

语言测试的理论模式与发展取向

黄齐东¹ 季敏²

(1. 河海大学 外国语学院, 江苏 南京 210098; 2. 三江学院 大学外语部, 江苏 南京 210012)

摘要:作为应用语言学的一个重要分支,语言测试已成为外语教学体系不可缺少的环节。语言测试可分为潜能测试、水平测试、成绩测试、诊断测试等。语言测试的形式呈现多样化的趋势,大体可划分为基于任务的语言测试、计算机适应性语言测试和语料库应用于语言测试。语言测试是衡量外语教学的有效方法,好的测试题应具有效度、信度和实用性,并对学生有积极的引导作用。研究语言测试的理论模式与发展取向,对于试题设计的科学性、评定分数的可靠性乃至实现语言测试的现代化具有重大意义。

关键词:语言测试 模式 应用

中图分类号: H310.4

文献标识码: A

文章编号: 1671-4970(2008)01-0058-05

语言测试理论,又称语言测试学,已发展成为应用语言学领域中一个相对独立的学科。语言测试的用途是多方面的,如用于语言教学、语言研究或选拔人才。《大学英语教学大纲》明确规定,大学英语教学的目的是要让学生掌握语言知识,培养学生运用语言的能力,使大学生能以英语为工具参与交际并获取本专业所需要的信息。而语言测试的目的在于正确和客观地评价学生的语言能力,评估教学质量和教学效果,为教学提供反馈信息。因此,研究语言测试的理论模式与发展取向,对于试题设计的科学性、评定分数的可靠性、测试过程的准确性,乃至实现语言测试的现代化,使语言测试在教学中切实发挥它的指导作用具有重大意义。

一、语言测试的类型

语言测试的分类方法有多种,主要可有以下几种分类方法。

1. 根据目的和用途进行分类

根据目的和用途,语言测试可以分为以下四类:

①潜能测试:用来预示受试者学习某种语言的潜力和天赋,是对其语言能力的全面评估。潜能测试是选择语言类专业人才的依据,如外语院校要专门招收善于学习语言的人,硕士生、博士生复试,考查其外语水平时使用等。②水平测试:是衡量受试者语言应用熟练程度和整体语言水平的测试,目前,国内外著名的水平考试有 TOEFL、IELTS、UCCPE、EPT、

WSK、CET 等测试,其中 CET 考试更偏向于尺度参照性考试。③成绩测试:是考查受试者学习一门语言课程或参加某种语言培训班之后所达到的语言熟练程度的测试。这种考试一般要参考相应的教学大纲。其使用规模相对较小,多被用于阶段性测试或随堂测试。④诊断测试:用来发现教和学两个方面的问题,可以考查单个项目,也可以考查综合性的项目,其目的是为了改进教学方法,调整教学计划。

2. 根据测试题型和项目结构进行分类

根据测试题型和项目结构,语言测试可以分为以下两类:①分离式测试:考查单项语言知识。如:语音、语法、词汇等的语言测试。②综合性测试是一种要求学习者同时运用多种语言技能的测试。如听写要求学生同时运用听力、语法、词汇等多种技能,完形测试要求学生同时运用语法、词汇、语篇分析和逻辑推理等方面的能力。目前国内外举行的大规模语言测试中,往往是以上两种测试题型的混合,集多项选择题、完形测试题、阅读理解和写作于一体。

3. 根据测试内容进行分类

根据测试内容,语言测试可以分为两大类:①语言形式测试:其测试目的主要考察受试者的主动输出技能,如说和写和被动接收的技能,如听力和阅读。②语言功能测试:主要测试语言技能的应用,如完形填空、听写、翻译等。

二、语言测试的三种模式及其应用

目前,语言测试主要有三种模式,即心理测量学—结构主义语言学模式、心理语言学—社会语言学模式和交际语言测试模式。这三种模式有着各自的特点和适用范围。

1. 心理测量学—结构主义语言学模式

在当前各类测试中所广泛使用的多项选择题源于心理测量学—结构主义语言学模式。以 Bloomfield、Fries、Lado 为代表的结构主义语言学家重视外语与母语之间的系统差异,预见并描述外语学习中的难点。结构主义语言学理论用直接成分分析法对语言进行描述,把语言分离成不同层次的单位。该理论认为语言是一个由不同的语言单位构成的层级系统,最小的语言单位是音位,音位构成语素,语素构成词或词组,词或词组构成句子。基于可分性语言能力假设,心理测量学在结构主义的影响下,把语言当作系统参照式来进行测试。这个参照式可细分为语言技能(听、说、读、写技能)和语言成分(语音、语符、语法、词汇)^[1]。语言能力可以分解为许多微技能,语言能力便是这些微技能之和^[2]。心理测量学方法主要使用分立式试题,采用多项选择题,由题干和选择项构成。考生需要排除干扰项,选出正确答案。为了保证测试信度,分立式试卷往往要经过命题预测、项目筛选、内容组合等复杂程序。分离式测试在各类大规模标准化外语考试中普遍采用。

心理测量学—结构主义语言学模式指导下的分立式测试法具有明显优点,即测试项目的覆盖面广,所获得的数据易于量化,但是分立式本身也有缺点。受当时盛行的行为主义心理学的影响,结构主义语言理论认为语言学习过程是一个机械的习惯形成过程,语言学习就是通过反复地刺激、反应和巩固而逐步形成的一种习惯^[3]。为保证测试的信度,分立式测试将听、说、读、写分项测试,而且多采用客观题形式,因此,在分析语言结构基础上进行大量的脱离语言环境的句型操练成为外语教学的主要内容,忽视它们与交际语境的相互作用。

2. 心理语言学—社会语言学模式

尽管心理测量学测试信度高,但是它的效度存有争议,即对分离的语言成分(语言知识)的考查是否能有效地考查语言能力。与心理测量学观点截然不同,心理语言学—社会语言学模式认为语言具有心理现实性,这种心理现实性表现为不可分离的语言能力。语言应该作为语用期待语法来考查,如果一个人掌握了某种语言能力,他就应该知道某个词或某句话是否可以在某一特定上下文或情景中

出现^[4]。在语言测试中应该采用综合试题或语用试题来考察考生的这一不可分的语言能力。考生被要求同时具备应用语篇能力和有关外部世界的知识。心理语言学—社会语言学模式推崇的题型主要有两类:完形程序和听写。完形程序要求考生填入词项或可接受词项。听写题型要求考生将听写语篇中被删除的语言成分恢复完整,而这些被删除的语言成分在长度上要超出人们的短期记忆所能容纳的量。

心理语言学—社会语言学模式认为完形填空和听写之类的整体性综合测试测量考生综合各语言技能的能力,其方式更接近语言运用的实际过程。然而,完形填空乃至听写,又在多大程度上反映语言运用的真实情况呢?心理语言测试模式将语法视作所有语言技能的依托和唯一因素,将接受(或理解)和运用(或输出)等同起来,而实际上它们是两种不同的过程。在设计或开发一项新的考试时,应首先考虑语言测试行为要与语言的实际情况相一致。因此,必须能够证明被试的测试行为与在具体场景下的语言使用存在一致关系。要做到这一点,首先要考虑的是使目标语言使用任务与测试任务保持同步关系,而心理语言测试模式恰好忽视了这一点。

3. 交际语言测试模式

针对 Chomsky 的转换生成语法观,Hymes 首先提出了“交际能力”这一概念。Hymes 认为,交际能力包括语言形式的正确输出、语句运用的切实可行、语言表达的得体周全和交际目的顺利实现四个方面的内容^[5]。Halliday 随后提出了“语言功能学说”,他将语言的功能分为概念、操纵、启发和想象四大类。这些功能是相互交织在一起的,学习语言就是获得语言知识和技能来行使语言的这些功能^[6-7]。Halliday 的语言功能学说以及交际法教学思想对交际语言测试理论模式的形成和发展有着直接的影响。

20 世纪 80 年代初,Canale & Swai 提出交际能力包括四种能力:语法能力、社会语言能力、会话能力和策略能力,该理论影响了 80 年代对交际能力测试的研究^[8]。到了 20 世纪 90 年代初,这种理论被 Bachman 提出的理论所替代。Bachman 把交际语言能力分为三部分:①语言能力:即语音、词汇、语法的内化知识;②策略能力:即使用语言的社会变体和言语功能的能力;③心理生理机制:即语言交际中的各种应对手段,如怎样开始谈话、话题转换、结束谈话等等^[9]。以上交际能力说明,学习一门外语不仅仅是学习它的形式、系统,还要学习它的使用规则,而语言测试的目的就是要考察学习者对语言的掌握情况^[10]。Bachman 的理论对 20 世纪 90 年代有关交际能力测试方面的研究起了十分重要的作用。其中尤

为突出的是对一些交际语言能力测试方面的概念定义的研究,如对测试构想和交际测试定义。

交际语言测试反映了人们对语言本质的新认识,体现了一种科学的语言观和语言学习观。在交际能力测试上影响较大的是美国外交服务学院的面试(FSI)模式。在FSI考试中,考官向考生提问,考生的回答即是向考官提供的本人的口语样本,随后考官根据考生语言的流利程度、准确性和得体性等几条标准来评判并打分。一些经过改良后的FSI式口试被用来评估学生的交际能力^[11]。在我国广泛进行的一些口试也采用了这种形式,如大学英语口语测试、全国公共英语等级考试体系的口语考试以及目前剑桥商务英语考试,便是交际语言测试思想的产物。然而,语言交际能力是由多种因素构成的一个复杂的综合体,如何对它进行准确、有效的测量一直是外语测试界亟待解决的问题。

以上三种测试模式各有优点和不足,但重要的不是比较它们的优劣,而是根据它们各自的特点加以充分利用^[1]。心理测量学测试因信度高、评阅迅速可靠,适用于大规模外语测试。但由于这种测试是以一定程度上牺牲效度来取得高信度的,因此应避免频繁使用心理测量学方法,防止其对教学产生不良反拨作用。心理语言学测试采用的是有限答案题型,设计简便,具有较高的准则关联性效度,评阅比较客观,适用于不需要诊断信息的语言测试,如新生入学后的分班考试及一些课堂考试^[12]。但由于它不能准确反映学生在学习中所碰到的问题,因此不可过高估计它在练习中的作用。交际语言测试是考察应试者语言运用准则的参照式测试,适用于为每一教学单元配置语言运用性练习,这种为掌握课程中每一选定的语言项目而设计的练习正好弥补准则参照式测试的不足。当然交际性练习只是课程练习的一部分,后者还应包括语言系统性练习。直接性系统参照性测试,即传统语言测试,可为语言系统性练习提供较为理想的题型。

三、语言测试的发展取向及理性认识

在迅速发展的语言测试理论的影响下,语言测试的形式呈现多样化的趋势。但从总体上看,大体可划分为以下3个方向。

1. 基于任务的语言测试

传统的语言教学方法把分离的语言形式教给学习者,让其反复操练。这种理念背后的假定是“输入”和“吸收”之间存在一种直接联系。而实际上,输入和吸收之间并没有直接联系,语言学习者并非完全按照呈现给他的东西来学习语言^[13]。于是,研究

者提出了“关注形式”^[14,15],其特点是兼顾语意和形,即语言教学仍以有目的交际和语意表达为核心,但不忽视语言形式的学习,尤其当学习者在交际活动中出现理解或表达困难时,需要将其注意力短暂地集中到语言形式上来。对基于任务的教学进行评估既应包括形成性的,又要包括终结性的,既有过程评估,又有结果评估,但核心是对学生的学习成绩作出评价,由此便产生了基于任务的语言测试(TBLA)。

Brindely和Norris把“任务”定义为“人们在日常生活中所从事的各种活动,但这些活动需要靠语言来完成”^[16,17]。这里所说的“任务”实际上就是一种真实世界的活动。Bachman & Palmer提出了设计和开发语言测试所必须遵循的一条重要原则,即语言测试任务特征必须和实际生活中的语言运用任务特征相一致的原则^[18]。这里涉及两个核心问题:①如何精确地确定、选择和描绘现实中的任务;②如何解释测试中使用的任务与现实中的任务之间的相关关系。对于任务的确定、选择和描绘,Bachman & Palmer提出的解决方案是确定目标语言使用域。目标语言使用域指的是应试者在非测试环境下可能遇到的一系列具体的语言使用任务,是测试者从中选择具体测试任务的基础。在实施TBLA时,如果测试者对应试者的背景了解清楚,目标语言使用域的界定就比较容易,测试任务的选择也相对轻松。Henning认为,真实性是语言测试的核心问题^[19]。这是因为只有根据真实性测试的结果,才能推测考生在目的语使用域中的语言使用能力,而且,测试的反馈作用能促使学生学习真实的语言材料,从而更有利于提高他们的交际语言能力^[20]。但如果应试者来自不同背景,使用语言的目的和需求不一,此时目标语言使用域的界定就很难,测试任务的选择就带有较强的主观性,测试内容的相关性和代表性就会受到影响^[21]。目标语言使用域确定以后,接下来的问题是选择测试中使用的任务。一些学者建议采用“需求分析法”^[17,22],但Bachman和Palmer认为,即使已经有了一个明确的语言使用域,从中选取具体任务作为测试任务的做法仍有问题^[18]。以现实生活域为例,直接将其中的任务作为测试任务有时就不合适,更重要的是,目标语言使用域中有众多任务,选择任务的标准是什么?如何保证选取的任务有充分的代表性?对这些问题需要进一步研究。

2. 计算机适应性语言测试

正如语言教学强调个性化和适应性一样,语言测试也开始注重体现适应性和个性化的特征,即语言测试要向以受试者为中心的方向进行发展,测试

的过程要在最大限度上适应受试者的个性特征(如受试的语言水平和认知特征等)^[23]。这种趋势促使研究者开始重视调查二语习得者的特征以及二语习得过程中的认知因素、学习策略和情感因素。不少学者已经开展了一系列理论与实验研究^[24-26]。计算机适应性语言测试为这种发展方向提供了可能。

计算机技术的发展不仅影响了外语教学的整个过程,而且也使得语言测试的技术手段和载体发生了革命性的变化。2001 年在美国举行的第 23 届国际语言测试研讨会的主题就是基于计算机技术语言测试新技术新方法。其中,最具代表性的是计算机适应性语言测试,它是指把计算机适应性测试的原理和方法运用到语言测试中去。随着 GRE、GMAT、TOEFL 等都已逐步改成机考,人们对量体裁衣式测试(CALT)这一术语也渐渐熟悉。

CALT 为量体裁衣式测试。这是因为计算机可以根据每个考生的不同能力生成不同的试题,以便在尽可能短的时间内精确地测试出每个考生的真实能力和水平。具体地说,计算机首先给考生选择一道对大部分被试者来讲具有中等难度的试题,如果考生答对了,计算机将自动选择下一道较难的考题;如果答错了,下一道考题则较易。在获得考生的其初始能力值后,计算机就会自动从题库中再选取另一组能够测试出该考生能力的项目。CALT 使测试的个性化得以加强,使考试成绩尽可能真实地反映出考生真正的能力。同时,CALT 测试作为一种无纸化考试,保密性和安全性也比较高。CALT 题库具有项目多、题型广、难度不一,因而可较全面地测试每个考生的语言水平。另外,CALT 有利于测试开发的规模化,国外有不少公司在这方面有相当不错的成果。我国的 NMET、CET、PETS 等大规模测试的组织 and 实施费时费力,而且还存在反馈作用不良等情况。CALT 的确是一条可行之路,我们应加强理论研究,适当引进国外现有的测试软件,加快建设新题库,进一步充实完善现有题库^[27]。

3. 基于语料库的语言测试

语料库是由语言学家与计算机专家基于语料库语言学所建立的人类自然语言数据的汇集,它是一个庞大的语言素材的集合体,主要用来观察、分析和研究目标语言的各种特征。语料库在建设初期主要用于语法分析和语言对比,但随着语料库在类型和用途上的多样化,语料库的主流用户也由原来的语言学家和词典编纂者发展为如今的语言教师和学习者^[28]。语料库分为标注语料库和非标注语料库。依据这个分类,可以建立许多相对不同类型的语料库,而这为语料库应用于语言测试提供了可能性。

语料库应用于语言测试具有以下三个明显的优势。①从题库设计与题型选择来看,它能提供真实口头语和书面语言材料,而且在标注语料库中,人们可以获取表现特殊的语言学特点的例句或在篇章中根据索引设计出排序重组题或选词填空题等;②从试题选择与传送来看,受试者或出题者只要指定所需材料的特点,计算机就能提供所有合适的材料供人们挑选。针对受试者的判断结果,语料库会直接地做上特殊的标记,然后计算机把应试者的答案与库中的记载相比,识别对与错,以便为受试者下一步的选送试题打基础;③从能力评估来看,基于语料库的语言能力诊断评估系统可为系统输出合适的评估模板(包括相应的评估目的、评估内容、评估参数以及评估报告等具体的参数值),形成各类语言能力诊断评估模式,进行词汇、语法、句法、语用能力的评估。TOEFL 和 GRE 就是语料库运用于语言测试的最典型、最成功的例证。这两种测试在实践上证明了语料库用于语言测试的可能性。从 1999 年 10 月份开始,GRE 在我国许多大中城市也实行了受试者机上测试,由于机上考试完全符合测试评估的三个方面的要求,即有效性、可靠性和可行性,因此其结果深受受试者好评。这也就说明语料库应用于语言测试不仅有其理论的可能性,而且也应该有其实际运用的可能性。

语言测试模式的变化反映了语言学界对语言本质属性研究的进展,体现了测试设计者对测试结构的深入了解。分列式及知识性题目的减少,综合性及语言运用题目的增多,反映了测试开发者对测试内容的进一步理解。总的发展趋势是,知识型测试向技能型转变,死记硬背让位于语言运用^[29]。在改进测试方面,英国文化委员会的经验值得我们借鉴。20 世纪六十年代初,他们开发了传统型测试模式 BCSCA,采用五分制,但由于 BCSCA 的测试效果并不理想,因此后来被英语能力测试小组(EPTB)所取代。到 20 世纪七十年代中期,Hymes 提出的交际语言能力观对应应用语言学界产生了很大影响,使 EPTB 在理论上显得落伍了,他们又设计出新的英语语言测试服务(ELTS)。这项英语测试以“需要分析”为基础,属于专门用途英语测试,但实施几年后,ELTS 的语言水平和结构模式又受到质疑。后来,英国文化委员会、英国剑桥大学测试中心和澳大利亚高校国际开发署共同开发出面向国际的雅思(IELTS),颇有影响。其改革进程体现了语言学不断发展的理论思想。

五、结 语

语言测试研究承受了巨大压力,一方面,语言测

试要从语言学领域寻求理论依据,另一方面,它又要解决本领域的实际问题。众所周知,语言测试具有独特的信息反馈功能,它能准确地考核结果并产生敏感的试后反拨效应,因此是外语教学体系不可缺少的辅助手段。目前,我国语言测试仍存在一定的问題,广大语言教师天天都在使用测试手段,年年出題、阅卷、评分,然而,所设计的试卷是否科学,是否测出了应测的项目,是否具有延续性与继承性,测试目的是否实现,应引起广大语言教师的重视。因此,研究语言测试的理论模式与发展取向,对于正确和客观地评估外语教学质量和教学效果,为教学提供反馈信息,具有重大意义。

参考文献:

- [1] 王振亚. 现代语言测试模式及其应用[J]. 外语教学与研究, 1994(3): 7-9.
- [2] LADO R. Language testing: The construction and use of foreign language testing[M]. London: Longman, 1961: 15-16.
- [3] RICHARDS J, RODGERS S. Approaches and methods in language teaching: A description and analysis[M]. Cambridge: Cambridge University Press, 1986: 126-125.
- [4] 戴曼纯, 肖云南. 现代语言测试发展: 理论与问題[J]. 湖南大学学报(社会科学版), 1999(2): 33-36.
- [5] HYMES. On communicative competence[M]. Harmondsworth: Penguin, 1972: 89-90.
- [6] HALLIDAY M A K. Explorations in the function of language[M]. London: Arnold, 1973: 64-66.
- [7] HALLIDAY M A K, HASAN R. Cohesion in english[M]. London: Longman Group Limited, 1976: 55-57.
- [8] CANALE M, SWAIN M. Theoretical bases of communicative approaches to second language teaching and testing[J]. Applied Linguistics, 1980(1): 1-47.
- [9] BACHMAN F L. Fundamental consideration in language testing[M]. Oxford: Oxford University Press, 1990: 25-26.
- [10] 施卫民. 语言交际能力与测试[J]. 西安外国语学院学报, 2001(5): 35-39.
- [11] 鄧申. 试论口语测试的真实性[J]. 外语界, 2001(3): 74-76.
- [12] BIKER D. Language testing: A critical survey and practical guide[M]. London: Edward Arnold, 1989: 120-121.
- [13] ELLIS R. The study of second language acquisition[M]. Oxford: Oxford University Press, 1994: 101-105.
- [14] DOUGHTY C, WILLIAMS J. Focus on form in classroom second language acquisition[M]. Cambridge: Cambridge University Press, 1998: 71-72.
- [15] LONG M H. Focus on form: A design feature in language teaching methodology[M]. London: K de Bot, 1991: 33-35.
- [16] BRINDLEY G. Task-centered assessment in language learning: the promise and the challenge[C]//John. Papers presented at the annual international language in education conference. Hong Kong: Language and Learning pub, 1994: 51-52.
- [17] NORRIS J M, BROWN J D, YOSHIOKA J. Designing second language performance assessments[M]. Honolulu: Second Language Teaching and Curriculum Center, University of Hawaii at Manoa, 1998: 136-137.
- [18] BACHMAN F L, PALMER A S. Language testing in practice: designing and developing useful language tests[M]. Oxford: Oxford University Press, 1996: 36-37.
- [19] HENNING G A. Guide to language testing: Development evaluation research[M]. Cambridge: Newbury House Publishers, 1987: 41-42.
- [20] 李清华. 论交际测试中的真实性[J]. 外语界, 2001(3): 68-75.
- [21] 韩宝成. 语言测试的新进展: 基于任务的语言测试[J]. 外语教学与研究, 2003(6): 352-358.
- [22] LONG M H, NORRIS J M. Task-based language teaching and assessment[M]. London: Routledge, 2000: 38-40.
- [23] 曾用强. 个性化自适应测试探索[J]. 外语教学与研究, 2002(2): 278-283.
- [24] LUNZ M E, BERGSTROM B A. An empirical study of computerized adaptive test administration conditions[J]. Journal of Educational Measurement, 1994(3): 251-263.
- [25] NEVO N. Test-taking strategies on a multiple-choice test of reading comprehension[J]. Language Testing, 1989(2): 199-215.
- [26] RICHARDS J, RODGERS S. Approaches and methods in language teaching[M]. London: Cambridge University Press, 1995: 98-99.
- [27] 孔文, 李清华. 大规模语言测试的方向: 计算机适应性语言测试[J]. 外语界, 2002(2): 76-79.
- [28] 肖依虎, 潘翠琼. 语料库在语言测试中的应用[J]. 外语教学, 2002(1): 93-96.
- [29] 肖云南, 罗晓英. 关于英语测试的改革[J]. 外语教学与研究, 2002(3): 23-26.

