阅前培训和阅后考核，在阅卷过程中由阅卷组长随机抽查阅卷员的评分质量，要求阅卷员严格掌握评分标准并且在整个阅卷过程中保持一致，要求每名阅卷员所评的作文分与对应考生客观题得分的相关系数必须达到一定标准，保证在阅卷过程中作文卷的随机分配、阅卷结束后的作文分计算机调整，以及建立稳定、合格、能胜任的阅卷员队伍。以上措施确保了CET作文评分的质量和阅卷的高信度（包括阅卷员本人的一致性、阅卷员之间的一致性、阅卷点之间的一致性），取得了良好的效果（杨惠中、Weir, 1998：36-41，126-150）。

上述措施具有较强的实用性，但会议阅卷存在下列不足：（1）试卷的随机分发只能在各阅卷点内部实现，难以在全国范围内实现。（2）对阅卷员的监督、抽查工作还有待完善。（3）对阅卷的监控和阅卷过程的质量

控制也难以做到与阅卷同步进行。而网上阅卷方式的优点在于，试卷可随机分发，阅卷员们可以在集中的阅卷点，也可以在办公室甚至家中，在电脑屏幕上通过阅读考生作文（电子文本或影像文本）而完成评分，阅卷负责人可通过网络进行质量控制，对分散在各处的阅卷员进行实时监控，提高阅卷效率。

本研究的目的有三：一是了解CET作文阅卷情况及其存在的问题；二是设计开发CET作文网上阅卷系统软件（CET Online Marking System，以下简称OMS）；三是在局域网上试用OMS，并检验阅卷员使用OMS进行网上联机阅卷的评分信度。

研究分三阶段进行。第一阶段在充分了解作文考试，特别是CET作文考试的基础上，进行了两项带有调查性质的实验，其内容涉及与CET作文网上阅卷有关的阅卷员心理决策过程和作文评分过程中的认知作用问题。第一个实验于2000年进行，从3个阅卷点随机抽取的24名阅卷员批改了2,380份CET-4作文卷（其中60份为共同卷），这一实验包含了一项内省法出声思维研究、一项反省研究以及与阅卷和心理认知有关的问卷调查。第二个实验于2001年进行，从3个阅卷点随机抽取16名阅卷员批改了120份作文卷，CET-4和CET-6各60份。通过对这两个实验产生的数据的定性和定量分析，确定了阅卷员在网上批阅作文时所遇到的问题和困难，并由此构建了一个CET作文网上阅卷的初步模式。此模式构成了设计OMS的基础。

第二阶段在总结前述工作和广泛了解国内外有关作文网上阅卷研究（包括自动阅卷和人工阅卷方面的研究）的基础上，设计并开发了OMS。

第三阶段主要是在局域网上试用OMS，并调查阅卷员使用OMS进行实际网上阅卷的评分信度。2002年进行了本研究的第三个实验。从上海阅卷点240名会议阅卷员中随机抽取14名，随机分为两组，在同一局域网上应用OMS批阅1,341份作文（726份CET-4作文，615份CET-6作文）。在正规会议阅卷时这些阅卷员则分散在8个房间。所有14名阅卷员都在网上阅卷和正规会议阅卷两种方式下独立批阅了20份共同卷，以比较阅卷员在两种不同方式下的评分信度差异。实验过程中和结束

后，对阅卷员进行了个别和群体访谈，以发现和解决问题，并调查他们对CET作文网上阅卷方式的评价。

本研究分析第三个实验中20份共同卷的评分信度采用两种分析软件。一是SPSS统计分析软件；二是近年来受到教育测量界所重视的多层面Rasch模式统计分析软件FACETS。除此之外的其他数据则只用SPSS软件分析。

使用SPSS分析上述20份共同卷的评分信度所采用的模型为：从评分一致性、正确性、作文分与对应考生客观题得分之间的相关系数三方面来分析和比较阅卷员在网上阅卷方式和传统会议阅卷方式下评出的考生作文分数据。

多层面Rasch模式分析方法应用FACETS软件，采用第三个实验中20份共同卷所产生的数据，以两种方式来分析阅卷员的行为。第一种方式从阅卷员严厉度（rater severity）和阅卷员一致性（rater consistency）两个方面来分析和比较阅卷员在网上阅卷方式和传统会议阅卷方式下的评分差异。第二种方式运用FACETS软件所提供的偏差分析（bias analysis）功能来分析和比较在两种阅卷方式下阅卷员和作文卷之间，以及阅卷员和阅卷方式之间的交互作用情况。

在本研究不同阶段所进行的三个实验的数据显示，阅卷员在网上阅卷方式下评出的作文分的信度都高于在传统会议阅卷方式下评出的分数的信度。尤其是第三个实验证明，在网上阅卷方式下阅卷员之间的评分一致性要高于他们在传统会议阅卷方式下的评分一致性（阅卷员间的评分相关系数值分别为.84和.77，alpha值分别为.98和.90）。另外，网上阅卷方式下的阅卷员内部一致性也很高（阅卷员本人在前后不同时段的评分相关系数值为.87，alpha值为.92）。阅卷员在两种阅卷方式下评出的四份控制卷数据经与专家分比较，发现在网上阅卷方式下阅卷员的评分正确性要高于他们在传统会议阅卷方式下的评分正确性。在网上阅卷方式下评出的作文分与对应考生客观题得分之间的相关系数（.67）也高于在传统会议阅卷方式下评出的作文分与对应考生客观题得分之间的相关系数（.58）。此外，作文分均值差异分析证明网上阅卷能提高作文的评分信度。用FACETS分析的结果与以上结论一致。

本研究对引起阅卷员在两种不同阅卷方式下评分信度差异的可能原因进行了调查，发现OMS对网上阅卷过程和阅卷员所采用的各项质量控制和实时监控措施是产生这种差异的重要原因。

通过本研究产生了一个在使用OMS环境下的CET作文网上阅卷模型。该模型具有四大显著特点：① 随机分发试卷；② 实时监控阅卷员；③ 控制阅卷质量；④ 高效记录分数和数据。这些特点相应构成了CET作文网上阅卷相对于传统会议式阅卷所具有的四大优势。

在本研究结果的基础上，CET考委会于2003年、2004年进行了两次各10万考生规模的CET作文网上阅卷实验，并于2005年6月考试结束后进行了20万考生规模的正式网上作文阅卷。从2007年起，全国范围的CET作文网上阅卷正式实施。

必须指出，从常规会议阅卷到网上阅卷，只是阅卷方式的改变，这一改变本身并不能提高阅卷信度。要保证网上作文阅卷的信度还须加强阅卷过程中各环节的质量控制，只有这些控制措施得到严格实施，才能确保大规模考试中作文评分的质量，保证考试的信度和效度。这一点对于考试结果可能对考生的未来产生重大影响的大规模考试来说尤为重要。

在以英语为外语（EFL）的作文考试的研究框架内，作文网上阅卷还有进一步研究的必要，这样的研究必将为作文评估研究做出理论和实践上的贡献。

致谢

本书源自我的博士论文。首先衷心感谢在我博士论文写作期间给予我指导、帮助和支持的师长、同事、朋友和家人。

杨惠中教授是我终身感激不尽的恩师。杨先生在我攻读博士学位期间不仅给予我学术和科研上的指导，还在经济和生活方面给我无私的帮助和支持。没有杨先生的督导与鼓励，就没有我的博士论文，也就没有今天这本著作。

感谢全国大学英语四、六级考试委员会提供研究经费，上海交通大学语言文字工程研究所提供研究场所和设备，并特别感谢金艳、刘鸿章、卫乃兴、朱正才、谢善路、巴源、杨浩然等教授的大力支持。

复旦大学已故教授董亚芬先生、上海海事大学已故教授温致义先生、上海外国语大学邹申教授曾认真细致地审阅了我的博士论文，并提出了许多高屋建瓴的批评意见和细致入微的修改建议。他们深厚的英文功底和严谨的治学精神都使我由衷地敬佩。

上海海事大学是我的母校，在我读博期间，母校的领导和同事给了我多方面的支持与鼓励，使我得以顺利完成学业。感谢金永兴书记、外国语学院张志军书记以及左彪、王大伟与夫人、王菊泉、吴慧等教授。

感谢参与本研究的所有教师，他们来自于全国30所高校，工作繁忙，但都十分认真地完成了所参与的工作，并提出了大量有益的建议，对于作文教学与评估极富启示意义。

当然还要感谢我的妻子和女儿。在我攻博期间，她们承受了太多的忧虑和困难，但还是一如既往地支持我，给我爱的支柱和力量。

此外，特别感谢匿名评审专家对本书初稿提出宝贵修改意见，感谢我现在的工作单位上海杉达学院外语学院美籍专家Mary Riordan和Michael J. Riordan极为细致地审读本书二稿并提出许多可贵的改进建议。文中不当之处由作者全权负责。

最后，特别感谢上海外语教育出版社的大力支持和各位编辑的辛勤劳动，使本书得以出版。

Contents