评价与测试

计算机化语言测试

田文燕

摘　要： 本文围绕测试题型、试题库建设和计算机顺应性测试几个方面，探讨了计算机化语言测试的问题，提出计算机顺应性测试是理想化的测试方式，并就计算机顺应性测试的含义、试题设计及面临的困境展开了讨论。

关键词： 计算机化语言测试；计算机顺应性测试；测试方式
中图分类号： G623.31　**文献标识码：** C　**文章编号：** 1009-2536（2006）03-0078-04

一、引言

基于因特网的语言测试（Internet-based language testing，下文简称IBT）* 从提出到最终实施已近20年了。Canal（1986）曾预言计算机化语言测试（computerized language assessments）的时代即将到来。这也预示了一场测试方法的变革——由纸笔测试（pencil-and-paper tests）向计算机化测试的转变。计算机辅助教学（CAI）逐步深入课堂的趋势必然会影响语言测试的手段。2006年新托福考试将全面实行网上测试（ETS，2005），成为语言测试领域的一个亮点。本文结合语言测试的题型、试题库建设、测试方式等几个方面的进展情况对计算机化语言测试与计算机顺应性测试的现状做一探讨，以便广大英语教师和学习者对这一领域的前沿信息有一定的了解，并指出这一领域目前所面临的主要问题。

二、IBT的题型

（一）多项选择题

在传统的纸笔测试中，旨在测试学习者的词汇量、语法知识和阅读能力的多项选择题和填空题极其易于用到IBT中。因为语言能力常常体现在词汇、语法、阅读理解、听力理解的分项测试中，于是这类题型便形成了IBT题型的雏形。当然，这种题型与能够采用简单的评分系统有关。这在国外许多语言测试的网站上均可见到。这些网站所提供的试题，基本上是一些技术含量较低的试题拼盘。通常，学习者答完试题后会很快得到成绩报告单。也有一些专业网站可提供一些较系统的语言测试，给学习者提供两套测试词汇试题，可用来评估学习者的英语水平；Elite Skills Ltd.（http://www.wordskills.com/level）为学习者提供了三个水平段的测试。Roever（2001）认为，这种测试是一种低风险的测试，非常适合那种只想检测一下学习进展情况或为真正的考试作准备的学习者使用。在这种测试中，学习者毫无理由或根本不会作弊或让他人替考，考生可以选择方便自己的时间、地点，并按照自己的节奏答题。

（二）非选择答题类题型

客观题主要采用多项选择题，而且往往只有一个正确答案。考生只需在所要选的选项上点击

* 本文IBT和computerized language assessments，按照新托福的说法，意思是一样的。因此，这两个术语都译成“计算机化语言测试”。

一下即可，要么得分，要么不得分，但这种试题会影响测试的效度。为了有效地控制考生因猜答案而引起的测试效度问题，命题专家们设计出了非选择答题类题型（constructed-response items），包括完型填空（cloze）*、简答题、口试（interview test）等（Davies，2002）。这种题型要求考生根据对阅读或听力短文的理解，回答一些问题，采用多级评分制。例如，每小题得分值为0—3分，这就要求根据考生答题情况酌情给分。对于这种题型，用计算机软件来评分是一项极严峻的挑战，但这项技术的开发使用现在已有很大进展。以TOEFL所使用的E-Rater** 为例，目前，研究者们正在研究用E-Rater作为新托福的独立写作部分的辅助评分工具的可行性。其特点是基于几种分析来进行评分，既运用了句法、语篇、主题和词汇的分析，也考虑了按照词数来测量短文长度这一特征。因为短文的长度在人工评分时被看作是一个非常重要的变量。1999年以来，E-Rater被用于工商管理硕士入学考试GMAT的分析写作测试工具。学者们认为由于计算机化语言测试和纸笔测试给考生呈现任务的方式不同，从而会影响考生在考试中的表现（Alderson，2000），因为IBT要考察的是考生的语言能力，而不是计算机熟练掌握的程度。

三、IBT的试题库建设

要实现IBT，必须要建一个既有理论依据，又便于操作的试题库。语言测试里的试题库有别于我们平时所说的数据库，因为它不是试题的简单拼盘。试题库里的每一道题都必须在一定的理论模式下进行模拟试测，对题目赋予若干参数，进行必要的等值分析。目前国外试题库建设多只用到单参数项目反应理论模型，即只考虑题目的难度值。专家们也在研究在多参数项目反应理论（考虑题目的难度值、区分度以及答案的可猜测度）指导下的试题库的开发。理想的试题库里的试题应表现在分数的可比性、组织上的严密性、内容上的广泛性、对考生的预测性以及经济上的可行性。因此，试题库是一个具有较大信息量的试题的科学组合，试题库建设是一项系统工程。毋庸赘言，试题库是实现IBT的先决条件，也是IBT所面临的一个挑战（张权，2004）。

四、IBT的理想测试方式——计算机顺应性测试

（一）计算机顺应性测试的含义

计算机顺应性测试（computer adaptive test，简称CAT），指对每一个考生提供难易度合适的测验项目的一种测试方法。其基本过程是：考生首先回答一个中等难度的测验项目。如果考生正确回答了这一初始项目，那么下一个测验项目的难度就要增加；如果考生答错了，那么下一个测验项目的难度就要降低；考生以后每回答一个测验项目，计算机就给出一个相应的能力估计值。这个能力估计值就成为选择下一个测验项目的水平依据（Anastasi，1990）。可见，CAT更注重考试过程，提高了测试精度（张权，2004），可谓理想的计算机化的考试。

因而，IBT的理想测试方式是根据CAT，让每一考生都可从中等难度的题目开始接受测试。通常，只需显示十几道题即可很快评估出考生的实际能力和水平。CAT可灵活掌握开考时间，并缩短实际考试时间，从试题库中随机抽取考题，且试题难度相同，但内容各异。这样，先参加考试的考生无法对后去应试的考生提供任何有关测试题的信息。IBT可将报考时间和开考时间定为一个时间段，而非一个时间点，以确保考生是在其最佳竞技状态下参加考试，从而使考生的水平得以正常发挥。由此可见，CAT更符合认知科学的精髓，是一种以人为本的测试。但是要实现CAT，也面临着极大的挑战，如CAT成本增加的问题、适合于CAT使用的试题开发以及口语测试在CAT中的处境。

（二）CAT的“增值”

在采用CAT前，需要确定它的难度和内容。因为试题功能是通过试题参数来体现的，即：试

---

* 这里的cloze没有给出可供选择的答案，而是要求受试者填入一个词或短语。

** E-Rater是一个功能强大的作文评分软件。ETS从1999开始把这个软件用于托福、GMAT的试卷评阅，目前有2.0版本。

题难度、区分度和题目答案的可猜测因素。评估这些参数的测量模式是项目反应理论（Item Response Theory，简称 IRT）。有些 CAT 只使用一个参数——试题难度，有些则使用三个参数。为了得到稳定的项目参数，需要抽取有代表性的考生样本进行试测。对于 CAT 来说，试测很有必要，但其成本却较高。对于这一点，研究者从理论上给考生做出了这样的解释：CAT 给考生提供了更适合他们能力的试题，考试有一定的趣味性，评分也比较准确（Jamieson，2005）。因此，如前文所述，还是有许多测试在采用 CAT 方式，如 TOFEL，GMAT 等。

（三）CAT 的试题设计

目前的大部分试题并不符合 CAT 的要求，因为 CAT 是需要试测来获取试题的评估参数的，而试测常需要极高的成本和考生样本。因此，命题人员在积极努力改编原有的试题。改编后的试题具有自我测试和分级测试的功能，更加符合考生的兴趣和能力。同时，为了降低成本，研究人员采用了两项技术：（1）让考生先回答一些调查性的问题来选择他们感兴趣的内容；（2）让他们做一些分级测试里的试题来评估他们的水平。这两项技术的使用是为了选择适合他们水平的试题（Jamieson & Chapelle，2002）。设计这种低风险的测试，目的是给英语学习者提供一些有趣的学习经历，给他们提供一些有关其英语水平的反馈及改进其学习英语的建议。测试通常是先问一些简单的问题，如“你为什么想学英语？”“你是做什么的？”等。问题的回答会作为考生选择特殊用途英语（English for Specific Purposes）还是普通英语（English for General Purposes）方面的内容的依据。然后提供有关词汇和语法方面的测试，从回答问题到测试结束，大概需要十五分钟。这种测试结果就会被作为选择适合于他们的水平试题的参照性依据（如初级、中级、高级）。例如：DIALANG（http://www.dialang.org/）就采用自我测试和分级测试来给个体学习者提供相应水平的试题。通过测试，个体语言学习者能够判断自己在词汇、语法、写作、阅读和听力方面的水平。

由于许多试题无法进行试测或从实证的角度来确定它的难度，命题人员也常常依据理论来编制不同水平的试题。例如：词汇试题的编撰依赖于词频数，语法试题依赖于课本编排顺序，书面表达依赖于对课文篇章的分析，阅读和听力理解题依赖于情景加工，而对课文篇章的分析则是基于语料库语言学理论和经验丰富的教师的判断。

（四）口语测试在 CAT 中的两难境地

目前的口试还无法采用 CAT。一般来说，口语测试由人工来评分。而 CAT 要求考生完成一道题或一项任务后，会自动评分，然后根据考生的答题情况，帮助其选择下一道适合考生水平的试题。这在口语测试中，计算机目前是难以做到的，而且命题人员在选择考试任务时，常将口语的“灵活性”（flexibility）和“适应性”（adaptivity）考虑进去。然而，研究人员还是把计算机引入到口试中。计算机化口试（Computerized Oral Proficiency Interview 简称 COPI）是由应用语言学中心研究出来的最新口语测试形式，它是在模拟面试型口试——磁带录音法（Simulated Oral Proficiency Interview—tape recorded 简称 SOPI）和口语测试——人工法（Oral Proficiency Interview—humans 简称 OPI）的基础上研究出来的。

起初，COPI 使用的材料来自 SOPI，并建成了 COPI 试题库。COPI 的测试范围从中级到高级，共有七项测试任务。前四项系自测，后三项会高于自测水平（Kenyon et al，2001）。在对比了考生在 SOPI 和 COPI 中的反映后，有报道说：考生感觉 COPI 比 SOPI 难度低，COPI 更趋向于自然，而且测试任务的难度更接近于考生的实际水平（Kenyon et al，2001）。然而，也有研究者认为，这项报道需要进一步调查研究。我国目前对计算机口语考试的实证研究已有报道（蔡基刚，2005）。如果 COPI 能在我国大规模的普及，将会极大满足考生的需求，但它仍面临着极具挑战性的问题，即口语语料库的开发以及如何更好地解决人机互动的问题。

五、结束语

IBT 的历史虽然只有不到二十年的历史，但从它在国外的研究和应用来看，已向我们展示了广

阔的应用前景。身处计算机时代的外语教师有必要了解 IBT，因为它不但有若干项功能可用于操作测试程序上，如题干的编制与展示，答案的收集与评分，测试结果的统计分析、储存、传送及信息的提取等（Burstein et al，1996），而且还能把广大的英语学习者带入英语学习的另一个广阔天地：充分利用网络来辅助学习英语。由此看来，计算机在测试中的作用是巨大的，但 IBT 在国内的发展仍将需要付出长期的努力。

参考文献：

Alderson，J. C. 2000. *Assessing Reading* [M]. New York: Cambridge University Press.

Anastasi，A. 1990. *Psychological Testing* (6^{th} *edition*) [M]. New York: Macmillan.

Burstein，J. & L. Frase. 1996. Technologies for Language Assessment [J]. *Annual Review of Applied Linguistics*，16.

Canal，M. 1986. The Promise and Threat of Computerized Adaptive Assessment of Reading Comprehension [A]. In Stansfield，C.（eds.）. *Technology and Language Testing* [C]. Washington，DC: TESOL Publications.

Davies，A. & Annie，B. 2002. 语言测试词典 [M]. 北京：外语教学与研究出版社.

Educational Testing Service（ETS）. 2000. *The Computer-Based TOEFL Score Use Guide* [M]. Princeton，N J: Author.

Jamieson，J. 2005. Trends in Computer-Based Second Language Assessment [J]. *Annual Review of Applied Linguistics*，25.

Jamieson，J. & Chapell，C. A. 2002. *Longman English Assessment* [M]. New York: Pearson Longman.

Kenyon，D. & Malabonga，V. 2001. Response to the Norris Commentary [J]. *Language Learning and Technology*，5.

Roever，C. 2001. Web-Based Language Testing [J]. *Language Learning & Technology*，5.

蔡基刚. 2005. 大学英语四、六级计算机口语测试效度、信度和可操作性研究 [J]. 外语界，4.

张权. 2004. 语言测试中的项目分析与等值技术：研究与应用 [M]. 北京：高等教育出版社.

收稿日期：2006-03-12

通讯地址：730050 兰州工业高等专科学校外语系
电子信箱：tianwenyan6508@126.com