

《计算机技术条件下的语言测试》述评

刘永兵

(东北师范大学 外国语学院, 长春 130024, 吉林)

Carol Chapelle & Dan Douglas. 2010. *Assessing Language through Computer Technology*. 北京: 外语教学与研究出版社/剑桥大学出版社. F21 +136 pp.

本书是外语教学与研究出版社精选并引进的剑桥大学出版社设计的“剑桥英语教师丛书”之一, 由国际知名语言测试专家 Carol A. Chapelle 和 Dan Douglas 合作撰写。他们在二语测试、二语习得和应用语言学领域研究著述颇丰, 并在世界各地讲学, 在二语课堂教学评价和教师培训领域经验丰富。本书是为二语教师、二语测试及其相关专业人士撰写的一本有关计算机技术辅助语言测试指导性著作, 从理论和实践的双重角度较系统地讨论了计算机辅助语言测试的软件应用、计算机辅助语言测试的评估以及计算机辅助语言测试的有效性等问题。以下对本书的主要内容作一简评。

1. 内容简介

本书共六章, 主要对最新技术发展、计算机辅助语言测试与传统测试的差异、测试类型、测试评估等因素进行了探讨, 同时对计算机辅助语言测试的优越性和局限性进行了深入细致的分析。在第一章中, 作者首先指出世界上越来越多的第二语言学习者均要参加计算机辅助语言测试, 这已是不争的事实, 所以二语教师、二语测试设计者、二语测试研究者有必要掌握计算机辅助语言测试的专门知识, 了解计算机与信息技术的最新发展及其对二语测试的影响。作者接着对二语教师应该具备的三个主要方面知识和技能做了简要的概述, 如测试选择 (selection of tests)、课堂评价 (classroom assessment) 和试题开发 (test development)。作者在概述中指出, 二语教师至少要学会在各种各样的计算机辅助语言测试中做出最恰当的测试选择, 根据测试对象利用相关软件创建自己课堂教学测试题目或设计课堂小测验中的个别问题, 以了解学生学习状况并能快速得到详细的信息反馈。最后作者介绍了目前最新的测试方法: 如计算机自适应测试 (computer-adaptive testing)、多媒体测试任务 (multimedia tasks)、自动反应分析 (automatic response analysis)。他们指出这些新方法为二语测试设计者提供了新的语言测试选择, 但是同时也为二语测试设计者提出了如何进行测试评估的新问题。作者还指出, 计算机辅助语言测试新技术发展也为二语研究者建构二语测试理论和二语测试研究提供了新的机会。这些机会包括: 对二语测试理论的重新审视、对二语测试理论

的全面建构、对二语测试效度进行重新思考、二语测试与二语习得理论的联系、跟踪测试结果等。

第二章主要讨论计算机辅助语言测试手段与传统的笔试及其面试之间的差异。作者概括出计算机辅助语言测试与传统语言测试五个方面的差异：1) 地点性和时间性 (**physical and temporal**): 传统的测试要求考生在指定时间到达指定地点, 而计算机辅助语言测试可以随时随地为考生提供服务, 大幅度提高测试的便捷性; 2) 评分标准与测试指令 (**rubrics and instructions**): 传统测试中的评分标准与测试指令会因考试时间 (如上午或下午)、地点 (如有无干扰)、监考员 (如发布考试指令声音质量、语速)、评分员 (标准掌握程度) 等的不同而影响考生的实际成绩, 而计算机辅助语言测试可以克服所有这些变异, 从而大大增加测试的公正性; 3) 输入和预期反应 (**input and expected response**): 计算机可以用来展示和快速加工大量数据信息, 可以让考生在接受语言测试时, 获得丰富的语境信息, 如图片、声音、动画, 在这些信息输入和反应方面传统的测试是无法做到的, 从而大幅度提高测试的真实性; 4) 输入与应答之间的互动 (**interaction between the input and response**): 计算机辅助语言测试可以达成输入和反应之间的互动, 通过计算机技术的运用可以进行计算机自适性语言测试并获得即时的信息反馈, 大幅度提高测试的信度和效率; 5) 评价特点 (**characteristics of assessment**): 自然语言加工技术 (NLP) 使计算机可以对复杂的答案进行自动计分, 大幅度提高测试评分的灵活性和准确性。在该章讨论中, 比较传统测试方法作者主要指出了计算机辅助语言测试五大主要差异, 从而说明计算机辅助语言测试的五大主要优点。作者在讨论中突出强调, 计算机新技术的发展使计算机辅助语言使用能力测试成为可能。当语言测试考查的是在不同语境中如何使用语言的问题时, 测试者就必须认真对待测试的语境因素。测试设计要尽可能真实地反映语言的实际使用情况, 这样的测试成绩才能测试出考生在实际生活中使用语言的能力。新技术促成了对语言使用能力的计算机辅助语言测试的可能性, 因此作者认为, 二语教师和语言测试者了解计算机辅助语言测试与传统语言测试的差异, 认识新技术对语言使用能力的测试方法及其对成绩的解释就变得至关重要了。

第三章指出信息技术为二语教师、二语测试设计者和研究者提供了评估语言能力的一系列可能性, 然而同时也产生了新问题。在本章中作者主要概括讨论计算机辅助语言测试中的六个问题或威胁: 1) 不同的测试表现 (**different test performance**)、2) 新任务类型 (**new task types**)、3) 自适题目选择的局限 (**limitations due to adaptive item selection**)、4) 自动反应评分的非准确性 (**inaccurate automatic response scoring**)、5) 妥协的安全性 (**compromised security**)、6) 负面后果 (**negative consequences**)。本章主要讨论了语言测试技术和测试效度之间的关系及其可能出现的问题, 并对所提出的问题或威胁提出了一些解决办法, 但是作者强调, 由于我们目前在计算机辅助语言测试方面的知识和研究水平的限制, 提出的问题比回答的问题要多。然而, 作者认为发现问

题本身就是向问题的解决迈出了第一步。

第四章伊始作者便一方面强调语言测试取决于不同的测试目的, 另一方面强调由于技术的发展计算机辅助语言测试的软件层出不穷, 数量甚多。作者认为不可能也没必要对各种计算机辅助语言测试的软件一一介绍, 他们认为教学软件 **WebCT** 最具有代表性, 所以本章以软件 **WebCT** 为例, 着力介绍了计算机语言测试软件的概念、应用、测试的实施, 同时指出软件设计者所面临的问题以及有可能采取的策略。**WebCT** 是为没有软件专业知识的使用者设计的计算机软件, 该软件可以用来开发完整的网络课程, 将当前开设的课程材料上传, 并可在网上编制测试题和实施测试。例如使用其相关软件程序可以设计多项选择题 (**multiple-choice**)、正误题 (**true and false**)、配对任务 (**matching tasks**)、简短应答 (**short-answer**)、段落写作 (**paragraph-writing task**) 等测试内容。作者以本人所在爱荷华州立大学对 **WebCT** 的使用开发利用为例, 采用网上截图方式较详细、直观地描述了网上语法测试题的编制、上传和施测; 同时对网上文本显示、学生进入测试系统、施测、测试结果报告及其相关测试记录等进行了详细演示说明。这对无计算机软件知识的二语教师学会使用 **WebCT** 软件进行网上计算机辅助语言测试题的编制、试题上传、对学生施测, 或编制自己课堂教学测试题目, 或设计课堂小测验中的个别问题进行施测等都具有非常直观的指导意义。作者还对其它几个相关软件, 如 **Hot Potatoes**, **Discovery School Quiz Center**, **Blackboard** 软件做了扼要的介绍。

第五章指出计算机辅助语言测试与其它测试的形式不同, 所以对其评估的标准也就不同。也正因为如此, 有关学者对计算机辅助语言测试的评估提出了很多不同的评估标准。这里作者首先对研究者在计算机辅助语言测试研制与评价中为提高测试质量而提出的“纲要与建议”进行了概括总结。作者通过列表比较的方法, 主要对 **Noijons** 和 **Fulcher** 提出的评价点进行了详细的引介性描述, 同时指出 **Noijons** 的考前 (**before**)、考中 (**during**)、考后 (**after**) 框架中的主要评价点均涉及试题研制者和测试评估者视角的接口问题; 而 **Fulcher** 从试题研制者视角出发提出的开发计算机辅助语言测试的三段论 (即计划与初始设计、可用性测试、实地试测与微调) 虽然不是出于评估研究目的, 但是为计算机辅助语言测试的评估奠定了基础。作者在本章讨论中指出, “纲要与建议”性的讨论并未对计算机辅助语言测试评估的实证研究提出具体的研究途径和方法, 甚至也未对研究者提出相关的实证问题。接下来, 作者便对计算机辅助语言测试评估的实证研究文献进行了梳理, 把计算机辅助语言测试的评估研究分为三个主要方面, 即质量与效率研究、语言反应评价研究、对情感、效果和效率的影响研究。作者在对这三个方面研究综述的基础上, 指出以往的研究焦点多集中于测试开发者关心的问题, 而且每个个体研究只涉及计算机辅助语言测试评估的某个方面, 见木不见林。基于这一认识, 作者根据自己的研究概括性描述和讨论了一个比较全面的计算机辅助语言测试评估

研究的理论框架, 该框架涉及六个主要维度: 1) 信度 (reliability)、2) 试卷结构效度 (construct validity)、3) 真实性 (authenticity)、4) 互动性 (interactiveness)、5) 影响力 (impact)、6) 实用性 (practicality)。另外, 作者还在本章中介绍和讨论了计算机辅助语言测试评估研究领域新出现的几个理论或研究框架, 即计算机自适应测试“效度验证理论 (validation theory)”、“基于论辩的结构 (an argument-based structure)”, “使用驱动框架 (a use-driven framework)”。

第六章首先指出, 尽管计算机技术在教育评价方面 (包括第二语言) 已经取得了重大进展, 网上试题设计、施测、自动评分等都已成为现实, 但是计算机辅助测试目前尚为达到人们预期的革命性变革, 还存在很多有待研究与解决的问题。作者在本章中主要指出了计算机辅助二语评估中存在的一些问题, 如基于机器模拟、真实语料试题生成等的语言使用能力评价尚未大面积使用, 运用新技术的测试目标也无显著变化, 计算机辅助二语测试在课堂教学中的使用还非常有限等等。另外我们对语言、语言使用、语言学习理解的局限性也在制约着计算机辅助二语测试技术的开发和利用。但是作者认为, 尽管如此, 今天在二语测试方面新计算机技术不仅正在改变着我们的语言测试观念, 也在改变着我们的语言观、语言教学和学习观念。所以研究者应该注意计算机辅助语言测试中长期存在的一些问题, 尤其需要关注相关的语言理解、语言运用和语言习得方面的理论研究。作者在本章中继而论述了应用语言学、二语习得研究与计算机辅助语言测试的关系。他们认为计算机辅助语言测试的开发、利用和评价对应用语言学和二语习得研究提出了挑战, 同时也扩大了应用语言学家的想象空间。以技术为媒介的互动交流 (如人机互动) 要求应用语言学和二语习得研究一方面要重新定义二语能力的本质及其内涵外延, 一方面需要重新思考二语测试的结构问题。

2. 简评

本书在内容结构安排和体裁设计上都很有特色, 是专门为二语教师、二语测试设计者、二语测试研究者撰写的一本有关计算机技术辅助语言测试指导性著作。其主要特色如下:

行文结构脉络清晰, 理论与实践并重是本书的突出特色。本书以计算机辅助语言测试的技术为线索, 首先描述了计算机辅助语言测试技术的最新发展, 同时分析了计算机技术条件下语言测试的优越性和潜在的问题。在澄清了学者关于计算机辅助语言测试的利弊之争之后, 本书较详细地讨论了如何实施和评估这种测试以及这种测试所产生的影响。这样就将理论与实践有机地结合起来, 从可行性到实践性, 详细客观地分析了计算机辅助语言测试的优势、局限及其对语言测试领域产生的影响。

高度概括、大量地使用图表是本书的另一个特色。作者一方面通过列表概括的体裁方法对计算机辅助语言测试的优点和缺点、存在的问题与相应的解决办法等一一进行对比分析, 言简意赅, 其态度观点一目了然; 另一方面通过网络截图对计算机辅助语言测试软件使用的步骤一一列举, 简单

易懂，便于掌握，具有很强的可读性和实用性。

内容丰富、指导性强是本书的第三个特色。与其他相同题材专著不同，可以说本书涵盖了计算机技术辅助语言测试主要领域所涉及的全部内容，即计算机辅助语言测试的技术发展、与传统测试的差异、存在的问题、测试的实施、评估和影响。作者巧妙地将该领域的六个主要构件融于一书，前后连贯，对初涉计算机技术辅助语言测试的二语教师、二语测试人员及其相关专业人士具有很强的指导性。

然而由于本书所涉及的内容非常宽泛，作者不可能就某一个方面有深入细致的详尽论述和讨论，也不可能对计算机辅助语言测试的很多软件或试题作全面的分析评价。正如作者反复指出的那样，计算机辅助语言测试采用什么样的软件取决于测试目的，另外只有对今天的计算机辅助语言测试的问题进行深入研究，才可能恰到好处地在二语评价中运用计算机技术；只有研究计算机与信息技术的最新发展，才能扩大应用语言学和二语习得研究的想象空间，充分发挥计算机辅助语言测试在二语教学中的积极作用。

收稿日期: 2011-01-10

作者简介:

刘永兵: 博士，东北师范大学外国语学院教授、博士生导师。 主要研究方向为应用语言学和话语分析。

Email: liuyb008@nenu.edu.cn。