

# 汉语测试的理论与 实践创新研究

邢红兵 王佶旻 主编

世界图书出版公司

北京·广州·上海·西安

图书在版编目(CIP)数据

汉语测试的理论与实践创新研究/邢红兵,王佶旻  
编. —北京:世界图书出版公司北京公司,2014.1  
ISBN 978-7-5100-6957-4

I. ①汉… II. ①邢… ②王… III. ①汉语—  
测试—文集 IV. ①H195-53

中国版本图书馆 CIP 数据核字(2013)第 232548 号

---

汉语测试的理论与实践创新研究

---

主 编:邢红兵 王佶旻

责任编辑:武传霞

---

出 版:世界图书出版公司北京公司

出 版 人:张跃明

发 行:世界图书出版公司北京公司

(地址:北京朝内大街137号 邮编:100010 电话:64077922)

销 售:各地新华书店和外文书店

印 刷:北京博图彩色印刷有限公司

---

开 本:787 mm × 1092 mm 1/16 印张:12.5

字 数:180 千

版 次:2014 年 1 月第 1 版 2014 年 1 月第 1 次印刷

---

ISBN 978-7-5100-6957-4

定 价:35.00 元

---

如发现印装质量问题,请与本公司联系调换  
版权所有 翻印必究

# 前 言

北京语言大学汉语水平考试中心是一家集“产、学、研”于一体的专业考试机构。二十几年来,汉考中心汇聚了一批致力于语言测试研究和实践的专业人才,他们当中有很多是四十岁以下的年轻人,思想活跃,富有创意。这部论文集的多数作者就是汉考的青年科研工作者,他们在语言测试理论和实践领域做出了创新和尝试。

论文集分为理论篇与实践篇两大部分,理论研究部分涉及语言测试理论研究的核心领域——语言能力研究和效度研究,也涉及语言测试的历史发展以及与测试研究相关的语料库建设。实践篇涵盖了语言测试研究领域的诸多方面,包括命题技术、题型研究、考试质量分析、考务管理系统研究以及单项语言技能研究等。这些研究从不同角度对语言测验的设计、开发和质量控制做出了探索。

北京语言大学汉语水平考试中心是中国汉语水平考试(HSK)的发祥地。二十几年来研发和实施了 HSK 和 C. TEST(实用汉语水平认定考试)等标准化考试。论文集中很多文章涉及对这两种考试的研究和分析。这些论文对改进和发展 HSK 和 C. TEST 起到了一定的推进作用。

语言测试是应用语言学的一个分支。她是一门交叉学科,语言学和心理学是其主要支撑学科;她也是一门特色学科,兼具人文科学的眼光和自然科学的精神。论文集的出版得到了北京语言大学语言学及应用语言学特色重点学科建设经费的支持。学科建设离不开科学研究和人才培养,我们也希望通过自身的努力,把语言测试的学科建设做好,更好地服务于语言学及应用语言学学科发展和社会需求。

编 者  
2013 年秋

# 目 录

## 理 论 篇

现代语言测验的历史发展及其对汉语水平考试的影响 .....	王 佶 旻 (3)
关于汉语中介语语料库建设的基本原则 .....	张 宝 林 (16)
C. TEST 口语面试引导和评分摭谈 .....	黄 理 兵 (27)
效度理论研究的主流与非主流观点 .....	赵 琪 凤 (41)
语言能力理论再思考 .....	韩 阳 (55)

## 实 践 篇

汉语水平考试中的词汇测试刍议 .....	龚 君 冉 (71)
HSK (高等) 考试长度合理性的多元概化理论研究 .....	陆 一 萍 (81)
美国学生实用汉语阅读能力的训练与评价 .....	李 慧 (91)
实用汉语水平认定考试 (C. TEST G 级) 管理系统数据库设计与实现 .....	黄 伟 (101)
基于 BPM 的 HSK 考务业务流程管理的研究 .....	刘 喆 (120)
HSK 考生汉字书写能力分析 .....	李 郁 (132)
一种新听力理解题型的命题研究 .....	朱 雯 静 (152)
HSK (改进版) 中高级试题质量分析 .....	王 娟 (164)
HSK 入门级阅读理解第一部分题型研究 .....	石 晨 (179)



# 理 论 篇



# 现代语言测验的历史发展 及其对汉语水平考试的影响<sup>\* 1</sup>

王佳旻

**摘 要** 现代语言测验的历史发展分为三个主要阶段：前科学时期、心理测量——结构主义时期以及心理语言学——社会语言学时期。前科学时期是客观性考试的萌芽时期。心理测量——结构主义时期是测验的独立时期，它的标志是使用分立式考试。心理语言学——社会语言学时期是测验的发展时期，它的特点是关注测验的理论构想，追求测验的真实性和交际性。汉语水平考试（HSK）的产生和发展受到现代语言测验理论发展的影响，经历了从分立式考试向心理语言学——社会语言学时期转变的过程。对 HSK 的未来发展，本文提出三点建议：（1）坚持科学、可靠、稳定的质量标准，（2）以 HSK 为基础制定汉语考试的行业标准，（3）在汉语能力标准的框架下开发考试。

**关键词** 汉语水平考试，心理测量学，分立式测验，语言能力标准

---

\* 本研究得到国家社科基金青年项目（12CYY028）和教育部人文社科重点研究基地重大项目（06JJD740005）的支持。



# 引 言

汉语水平考试（HSK）是汉语作为第二语言的能力测验，与其他二语测验一样，它的产生和发展受到现代语言测验理论发展的影响。现代语言测验主要是指 20 世纪发展起来的客观性考试。它的发展经历了三个时期：前科学或传统时期、心理测量——结构主义或现代时期以及心理语言学——社会语言学或后现代时期（Spolsky, 1995）。在语言测验发展的三个历史时期中，前科学时期是客观性考试的萌芽时期。心理测量——结构主义时期是测验的独立时期，语言测验真正成为一门相对独立的学科也正是在这一时期，这一时期的标志是使用分立式考试。心理语言学——社会语言学时期是测验的发展时期，在这一时期，人们开始更多地关注测验的理论构想，追求测验的真实性和交际性。

## 1 语言测验三个历史时期的发展特点

### 1.1 前科学或传统时期

前科学时期是从考试制度的确立一直到上世纪 50 年代，这个时期最大的特点是心理测量学的引进与客观性语言测试的产生。前科学时期人们普遍使用作文和开放的考试，并由一位被授权的主考凭直觉给出测验的结果。在考试发展的过程中，人们逐渐注意到，作文和答案开放的题目缺乏稳定性，同时，不同的评分者之间也缺乏一致性。为了解决技术不可靠这个问题，研究者们引进了心理测



量学的概念。心理测量学对测验的影响使得测验开始走向客观化和科学化，语言测验也在这样的背景下发展起来。和其他的客观性测验一样，现代语言测验也是在一种冲突中产生的。一方面，考试具有明白无误的效力和令人敬畏的力量；但另一方面，考试的分数却显然又具有不确定性。这种矛盾和冲突最终在美国得到了全新的阐释，客观性语言测验真正登上了历史舞台，并且很快就变成了美国教育测量的主导力量之一。客观性测验有两个突出的特点：第一，能在很短的时间里测出学习者在一年甚至几年时间里学习的语音系统、语法形式或者词汇的代表样本；第二，可以用速度和难易程度来评分，并且可以采用机器评分，这样就为大规模考试创造了条件。

上世纪 20 年代前后，一大批语言测验在美国编制出来并予发表，到 20 年代末，已发表的测验大约有 50 多种。这其中就有 Frederick Kelly 发表的第一份多项选择题测验和 Charles Handschin 编制的第一个标准化现代语言测验。不论这些测验用在学生个体时有什么弱点，但和以往相比，这些研究在整体水平上更为全面和精到。在这个时期，涌现出一批职业的语言测验开发者，他们之间的交流非常频繁。尽管在如何设计测验上从来没有一致过，但在测验应该如何检验和证明方面，人们找到了共同的基础，也就是通过试验和统计建立起测验的信度和效度，这正是心理测量学带给人们的进步。

尽管客观性语言测验逐渐发展起来，但它有两个突出的问题。第一是和语言学脱节。当时，语言学的功能是进行哲学探索，而不是思考语言的性质以及如何学习或习得语言。当人们探索语言测验理论的时候，语言理论的重要性却没有显现出来，并且，语言测验



和语言理论的分离状态延续了很多年。这一时期测验先驱者的另一个弱点在于，他们没能解决如何建立效度的问题。他们从未注意到，新的测量工具和学生接受外语教育的方式之间缺乏磨合。他们也没有寻求创建任何关于语言能力的理论或观念，没有去思考在语言测验中哪些要素该测及怎么测。

## 1.2 心理测量——结构主义或现代时期

心理测量——结构主义时期大约是从上世纪 50 年代到 70 年代。50 年代以前，人们关心和争论的是如何确立标准化的客观性考试的地位；50 年代以后，人们着手解决实践和理论中的问题。这一时期的测验深受心理测量学和结构主义语言学的影响，主要测量独立的语言结构和规则。

上世纪 50 年代到 60 年代，语言理论和语言测验理论首次结合起来，开始为语言测验的实践提供基础，在语言理论方面，它是和美国结构主义语言学连在一起的。虽然常有分歧，但人们都倾向于认为，语言应被看成是由独立的结构组成的，因此，最好的办法是研究语言的组成成分。在方法论上，研究者对分析和分类有共同的信念，希望把他们研究的语言要素分离并列举出来。

语言学的结构主义理论，要用于教学和测验，必须得有相应的学习理论，但结构主义语言学本身并没有直接给出这样的理论。于是，应用语言学界便从行为主义心理学那里引进学习理论。行为主义心理学是非常重视人类学习规律的心理学流派，其创始人华生和发展者斯金纳都提出过自己的学习理论，特别是斯金纳的操作学习理论对心理学和教育学都产生了很大的影响，这也是应用语言学领域采纳行为主义作为学习理论的支柱的原因之一。



语言学的结构主义和心理学的行为主义结合，是这个时期语言测验的主要理论基础。在这样的理论影响下，分立式测验登上历史舞台，从那时起，分立式测验几乎成了主要的测验形式。经典分立式测验理论可以认为是由两条轴组成的：一条轴包括被测的技能，也就是听、说、读、写四种主要技能；另一条代表四种主要语言成分，即音位、词法、句法、词汇。在第二条轴上，语言可按分立式方法无限制地分下去。测验内容可以认为是两条轴的交叉点，在这一点上，特定的语言知识表现为特定的语言行为。分立式测验流传很广，至今仍不同程度地出现在各种测验中，其中最著名的就是美国教育测验服务中心（Educational Testing Service, ETS）在上世纪60年代研制的托福考试。

心理测量——结构主义时期的主流测验形式是分立式测验，但几乎是在分立式测验诞生的那一刻起就有人提出了不同的观点。Carroll（1961）提出了综合式测验的概念，在他看来，综合式测验意在测量“一个话语的总的交际性效果”。Carroll的观点是，综合法有很多优点，因为它允许“在语言项目的整个范围内更广泛地取样”。但可能人们当时热衷于分立式测验，这个观点被冷落了差不多十年。今天，我们看到，综合式测验越来越受到人们的重视，逐渐成为测验的新趋势。

### 1.3 心理语言学——社会语言学或后现代时期

心理语言学——社会语言学时期从上世纪70年代一直延续到现在，它是在对心理测量——结构主义观点的反驳中诞生的，它发展了语言测验领域综合的——社会语言学的观点。这种观点把语言表现和语言能力联系起来，也考虑到语言在各种不同的交际情景中



的使用。

后现代时期语言测验的进步，表现在人们在不同方向上对语言能力的探索。首先是 John Oller 提出“一元能力”说（Unitary Competence Hypothesis, UCH）。他认为语言能力是一个单一的能力，并用因素分析证明了这个单一的能力。不难看出，Oller 的一元能力说与心理学家 Spearman 所提出的一般因素（g 因素）观点极为类似。Oller 的理论引起了广泛的讨论，70 年代末到 80 年代初，至少有二十多人做过因素分析以证明或证伪 Oller 的假说，至于参加讨论的人就更多了。Oller 的假说产生过很大影响，但基本上是一个不很成功的假说。

70 年代末以来，美国兴起了一场“能力运动”（proficiency movement）。“能力运动”认为，以往在教学法上的论争是“选错了战场”，语言能力表现为有效的交际，语言教学和语言测验都应以语言能力为目标。于是，语言测验的研究者部分地放弃了结构主义语言学加行为主义学习论的模式，转而在 Dell Hymes 理论的影响下探索语言交际能力。先有 Canale & Swain（1980）的交际能力模型，到 90 年代，Bachman（1990）又提出了一个更精致的交际能力模型，并赢得了好评。

心理语言学——社会语言学时期语言测验更有意义的革新是在语言能力标准（或量表）的制定方面。能力标准旨在对不同的语言能力水平等级进行定义和描述，从而为语言教学、学习和测验提供统一的标准参照体系。在能力标准的框架下，语言测验在定义考生能力范围、评价考生语言表现和丰富分数解释方面都有了统一的标准。

在 70 到 80 年代的英国，相继出现了 Wilkins 的《功能意念大



纲》和 Munby 的《交际大纲》，这两个大纲是语言能力标准的雏形。80 年代，美国制定了一个非常重要的语言力量表——ACTFL 大纲，大纲对听、说、读、写四项分技能做了总体等级描述。以大纲为依托开发了三项语言测验，即口语面试（Oral Proficiency Interview, OPI）、写作能力测验（Written Proficiency Test, WPT）和综合表现测验（Integrated Performance Assessment, IPA），其中 OPI 和 WPT 是分别根据 ACTFL 大纲中口语和写作能力等级标准而设计开发的。进入 21 世纪，欧洲推出了一个比较成熟的能力标准《欧洲语言共同参考框架》（CEF）。CEF 对外语学习者能力水平进行了界定、划分和描述，对欧洲各国的语言教学、学习以及测验编制都起到了纲领和规范的作用。在 CEF 诞生后，包括托福考试在内的许多著名的语言测验都纷纷把它作为分数比照和解释的参考框架，这也在一定程度上体现了一个成熟的能力标准所具有的作用和影响力。

## 2 语言测验理论的发展对 汉语水平考试（HSK）的影响

汉语水平考试（HSK）是和对外汉语教学一起成长起来的。上世纪 80 年代，留学生数量增多，各个学校对选拔性的汉语考试的需求也就逐渐凸显出来。1984 年，北京语言学院（今北京语言大学）开始着手研制汉语水平考试，主要用途是考查留学生是否达到进入中国大学学习的基本语言水平。那个时候研发人员可以借鉴的同类语言测验中最具影响力的就是美国的托福考试，托福考试从 1964 年开考到本世纪初改版之前，在测验设计理念上深受结构主义语言学和行为主义心理学的影响，是典型的分立式测验，或者说是



心理测量—结构主义时期语言测验的代表作。汉语水平考试的研发在很大程度上学习和借鉴了托福考试的经验，在形式上也属于分立式的语言测验，它的主要特点有：

第一，在测验设计上，对语言技能和语言要素进行独立测量，这种思路体现了结构主义语言学对语言能力的理解。

第二，在分数报道和解释上，采取常模参照的办法，只解释考生在全体被试中的相对地位，不描述其在现实生活中的语言表现水平；报道分数时既报道合成的总分，又报道各个分测验的单项分，在分数报道上也体现了对语言技能分别测量的思想。

第三，全部采用真实语料，这是 HSK 不同于其他许多考试的特点，也是我们一直坚守的命题准则。语言测验的特点是以语言来考查语言，因而语料的真实性直接影响到测验的外部效度。测验说到底是一种取样测量的过程，是一种实验室的环境，要从测验的分数去推论考生现实生活中的语言表现，就需要保证取样的代表性和真实性。因此我们不赞成使用“人造”语料或“改造”语料。

第四，试卷的标准化和客观化程度高，测量的信度优良，分数的稳定性比较好。以 HSK（初、中等）为例，从 1988 年开考至今，试卷的信度（ $\alpha$  系数）一般都在 0.95 以上，试卷的难度（通过率）稳定在 0.53 左右，这两组数据说明 HSK 的测量精度很高，稳定性很好，在考试的质量上一点也不逊于托福考试。

语言测验的发展进入到心理语言学——社会语言学时期以后，人们对分立式考试的质疑越来越多，以托福考试为代表的分立式测验面临雅思等综合式考试的挑战。雅思考试的特点是注重语言的交际性和真实性，重视口语等主观性测验，因此具有比托福考试更高的测验表面效度和社会认可度。从 1993 年开始，为了进一步提高



测验效度，迎接雅思所带来的市场挑战。美国教育测验服务中心（ETS）着手进行了托福改版工作。新版托福已于2005年9月推向市场。新托福的推出体现了语言测验的新趋势，即注重综合性和交际性。

与托福考试一样，HSK在发展过程中也受到了测验理论发展的影响以及汉语教学发展的挑战，这就引发了我们对如何更有效地测查出考生的真实能力水平的思考。这里包含两方面的问题，第一是如何重建测验的构想，第二是如何更好地划分和解释测验分数。这两个问题的解决都归因于我们对语言能力的认识，对能力问题的关注与研究是心理语言学——社会语言学时期语言测验理论发展的标志。2007年HSK（改进版）的问世是我们对测验构想和分数等级进行思考和验证的结果，它标志着汉语水平考试告别了心理测量——结构主义时期的分立式测验，开始走向更注重交际能力的心理语言学——社会语言学时期。

改进版HSK是一种新型的考试，它不像HSK初创时期一样去模仿托福考试的形式与内容，也没有一味地追求测验的表面效度而牺牲测量的精度，而是批判性地继承和发展了原有HSK的性质与特点。从测验的形式上讲，它不是单纯的分立式测验，也不是完全的综合式测验，而采取以综合式为主，分立式为辅的形式。从考查重点来看，它注重考查语言交际能力又同时兼顾语言知识（北京语言大学汉语水平考试中心“HSK改进工作”项目组，2007）。

### 3 对汉语水平考试未来走向的思考

改进版HSK的推出不是汉语水平考试改革与发展的终点，相



反，它只是我们探索考试新理论与新方法的一个起点。要使汉语水平考试更好地服务于汉语国际教育，就要充分考虑以下几个问题。

第一，汉语水平考试（HSK）应是科学、可靠和稳定的测量工具。考试说到底是一种教育测量工具，我们使用考试的目的是为了科学、准确地测量出学习者的真实能力，在这一点上教育测量与物理测量没有本质区别。打个比方说，我们平时用秤称体重，目的显然不是为了使自己苗条，而是能够客观地测量出目前的体重水平。因而对考试来说，测量的可靠与准确是第一位的。作为一种证书考试，HSK 还有一个稳定性的问题，从历时的角度看，证书考试必须在最大程度上保证证书的等价性和获证率的稳定性，也就是说持有同等证书的人应该具有同样的语言水平，而具有同样语言水平的人应该获得同一等级的证书（不管他/她于何时何地参加考试）。这一点可以说是证书考试的生命线，它要靠考试难度的稳定性和分数等值的精确性来保证。

从考试理论发展的历史可以看出，考试制度的确立和发展离不开心理测量学的支撑，测量技术的主要作用就是估算和控制测量误差，提高测量精度。现代语言测验理论不管如何变化，也不论采用的是“以能力为中心”的测验，还是“以任务为中心”的测验，都必须时刻关注测验的测量精度。汉语水平考试（HSK）作为一项国家级考试，更应该注重测量的科学性、可靠性和稳定性，因为这是考试得以可持续发展的基本保障。

第二，制定行业标准，保证 HSK 及其他汉语考试产品的操作规范与质量水平。随着汉语教学的快速发展，各种汉语考试应运而生，考试市场的多元化格局也逐渐形成。考试是一种教育测量产



品，同其他产品一样，它的质量是需要检验并公之于众的。为了保证汉语考试的健康发展，应以 HSK 考试为基本参照，建立汉语考试的行业标准，用以指导汉语考试的研发与实施，同时使汉语考试市场更加规范和有序。这种行业标准至少要包括开发新考试的基本程序、考试质量评价体系以及向考生和社会公布的年度技术报告 (technical manual) 等。

第三，在语言能力标准的框架下开发考试，充分发挥汉语水平考试作为“标准”的尺度作用，并使其与国际接轨。二十多年来，HSK 被当做对外汉语教学的一把尺子，用以评估学生的汉语水平和教师的教学水平，无形中承担了一种“标准”的作用。新时期的汉语水平考试要继续发挥标尺的作用，进而走向国际市场，就需要在语言能力标准的框架下开发考试，实现与欧洲语言能力框架的兼容。

当代语言测验发展的显著特点之一是测验的开发在能力标准的指导下进行，能力标准用于指导语言测验可以发挥以下四个方面的作用。

首先，能力标准可以为定义语言测验的考生对象提供统一的参照标准。测验开发的第一步就是要对考生对象进行定义，根据考生的能力水平来开发合适的测验。一直以来我们在对考生对象进行定义时缺乏统一的标准，各个研发部门的定义没有可比性。在能力标准参照体系下，考生对象的定义有了规范和统一的描述，我们可以知道某个考试是为哪个能力等级的考生设计的。

其次，能力标准可以为测验内容的设计提供良好的素材，特别是对于口语考试和作文考试。能力标准对每一个等级在现实交际活动中能够做什么都给出详细的描述，这些描述将为口语和作文考试

