



汉语水平考试 研究 (续集)

刘英林 主编



现代出版社

序

语言水平测试的制定是一项非常复杂的工作，具有相当重要的理论和实践指导意义。国外对语言水平测试的研究已有几十年的历史，积累了相当丰富的经验，像美国测试留学生英语水平的 TOEFL（托福）和测试研究生入学水平的 GRE 都是有世界性影响的权威性语言水平测试。我国这方面的研究起步较晚，但进展相当迅速。1984 年北京语言学院开始组织人力研制汉语水平考试（HSK），成立了汉语水平考试中心，1990 年汉语水平考试通过专家鉴定，1992 年正式成为中华人民共和国国家级考试，前后一共不过八年的时间。这八年间，汉语水平考试中心在刘英林教授的领导下，借鉴国外经验，收集大量资料，进行多次试验，完成了这项极其有意义的工作，他们这种勇于开创和刻苦钻研的精神是很值得钦佩的。目前，已有 110 个国家和地区的两万多外国朋友参加过汉语水平考试。HSK 已经走向世界，开始在全世界树立起它的权威性。

汉语水平测试是以当前对外汉语教学实践为基础制定的，因此符合客观的实际情况；但是它又高于当前对外汉语教学，因此能对当前对外汉语教学有一定的指导意义。这后一点也很重要，以汉语水平考试为依据的“词汇等级大纲”为例，这是汉语水平考试中心的研究人员根据科学的统计和对外汉语教学的需要编制出来的，把应该选收的 8000 个词按难易分为四级，这四级词已经成为当前编写对外汉语教材选词的指导，彻底改变了过去编教材主要凭主观印象选词的局面。

汉语水平考试中心的研究人员在制定汉语水平考试的过程中边实践，边总结，写出了一批很有特色的论文，这些论文不只是对语言水平测试做出了贡献，对对外汉语教学的总体设计和教材编写等方面也有很重要的参考价值。1989 年曾由刘英林教授主

编，选出其中十几篇编成《汉语水平考试研究》一书出版，受到了国内外语言教学工作者的普遍重视，并获得北京市第二届哲学社会科学优秀成果二等奖。现在，刘英林教授又抱病编成这本《续集》，从收入《续集》的论文可以看出，随着四年来研究的深入，无论在理论方面还是在实践方面都有了新的发展，所收入的十篇国内外用户报告可以说是对 HSK 最客观、最实际的评价，对今后 HSK 的进一步完善将是有力的促进。

汉语水平考试虽然已经取得如此令人瞩目的成绩，但还远远没有达到完善的程度，难度较大的口语测试和高等汉语水平测试还有待于研制。《续集》收入了几篇专门讨论高等汉语水平考试的文章，还有讨论计算机辅助自适应汉语水平考试的文章，说明汉语水平考试中心的研究人员正在向新的高度攀越，相信他们一定能够成功，为汉语水平考试做出更大的贡献。

林 焘

1994 年 1 月
于北京大学燕南园

目 录

序 林 燕

论 文

汉语水平考试 (HSK) 述略 刘英林 (1)

汉语水平考试 (HSK) 研究的新进展

..... 刘英林 郭树军 (12)

关于“汉语水平考试”的研究报告 刘英林执笔 (32)

汉语水平考试 (HSK) 概观 刘英林 (45)

汉语水平考试的等值问题 郭树军 (51)

汉语水平考试的分数体系 谢小庆 (72)

汉语水平考试结构效度初探..... 张 凯 (91)

汉语水平考试的信度、效度分析报告 何 芳 (106)

汉语水平考试 (HSK) 的理论基础探讨 刘英林 (117)

高等汉语水平考试的总体设计与理论思考..... 刘英林 (137)

高等汉语水平考试试测结果的统计分析

和对课程设置的评估..... 刘健力 李 明 (154)

HSK 题库建设综述 李 航 (166)

计算机辅助自适应汉语水平考试的设想..... 谢小庆 (184)

中国汉语水平考试的若干问题..... 刘英林 (195)

汉语常用字词的统计与分级 刘英林 宋绍周 (214)

汉语水平考试 (HSK) 与对外汉语教学 刘英林 (231)

中高级阶段对外汉语教学的理论探讨·····	刘英林 (244)
对外汉语教学论的几个问题·····	刘英林 (258)
教学与考试八则·····	刘英林 (265)
对外基础汉语教学法创新之路 ·····	刘英林 李景蕙 (274)

对外汉语教学的一项基本建设 ·····	程 棠 (295)
参照 HSK 改进课堂训练 ·····	北京语言学院 李 杨 (306)
北京大学 HSK 入学考试分班入系分析	

·····	北京大学 杨德峰 (318)
-------	----------------

华东师范大学留学生 HSK 考试评述	
·····	华东师大 吴永毅 徐子亮 (332)

北京语言学院 HSK 入学测试编班分析	
·····	北京语言学院 贾永芬 方玲 (352)

HSK 对对外汉语教学的重要指导作用	
·····	大连外院 徐甲申 (365)

建造以 HSK 为参照的教学链 ·····	天津大学 耿二岭 (376)
-----------------------	----------------

从 HSK 摸底分析看新疆民族中学的汉语教学	
·····	中央民族学院 赵学会 (388)

少数民族汉语水平考试探讨 ·····	延边大学 刘明章等 (399)
--------------------	-----------------

1993 年 HSK (高等) 试测结果统计与效度分析	
·····	北京语言学院 倪明亮 (406)

关于在美国举行 HSK 试考的分析报告	
·····	(中国) 潘兆明 (美国) 施旭东 (412)

考试办法和考试大纲

中国汉语水平考试 (HSK) 办法 ·····	(425)
高等汉语水平考试[HSK (高等)]考试大纲 ·····	(428)

HSK 鉴定书及评介

- 中国汉语水平考试 (HSK) 鉴定书 (434)
- 汉语水平考试通过鉴定 人民日报 (海外版) (439)
- HSK —— 汉语水平考试见闻
..... 人民日报记者 曹焕荣 (440)
- HSK, 中国特色的汉语水平考试
..... 中国教育报记者 于建坤 (442)
- 为 HSK 考试叫好 人民日报 徐兆荣 (445)
- HSK 走向世界 中央人民广播电台记者 李向明 (446)
- 汉语“托福”——进入中国之门 光明日报记者 唐 旬 (449)
- 后记 (452)

汉语水平考试(HSK)述略*

刘 英 林

提要 本文从七个方面介绍汉语水平考试(简称 HSK), 即: 基本性质和用途, 基本路子和格局, 命题依据和范围, 命题标准和原则, 信度和效度, 难度参照点和稳定性, 测试等级量表和常模。

1. HSK 的基本性质和用途

1.1 HSK 是一种水平考试, 又带有学能考试的某种特征。所谓水平考试, 主要是测量测试对象是否达到某种水平, 它着眼于现在。所谓学能考试, 则主要用来测试测试对象完成某种任务的能力倾向, 它着眼于将来。HSK 的主要目的是测量考生的汉语水平能否适应在中国学习、生活和交际的需要, 它明确划定了初等水平和中等水平等各级测试标准。

1.2 HSK 是目标(尺度)参照考试和常模参照考试相结合的产物。近代测试专家把标准化考试分为“目标(尺度)参照考试”和“常模参照考试”两种形态。一般的目标参照考试的主要着眼点是鉴别考生考试成绩是否达到了“规定的标准”, 以保证今后学习或工作的延续性、适应性。一般的常模参照考试的主要着眼点是确定每一考生的考试成绩在全体考生中的“相对的位置”, 以

* 原载《中国语文》1990年第4期。

保证选拔人才的合理性、公平性。面对对外汉语教学的现实，我们试图在“目标”和“常模”之间开辟一个相对独立的具有中介性质的标准化考试层次。HSK 的基本着眼点是测量考生的水平是否达到规定的等级标准，达到了哪一级标准，同时又能较好地区分每一考生在全体考生中的“相对的位置”，因而可以达到一考多用的目的。

1.3 HSK 是一种能力与速度相结合的考试。HSK 既要考查考生的答题能力又要考查考生的答题速度，它的着眼点是在规定的时间内准确地界定测试对象的汉语能力标准。HSK 界定的能力标准可以划分为以下五种类型：1) 界定国外非零起点的汉语学习者来华继续学习汉语时分班的能力标准。2) 界定外国留学生入我国高等院校学习专业所必须具备的汉语能力标准。3) 界定中国对外汉语教学（包括一年级和二年级）结业时应具备的能力标准。4) 界定有一定汉语水平的外国人和华侨子女获取《汉语水平证书》的能力标准。5) 界定国内少数民族学生入高校学习专业或获取《汉语水平证书》的能力标准。

答题速度的快与慢本身就是一种能力的体现。HSK 每一道题的平均答题时间为：听力理解，15——20 秒；语法结构，40 秒；阅读理解，80 秒；综合填空，45 秒。

2. HSK 的基本路子和格局

1984 年 12 月北京语言学院受国家教委委托开始研制 HSK。HSK 的基本路子和格局的主要特点是，以听力理解和阅读理解测试为主线，以语言能力测试为核心，以语言知识（知识的运用——本书编者注）测试为基础，以标准化试题形式为基本方向。

2.1 HSK 以听力理解和阅读理解测试为主线。HSK 最初的

设计思想之一，是以衡量应试者是否具备在中国高等院校入系学习和在中国生活的汉语能力为主要目标的。HSK 的基本路子和格局必须服从于入系学习、日常生活对汉语能力的基本要求。因此，听力理解和阅读理解是测试重点，对这两项技能（能力）的测试题共占 58.8%。

2.2 HSK 以语言知识（知识的运用——本书编者注）测试为基础，语言能力测试为核心。在研制过程中，我们始终强调 HSK 既要能有效地测试应试者的实际汉语水平，又要能对对外汉语教学发挥积极的反馈作用。基于此，我们将听力理解、阅读理解、综合填空等语言运用能力作为测试的核心，共占整个试题的 82%，同时也给语言知识（知识的运用——本书编者注）测试以一席之地，占 18%。

为更好地贯彻以语言能力测试为核心的思想，我们在选取语料时，一方面注意语言的真实性和常用性，以保障测试的内容都是应试者在日常交往、学习、生活和工作等特定环境中经常遇到、需要掌握的真实、自然、常用的语言；另一方面，又要注意测试内容的广泛性和代表性。

2.3 HSK 以标准化试题形式为基本方向。所谓标准化试题形式，即是应试者根据考试内容所提供的信息和现成答案，经过自己的分析、理解、判断，进行填空、选择正确答案等，以固定的形式回答主试者的问题。HSK 试卷共有 170 题，其中 1——154 题为多项选择题（均为四个选择项），属客观性试题，占 90%。155——170 题为汉字填空题，是半主观半客观性试题，占 10%。

3. HSK 的命题依据和范围

HSK 命题的主要依据有两个：一个是《汉语水平等级标准

和等级大纲》(简称《标准和大纲》),这是中国对外汉语教学学会编制、国家对外汉语教学领导小组办公室审订后在全国试行的一份重要的指导性文件,具有一定的权威性,是 HSK 命题的根本依据。另一个是《汉语水平考试大纲》,这是国家对外汉语教学领导小组办公室汉语水平考试部编制、并在全中国试行的另一份重要文件,它既是 HSK 命题的具体依据,又是考生备考的指导用书。

3.1 《标准和大纲》是 HSK 的主要命题依据。制订《标准和大纲》的主要目的,是在我国对外汉语教学体系中明确规定一个统一的教学终点,并逐步建立起一个科学的、规范的、可普遍接受的量化标准和等级。这个统一的教学终点和统一的量化标准和等级也是国外汉语学习者分别达到初等、中等和高等汉语水平的学习内容和达标要求(高等汉语水平即四、五级的学习内容正在研究制定中,《标准和大纲》暂缺此项内容)。

3.2 《汉语水平考试大纲》是 HSK 的具体考试依据。对整个考试起宏观导向作用,并指导 HSK 考试的具体设计和实施。考试大纲规定,HSK 共有四项内容:

听力理解,既有表层理解又有深层理解,主要测试考生能否听懂正常语速的句子、对话和一般性题材的讲话。1) 了解所听到的句子和简短对话的大意。2) 跳跃障碍,抓住其中的主要信息或某些重要细节。3) 根据听到的材料进行推理和判断。4) 理解说话人的目的和态度。

语法结构,既有判断指定词语的位置,又有语法结构的选择,主要测试考生对汉语普通话语法结构的掌握程度。1) 常见的量词、方位词、能愿动词、副词、介词、连词、助词等的用法。2) 动词、形容词和名词重叠。3) 几种重要补语、定语、状语的用法。4) 语序。5) 比较的方式。6) 提问的方式。7) 常用词组和习用语。8) 常用复句。

阅读理解，既有精读与复读，又有略读与跳读，主要测试考生的阅读能力和速度。1) 掌握所读材料的主要用意和大意。2) 了解所读材料的主要事实和信息。3) 跳跃障碍，捕捉所需的某些细节。4) 根据所读材料进行引申和推断。5) 领会作者的态度和情绪。

综合填空，既有综合性词语填空，又有综合性汉字填空，主要测试考生依据上下文综合运用汉语的能力和在综合运用汉语能力的基础上书写常用汉字的能力。

4. HSK 的命题标准和原则

4.1 评价试题质量的标准。评价 HSK 试题质量有三条基本标准：第一，能否准确地界定考生的实际汉语水平；第二，能否保持年度与年度试卷以及试卷内部项与项之间的难度稳定性和均衡性；第三，能否引导对外汉语教学向好的方向发展，促进整体教学质量的提高。要达到上述标准，命题时既要体现标准化考试的共性，同时又注意体现中国特色、汉语特色。

4.2 命题的总原则。

HSK 的命题原则是把标准化考试的理论与对外汉语教学的实践和汉语的特色相结合。根据汉语语音、语调和词汇等特点设计听力理解和阅读理解中的相关试题，针对汉语语法和语序的特点安排语法结构试题，围绕汉语书写的特点构思汉字填空试题，并依照文化、心理的特点选择和设置必要的“语言文化”要素。这种“语言文化”要素体现在以下三个方面：1) “交际文化”，如中国特有的称呼、问候语。2) “语感文化”，指两种语言中不能直接对译的那一部分表示本民族特有事物和现象的词语。3) “语境文化”，指在特定的语境中方可正确理解的词语或语言现象。

HSK 试卷字词范围的限制性原则。一般说来，HSK 的词汇

范围要限制在《HSK 常用词汇一览表》之内，即在对外汉语教学的甲、乙、丙三级共 5168 个词汇范围内。在部分较难和最难的试题中，词汇范围可以适当超出，但是，超出幅度不得大于 5%。HSK 试卷卷面和听力录音的总字数也有严格限制，前者一般为 6000——6400 字左右，后者一般为 1800——2200 字上下。整个 HSK 考试试题的总字数限制在 7800——8600 字之间（包括试卷中的阿拉伯数字）。

HSK 试卷语法范围的规定性原则。一般说来，HSK 试卷的语法范围要在《语法等级大纲》规定的限度之内，即在全国统一的甲、乙、丙三级共 589 项、点的语法范围之内，甲级基本语法共 133 项，这是对外汉语教学语法的主体部分，是一个比较完整的体系，它覆盖面大，使用度高，关系到全局，是 HSK 要测试的重点。乙、丙级语法合计为 449 点。乙级语法和丙级语法是对甲级语法的补充、扩展和深化，其中大部分是实用意义较强的常用词语及固定格式。在命题时，部分较难和最难的试题中的语法范围（特别是语法点）可以适当超出《语法等级大纲》，但超出幅度也不大于 5%，即 30 项、点。

HSK 试题语料的广泛性和多样性原则。HSK 语料选自现今中国国内几十种文字质量较高的报纸杂志。体裁包括新闻报道、小说、散文、政论、人物介绍等等，题材涉及中国当代政治、经济、社会、文化、史地、科普等多种领域。语料的广泛性和多样性，保证了 HSK 的信度。

HSK 试题内容的普遍性、公平性原则。即试题内容所涉及的知识必须是一般性的而非专业性的。试题材料的普遍性和公平性，可以增加 HSK 的效度。

HSK 试题内容的新颖性、趣味性原则。试题语料的知识内容要有新意，内容要么属于有趣的新常识，要么是从新的角度阐发自己独特的见解。

HSK 语料容量的多层次、多信息原则。这种多层次、多信息的语料，或者是根据时间先后叙述的相关事件，或者是对某一事件提供多方面的细节，或者是比较两个或两个以上的事件、人物或论点。

HSK 试题干扰项的平衡性、迷惑性原则。HSK 各项试题选择项（包括正确答案和干扰答案）都是精心编制的。正确答案有一定的隐蔽性，干扰答案（分心答案）设在考生最易出偏差的环节，具有一定的迷惑作用。

HSK 试题答案的单一性、简洁性原则。作为多项选择试题，要求正确答案是唯一的，不能有两个或两个以上的正确答案，或者说，同一试题中，不应出现两个或两个以上的可接受的选择项（或汉字）。选择项的语句的长短（字数）应尽可能保持一致；如果因命题需要不能这样，也要按字数多少形成阶梯式的等级。

5. HSK 的信度和效度

5.1 HSK 试题具有题量大、内容广博的特征，加之评分客观，因此使 HSK 的信度得到了有力的保证。1986——1988 年三年三套 HSK 试卷的内部一致性信度均在 0.95 以上，完全符合标准化考试的要求。

5.2 HSK 具有较高的效度，是一种有效的考试。目前使用三种方法检验 HSK 的效度：

1) 理论（观念）效度。当前，我国对外汉语教学正在形成语言结构与语言功能相结合，注意语言文化渗透的教学法体系。这一教学法体系已成为国内基础汉语教学的主导教学法。HSK 的指导思想、基本路子、测试重点、命题原则和特色，与上述教学法理论体系是完全相符合的。2) 内容效度。为确保 HSK 的

内容效度，我们严格按照全国统一的《标准和大纲》所规定的词汇内容和语法内容进行命题，对听、读、写等不同的语言技能进行考查，可以比较准确、有效地测量考生的实际语言水平。3) 预测效度。1985年和1987年，两次对参加过HSK考试的考生进入专业院校后的学习情况进行了追踪调查，调查结果表明，HSK基本上反映了考生的真实水平。北京语言学院、北京大学等主要使用单位也普遍认为HSK对考生有较好的区分能力。

5.3 建立预测制度是保证信度、效度的关键措施。为从预测中获得真实可靠的数据，必须有以下四个方面的保证。

1) 预测方法的科学性。主要表现在两个方面，一是所选预测观察点的稳定性，即这种观察点教学历史较长，教学质量稳固、学生来源稳定。二是分散预测的固定性，即每年定期在这些观察点上用基本相同的预测试卷，在相对稳定的系级进行试测。2) 预测对象的代表性。主要体现在两个同质性上：一是注意命题试测的同质性，即参加分数预测的考生与将来参加实测考生水平的同质性。二是分散预测中的试题与将来实测中的试题内容与难度的同质性。3) 预测条件的一致性。即预测条件与正式考试(实测)条件应基本保持一致，但是预测时，在考试时间上可适当放宽，让绝大多数考生都能作完绝大多数试题，以便获得较完整的数据和资料。4) 预测内容的保密性。除了规定必要的保密纪律外，还要做到：预测题目要尽量多；预测要分散在不同班级同时进行；预测最好提前一年进行。

1988年度试卷中中等难易度(0.3—0.7)的试题有145题，占总题数的85%；1986—1988年度三套试题中，85%的试题的区分度指数均在0.3以上。这就使HSK具有较好的区分能力，可以保证HSK的信度、效度及难易度的继承性。

6. HSK 的难度参照点和稳定性

6.1 HSK 难度参照点的“二元化”。HSK 主要为两类考生建立两个较为稳定的参照点。

一类考生是在中国正规学习汉语 800 学时左右的留学生，程度相当于目前国内对外基础汉语教学一年级的中上水平。另一类是正规学习汉语 1600 学时左右的留学生，程度相当于目前国内对外基础汉语教学二年级的中上水平。HSK 这种参照点的二元化特点，必然导致命题的“二元化”倾向，即命题中一类试题以“类似”一年级结业时的中上水平为假设目标，另一类试题以“类似”二年级结业时的中上水平为期望目标。

6.2 HSK 难度的稳定性和继承性。水平考试的总体难易度（平均难易度）最好在 0.50——0.60 之间。1986——1988 年三套试卷的难易度平均值都在 0.55——0.57 之间，说明 HSK 基本上是一种难度稳定的考试。

1986——1988 年一、二年级预测难度均值均为 0.53（一、二年级学生比例为 1:1，下同），一、二年级预测难度均值都在 0.53——0.57 之间，两者之间有比较稳固的数量递增关系，这说明我们建立预测制度并尽力保持 HSK 难度水平一致性和继承性的尝试是成功的。

7. HSK 的测试等级量表和常模

7.1 建立以原始分数为标记的等级分数。1988 年，初步界定了对外汉语教学初等和中等水平的 HSK 水平等级量表，其中包括一、二年级留学生学习结业标准、非零起点的新生编班标准、留学生入学（入系）标准和颁发 HSK 证书的标准。

HSK 水平等级量表，将基础阶段和基础后阶段的汉语水平

划分为两等 8 个级别，其中，0 等为 1—2 级，初等为 3—5 级，中等为 6—8 级。一年级最优秀学生的汉语水平可望达到 5 级水平，二年级最优秀的学生能够达到 8 级水平。两等水平的每一“等”都有一个界标，两个界标是依据几年来的实践结果决定的，因此，每一界标实际上均隐含着—个较为稳定的常模基础。

7.2 建立 HSK 常模。HSK 是一种特殊的标准化考试，因此，实现成绩评定的科学化，既要统一界定初等和中等水平测试等级标准，又要进一步解决并控制分数等值的问题。解决分数等值的问题，可采用经典测试理论，也可研究使用项目反应理论。目前，我们首先尝试运用经典测试理论建立 HSK 常模。

根据 HSK 考试规模小、测试用途多、应试者国别众多、程度复杂多变等特点，在建立统一的比较标准时，必须首先严格筛选和建立独特的标准样组。这种样组必须按照一定的抽样原则和方法，从考生总体中抽取具有代表性和稳定性的考生样本。要保证标准样组的代表性和稳定性，必须选好具有稳定的考生样本的抽样单位，必须保证适当的抽样人数和—、二年级两个层次考生在样组中的一定比重，还必须认真考虑国别的因素，这些都是由 HSK 的特点决定的。经 30 次计算机自动随机抽样结果显示，样本在构成上基本上是稳定的，各种标准样组值也是较为稳定的（如原始平均分均在 90—91 之内，原始标准差均在 38—39 之间）。

HSK 各种标准值的稳定性为建立 HSK 常模、为把原始分数转换成可比性更强的 HSK 分数以及以 HSK 分数作为标志的等级分数奠定了良好的基础。为了给考生和用户提供多层次的信息，满足多方面的需要，HSK 使用一种特有的分数体系，评定三种分数——原始分数，HSK 分数和 HSK 等级分数。HSK 分数（目前，总分也为 170）是由标准分数 Z 分数转换成的导出分数。由于 HSK 试卷基本上是稳定的，难度的浮动是有限的，因

此 HSK 分数与原始分数实际上有着比较稳定的对应关系。

从 1990 年开始，将按照以 HSK 分数为标志的等级分数量表，由国家对外汉语教学领导小组办公室统一颁发 HSK 证书（初等水平证书 ABC 三级，中等水平证书 ABC 三级）。HSK 证书既可作为外国留学生直接进入中国高等院校学习所要求的实际汉语水平的一种证明，又可作为聘用机构录用汉语人员的重要参考证明。

1990 年 2 月 20 日，HSK 通过了国家教育委员会和国家对外汉语教学领导小组联合主持的鉴定。这里引用专家鉴定意见中的一段话作为本文的结束：“HSK 研究已初步形成了具有自己特点的系统，达到了一定的标准化和自动化水平；HSK 的研究成果填补了我国汉语测试的一项空白，是我国对外汉语教学领域中一项大的突破，对于我国语文考试标准化具有借鉴作用，在国内同类研究中处于先进水平。”