

中国汉语水平考试 (HSK) 研究书系

漢

HSK

中国汉语水平考试(HSK)
研究报告精选

北京语言大学汉语水平考试中心 编

主编：谢小庆

Selected Research Reports on the
Chinese Proficiency Test (HSK)
of China

北京语言大学出版社



中国汉语水平考试 (HSK) 研究书系

A Series of Studies on the Chinese Proficiency Test (HSK) of China

HSK

中国汉语水平考试(HSK)研究报告精选



Selected Research Reports on the Chinese Proficiency Test (HSK) of China

北京语言大学汉语水平考试中心 编

主编 罗小芳

(京)新登字157号

图书在版编目(CIP)数据

中国汉语水平考试(HSK)研究报告精选 / 北京语言大学汉语水平考试中心编.
谢小庆主编.

—北京: 北京语言大学出版社, 2005

(中国汉语水平考试(HSK)研究书系)

ISBN 7-5619-1456-3

I. 中…

II. ① 北… ② 谢…

III. 汉语—对外汉语教学—水平考试—研究报告

IV. H195

中国版本图书馆CIP数据核字(2005)第067077号

书 名: 中国汉语水平考试(HSK)研究报告精选

责任印制: 汪学发

出版发行: **北京语言大学出版社**

社 址: 北京海淀区学院路15号 邮政编码: 100083

网 址: www.blcup.com

电 话: 发行部 82303648/3591/3651

编辑部 82303647

读者服务部 82303653/3908

印 刷: 北京北林印刷厂

经 销: 全国新华书店

版 次: 2005年7月第1版 2005年7月第1次印刷

开 本: 710毫米×1000毫米 1/16 印张: 17.125

字 数: 237千字 印数: 1—3000

书 号: ISBN 7-5619-1456-3/H.05068

定 价: 39.00 元

凡有印装质量问题, 本社负责调换, 电话: 82303590

中国汉语水平考试 (HSK) 研究书系

第一辑

《语言测试及测量理论研究》

主编：张 凯

《中国汉语水平考试 (HSK) 研究报告精选》

主编：谢小庆

《语言测试专业硕士论文精选》

主编：孙德金

《中国汉语水平考试 (HSK) 技术报告》

主编：郭树军

《中国汉语水平考试 (HSK) 工作规范》

主编：陈 宏

A Series of Studies on the Chinese
Proficiency Test (HSK) of China

Collection One

*Studies on Language Testing and Measurement
Theories*

Zhang Kai (ed.)

*Selected Research Reports on the Chinese
Proficiency Test (HSK) of China*

Xie Xiaoqing (ed.)

*Selected Theses of Master Candidates in Language
Testing*

Sun Dejin (ed.)

*Technical Reports on the Chinese Proficiency Test
(HSK) of China*

Guo Shujun (ed.)

*The Standards and Regulations of the Chinese
Proficiency Test (HSK) of China*

Chen Hong (ed.)

中国汉语水平考试（HSK）研究书系

编辑委员会

主 任

孙德金

委 员

陈 宏 郭树军 谢小庆 张 凯

中国汉语水平考试(HSK)自1984年在北京语言学院(今北京语言大学)开始研制,已经走过了二十多年的历程。她伴随着中国改革开放的脚步,伴随着对外汉语教学事业的发展,从无到有,从小到大,成为在国内外有广泛影响的汉语水平测量工具。今天我们编辑出版“研究书系”第一辑五本著作,既是对过去的梳理和总结,更是为了开拓更加广阔的未来,因为我们知道,汉语教学和汉语水平测试在中国国力不断增强的大背景下拥有可以预见的美好前景。

对外汉语教学,或称汉语作为第二语言教学,无论是作为一项事业还是作为一门学科,相对于英语来说,历史是短暂的。而作为对外汉语教学四大环节之一的汉语测试,其开发、研究的历史就更加短了。上世纪80年代初,国内以英语为代表的第二语言教学开始由重知识转向强调语言运用能力,英语的标准化考试TOEFL对听力和阅读能力的考查对英语教学思想的改变产生了重要作用。而在对外汉语教学界,人们当时关注的只是教学中的课程测试或称成绩测试,还没有一种类似TOEFL的标准化语言能力(水平)测试,大大落后于国际第二语言教学理论和实践的发展。在当时的背景下,北京语言学院(今北京语言大学)的一些有识之士敏锐地抓住时机,提出研制汉语水平考试的设想。在教育部、学校领导的支持下,于1984年底成立“汉语水平考试设计小组”,成员有:刘珣、黄政澄、方立、孙金林、郭树军。设计小组在极其艰苦的条件下,完成了各项研制任务。在1985年第一届国际汉语教学讨论会上,刘珣先生代表研制小组作了题为“汉语水平考试的设计与试测”的报告(见《第一届国际汉语

教学讨论会论文选》,北京语言学院出版社,1986:536~542),引起轰动。1988年6月首次在北京语言学院举行正式考试。1990年HSK通过了教育部组织的专家鉴定,鉴定书认为“HSK的研究成果填补了我国汉语测试的一项空白”。1991年HSK正式推向海外。随着对外汉语教学的发展,原本以汉语预备教育(通常学习时限为一至两年)为主体的局面被打破,短期强化教学(通常为半年左右)和汉语言专业本科学历教育(通常为四年)逐渐发展起来,因此对标准化的汉语水平考试产品提出新的要求。北京语言学院汉语水平考试中心顺应形势的要求,先后完成了HSK(高等)(1993年)和HSK(基础)(1997年)的开发研制,最初研制的汉语水平考试产品改称为HSK(初、中等),三者形成了由低到高的系列考试产品,基本满足了对各类母语非汉语学习者水平测试的需求。

语言测试作为一门交叉型应用学科,以语言学、教育测量学、心理测量学等学科为基础,有着自己的研究对象、概念范畴、研究方法和学科目标。简单地说,这个学科的终极目标是要寻找到对语言能力进行准确测量的理论和方法。而由于人类对自身语言能力的认识还处在极其初级的阶段,这一领域的研究就具有了极大的挑战性。我们强调这一点,是因为社会上对汉语水平考试(HSK)的认识还存在着通俗化或称非科学化的现象,换言之,还有人不把她视为科学,认为考试不就是出出题,弄份卷子嘛!因而我们经常听到这样那样的议论。以题目的难度为例,按照标准化语言测试的基本原理,一个题目的难度,不是由教师、应试者或者其他什么人确定的,而是由该题目在一个较大样本的考生群体中的反应决定的。通俗地说,如果这个群体中的多数答对了,该题目表现为容易,反之为难。此外,对题目的区分度,对试卷的信度、效度等,标准化考试都有严格的要求,因此,评价或衡量一个考试产品的质量,关键要看该产品在上述各方面的数据是否符合科学的要求,只要符合要求,一般来说该产品用来测量语言水平(能力)时就会是准的、稳定的。

汉语水平考试(HSK)多年来一直在科学性上不断追求。围绕着“测

什么”、“怎么测”等基本问题,汉语水平考试中心开展了大量的基础研究。“测什么”讲的是通过语言和语言内容等所表现出的语言能力。要确定能够表现语言能力的语言项目、语言内容,就必须充分考虑各个方面的因素,比如语言材料的题材、体裁、语体、风格等,以及汉字、词汇量、语法点项等。这方面的研究成果主要反映在《汉语水平词汇与汉字等级大纲》(1992)和《汉语水平等级标准与语法等级大纲》(1996)中。“怎么测”讲的是测量技术和测量手段。教育测量学和心理测量学的基本原理和方法是汉语水平考试(HSK)的测试理论和技术基础。无论是题目的设计、试卷结构,还是等级划分、分数解释,都有其依据和标准。在信度、效度、题目难度、区分度,以及公平性等方面都要有理论和数据的有力支持。这些方面的研究成果主要反映在本书系的各篇文章和《中国汉语水平考试(HSK)技术报告》中。可以肯定地说,没有在科学性上的不懈追求和努力,就没有汉语水平考试(HSK)的良好信誉和日益增强的影响力。

汉语水平考试(HSK)作为标准化考试,不仅要做到命题、拼卷、预测、等值等方面的标准化,还要做到考试实施的标准化,这是我们编辑《中国汉语水平考试(HSK)工作规范》的初衷。

本书系的五本书,既各自独立,又互有联系。基本的编辑意图可以表述为:以《语言测试及测量理论研究》表明汉语水平考试(HSK)的基础理论支撑,以《中国汉语水平考试(HSK)研究报告精选》展示汉语水平考试(HSK)的研究成果,以《语言测试专业硕士论文精选》介绍北京语言大学汉语水平考试中心的教学成果,以《中国汉语水平考试(HSK)技术报告》体现汉语水平考试(HSK)的技术实力,以《中国汉语水平考试(HSK)工作规范》说明汉语水平考试(HSK)的操作保证。

汉语水平考试(HSK)作为世界语言测试园地里的一棵小树,已经枝繁叶茂,与 TOEFL 等语言测试产品一道,服务于世界各国人民语言文化的学习和交流。和任何产品一样,汉语水平考试(HSK)并非十全十美,还



有改进的空间,这已经列入我们的工作计划。而改进同样需要研究,需要建立在科学论证的基础上。社会期待着改进后的汉语水平考试(HSK),我们也将推出改进版的 HSK 的同时,奉献出我们新的研究成果。

中国汉语水平考试(HSK)研究书系

编辑委员会

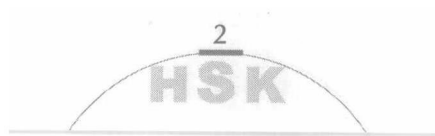
2005 年 6 月

目 录

- 1 HSK(初、中等)在变量关系控制方面的尝试 陈 宏
- 15 HSK 阅读理解试题的设计 郭树军
- 39 关于 HSK 信度的进一步研究 谢小庆
- 58 汉语水平考试(HSK)信度、效度分析报告 何 芳
- 68 关于考生团体异质程度对 HSK(初、中等)信度的影响
..... 陈 宏
- 89 语言能力自我评价的效度研究 王信旻
- 104 汉语水平考试结构效度初探 张 凯
- 123 汉语水平考试(HSK)项目内部结构效度检验 郭树军
- 151 语言能力测验的结构效度检验及其意义 陈 宏
- 169 汉语水平考试 J324 卷构想效度的验证研究 李 慧、朱军梅
- 178 汉语水平考试的等值问题 郭树军
- 197 经典等值与 IRT 等值的比较研究 谢小庆
- 217 对从 HSK 题库中计算机自动生成试卷稳定性的试验检验
..... 谢小庆 任 杰
- 226 中国少数民族考生与外国考生 HSK 成绩的公平性分析
..... 任 杰 谢小庆
- 236 中国境内外 HSK 考生成绩公平性的分析 任 杰

247 HSK 等级分数的问题 张 凯

265 后记



Contents

- 1 An exploratory study in controlling inter-variable relations
in developing HSK (elementary and intermediate)
..... Chen Hong
- 15 Basic principles of designing reading comprehension items
of HSK (elementary and intermediate) Guo Shujun
- 39 A further investigation in the reliability of HSK Xie Xiaoqing
- 58 A report on the reliability and the validity of Chinese
Proficiency Test (HSK) He Fang
- 68 The impact of heterogeneity of candidate subgroups on the
reliability of HSK (elementary and intermediate)
..... Chen Hong
- 89 A validation research in self-rating of language ability
..... Wang Jimin
- 104 Construct validation of Chinese Proficiency Test (HSK)
..... Zhang Kai
- 123 An internal construct validity investigation in Chinese
Proficiency Test (HSK) items Guo Shujun
- 151 Construct validation of language ability tests and
its implications Chen Hong
- 169 A confirmatory research in the construct validity of
J324 of Chinese Proficiency Test Li Hui & Zhu Junmei

178	The equating in Chinese Proficiency Test development	Guo Shujun
197	A comparative study of CTT vs. IRT approach to test equating of HSK	Xie Xiaoqing
217	An experimental research in the reliability of computer generated tests based on an item bank of HSK	Xie Xiaoqing & Ren Jie
226	An analytical examination of grading of HSK for DIF among Chinese minority nationality candidates and foreign candidates	Ren Jie & Xie Xiaoqing
236	A DIF investigation of grading of HSK used domestically vs. abroad	Ren Jie
247	Issues of grading of HSK	Zhang Kai
265	Postscript	



汉

HSK(初、中等)

在变量关系控制方面的尝试*

陈 宏

■ **内容提要:** 本研究的目的在于探讨能否通过相对提高每个题目与其所在分测验的相关来控制变量关系,并借此改善 HSK 的判别效度。这种尝试主要包括两个主要环节。第一个环节是在 HSK 原有的题目分析系统中增加了项目内部结构效度检验步骤,并相应地在题目特征指标中增加了一个新的指标:“最低相关差”。第二个环节就是修订题目筛选标准,并在拼卷时优先选择那些经过项目内部结构效度检验最低相关差取值高的题目。

■ **关键词:** 汉语水平考试 判别效度 最低相关差

Abstract: In the past years, some empirical studies of the discriminative validity of the Chinese Proficiency Test (HSK) revealed that many items more highly correlate with other subtest(s) than the one they belong to. This has even led researchers to question the measurability of some variables and the theoretical basis for the construct definition of the test. The present study is the first attempt to improve the discriminative validity of HSK. It is realized by the following two main procedures. The first is to add an approach verifying the item-interior construct validity to the existing item analysis system of HSK. Accordingly, a new param-

* 原载北京语言大学汉语水平考试中心编《汉语水平考试研究文集》,经济科学出版社,2000年5月。

ter named "the lowest correlated difference" is added to the item property parameters. The second is to modify the item selecting criterions. Under the direction of new criterions, the items that have high coefficients on "the lowest correlated difference" have the priority to be selected.

Key words: HSK, discriminative validity, the lowest correlated difference

零 引 言

HSK 结构效度研究方面的发轫之作是张凯先生的《汉语水平考试结构效度初探》一文,张凯先生在文中报告并讨论了对 HSK 的三份正式试卷进行因素分析的结果。他认为:就 HSK 的最初假设,即每个分测验测量一种特质或语言能力而言,他所得到的 HSK 的“数据对理论的拟合程度都还不十分理想”(张凯,1992)。

在张凯的研究之后,郭树军(1995)采用 Henning 介绍的检验项目与分测验相关关系的方法——内部结构效度法,对 HSK 的结构效度进行了更深入的探讨。这种方法的基本假设是,“若一个项目与自己所在的分测验的点二列相关系数高于与其他分测验的点二列相关系数,则此项目具有结构效度;反之,一个项目与本分测验的相关低于它与其他分测验的相关,则此项目缺乏内部结构效度”(G. Henning, 1987)。郭树军发现,总的来看,“HSK 各试卷各分测验中均有一定数量的题目缺乏内部结构效度”。鉴于问题主要表现在相当一部分题目要么与某一个分测验的相关高于跟本分测验的相关,要么跟两个甚至三个分测验的相关均高于跟本测验的相关,郭树军认为,“HSK 的项目有比较复杂的内部结构交叉现象”(郭树军,1995)。

作为相互印证,张凯和郭树军的研究的一个共同点,就在于它们都直接诉诸语言能力测验中一个容易为人疏忽的问题,即语言能力测验的设计和开发者必须为他们对语言能力的定义提供判别效度方面的证据。我

们知道,结构效度实际包含了聚敛与判别这样一对既矛盾又统一的两方面的要求。聚敛通过高度概括的方式反映语言能力,是测验在可靠性与同质性(或内部一致性)方面的量化特征,主要体现对测量工具本身质量的考虑;而判别是测验在目的性方面的量化特征,更多牵涉测验用途,偏重对语言能力内部结构和结构关系进行分类和区别定义。所以,在需要提供的各种结构效度的证据中,聚敛主要是一个对结构效度所包含的信度方面的要求。然而,语言能力测验对其宣称要测的语言能力或结构的实际测量的有效性同时需要各种能力或结构间彼此判别的证据。相对于测验的目的和用途而言,这方面的证据也许是更重要和更有实质性的。因为“如果我们要表明我们的测验分数是某种特定的语言能力的有效指标,我们就必须证明它们不仅与同一种能力的其他指标相关,而且与其他能力的统计量无关。”(Bachman, 1990) HSK 各分测验分别测量不同的语言行为,但我们却不能要求他们彼此毫不相关,因为这些语言行为通常被认为是彼此相关的,也因为 HSK 对所选语言行为样本的真实性要求很高,实际牵涉的语言能力要素很复杂,从而被试的应试行为之间必然彼此相关。然而问题在于,HSK 的一些设计和开发人员担心“这种相关如果超过了一定的限度,我们就无法从测验的分数中确认和分解出我们要测量的那种能力,就没有理由说分数有效地反映了测验假设的那种结构。”(陈宏,1997) 郭树军和张凯的研究结果说明这种担心并不是毫无理由的。

判别效度问题的提出说明 HSK 结构效度的建立可能主要面临两个相互连带的问题,一是 HSK (初、中等) 对它要测量的结构和结构关系的操作性定义甚至 HSK 有关语言能力结构的理论假设本身就有待完善,一是 HSK 未能充分实现其既定的操作性定义,即 HSK 试卷生成的某些环节还有缺陷。本文研究的问题是,在充分实现 HSK 现有的操作性定义方面,在技术许可的范围内,我们需要和能够做些什么,以及这样做是否有助于有效地测到我们打算测量的语言能力或结构。