

# Design Principles of New HSK 3.0 and Their Implications for Chinese as a Foreign Language Education

## 新汉语水平考试 (HSK 3.0) 的设计原则及其对外汉语教学启示

Sukma Tajuddin<sup>1\*</sup>; Sukma<sup>2</sup>; Leni Cahyati<sup>3</sup>

<sup>1</sup>Liberal of Arts, Hunan Normal University, Changsha, China

<sup>2,3</sup>Faculty of Cultural Sciences, Hasanuddin University, Makassar, Indonesia

Email: [sukmamengtian@hunnu.edu.cn](mailto:sukmamengtian@hunnu.edu.cn); [sukma@unhas.ac.id](mailto:sukma@unhas.ac.id); [lenichyt@unhas.ac.id](mailto:lenichyt@unhas.ac.id)

Received : 06 February 2026

Accepted : 28 March 2026

Published : 31 March 2026

Copyright © 2026 by Author (s)

This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

<https://creativecommons.org/licenses/by/4.0>



Open Access

### Abstract

The reform of the Chinese Proficiency Test under the New HSK 3.0 framework marks a significant shift in the assessment of Chinese as a Foreign Language (CFL). While previous studies primarily examine vocabulary expansion or standard alignment, the design principles underlying the reform remain under-explored. This study investigates the conceptual logic of the New HSK 3.0 examination system and its implications for CFL education. Using qualitative document analysis and multi-source literature review, the research analyzes key policy documents, including the Chinese Proficiency Grading Standards for International Chinese Language Education (CEFCIL) and the HSK Examination Syllabus (2025). The findings identify three core design principles: multidimensional competence construction, communicative authenticity, and receptive-productive differentiation in character learning. These principles are reflected in the expansion to a nine-level proficiency framework and the redistribution of vocabulary and character requirements across learning stages. The study provides insights into Chinese language assessment reform and its implications for curriculum design and language teaching in international Chinese education.

### Keywords

Chinese Proficiency Testing; Communicative Competence; CFL Education; Language Assessment Reform; New HSK 3.0

### 引言

语言水平测试不仅在衡量学习者语言能力方面发挥关键作用，同时也深刻影响课程设置、教学实践与学习策略。这一现象即学界广泛讨论的反拨效应，已在语言测试研究中得到大量实证支持(Cheng, 2004)在对外汉语教育领域，汉语水平考试(HSK)(Peng et al., 2021)已成为全球最具影响力的标准化测评工具之一。随着全球汉语学习者规模持续

扩大, HSK 日益成为衡量汉语水平、引领课程建设的重要参照标准。已有研究表明, 语言测试会对教与学产生正向与负向双重反拨效应 (Hu, 2025; Kong, 2024)。 (Zhao, 2025)在对 HSK 反拨效应研究的系统综述中指出, HSK 对学习者和教师均产生可观测影响, 其中对学习者的学习动机、学习策略与目标导向的作用更为显著。从宏观层面看, 应用语言学与交际语言教学理论的发展推动了全球语言测评实践的重大转型 (McNamara, T., & Roever, 2006)。《欧洲语言共同参考框架》(CEFR) 的广泛推行, 促使各国语言测评体系逐步突破传统离散式语言知识点考查, 转向更为综合的能力本位测评模式 (Council of Europe, 2020)。

汉语水平考试自 1984 年设立以来历经数次重大改革。2009 版 HSK 采用六级等级框架, 考查听、读、写三项技能。该体系虽具备较高内部信度, 但已有研究揭示其在构念效度方面存在若干短板。例如, (Peng et al., 2021)基于 (Bachman, L. F., & Palmer, 2010)提出的测评使用论证 (AUA) 框架分析发现, HSK 五级写作中的句子排序题型与阅读能力的相关性显著高于写作能力。此外, 现行考试未设置强制性口语测评, 也因难以充分体现交际能力构念而备受批评。这些结构性局限表明, 旧版 HSK 偏重语法知识与词汇识别, 对语言运用中的语用能力、互动能力与跨文化维度体现不足 (Hanban / 孔子学院总部, 2010; Jin, 2017)。

为弥补上述不足, 教育部中外语言交流合作中心于 2021 年发布《国际中文教育中文水平等级标准》(GF0025-2021, 简称中文国标), 作为汉语能力描述的新框架。以此为依据, 新版 HSK 考试大纲于 2025 年 11 月正式公布, 计划于 2026 年 7 月在全球实施。新体系将等级由六级扩展至九级, 分为初等 (1-3 级)、中等 (4-6 级)、高等 (7-9 级) 三个阶段, 并同步设立笔试、口试 (PTH)、院校中文水平认证 (CHBC) 三条测评路径。

相较于其他语言测评体系或旧版 HSK, 本研究之所以聚焦 HSK 3.0, 主要基于以下三方面的原因。首先, 从影响范围来看, HSK 作为全球最具规模的汉语水平测试, 其改革将直接作用于庞大的国际学习者群体, 并对课程设置与教学实践产生广泛而深远的制度性影响, 因此具有显著的研究代表性与现实意义。其次, 从改革性质来看, HSK 3.0 并非对既有体系的局部修订, 而是在等级结构 (由六级扩展至九级)、能力描述框架以及测评路径等方面进行系统性重构, 体现出从以语言知识为中心向以交际能力为导向的范式转变, 这为检验当代语言测试理论在具体测评体系中的应用提供了重要契机。再次, 从国际

对接视角看,新体系以中文国标为基础,并尝试与 CEFR 等国际语言能力框架建立参照关系,使 HSK 在全球语言测评体系中的可比性与解释力进一步提升。然而,尽管 HSK 3.0 在规模、结构与理论导向方面均具有重要研究价值,当前研究仍较少从整体层面系统探讨其底层设计原则及其对对外汉语课程政策与教学实践的综合影响。正如 (Xinsheng, 2020) 所言,当前仍缺乏将测评设计逻辑与教育影响相结合的整合性定性研究。

为弥补这一研究缺口,本研究聚焦新 HSK 3.0 改革的理论基础,解析新版考试框架所蕴含的设计原则,重点探讨这些原则如何呼应当代语言测试理论发展,以及其可能对对外汉语课程政策与教学实践产生的影响。

基于上述研究目标,本文提出以下研究问题:

研究问题 1: 从当代语言测试理论视角看,新 HSK 3.0 构念的核心设计原则是什么?

研究问题 2: 这些设计原则对对外汉语教育的课程政策与教学实践有哪些启示?

本研究贡献主要体现在三方面:第一,理论层面:整合 Bachman & Palmer 的交际语言能力模型、Young 的互动能力框架与 Nation 的接受-产出不对称理论,为解读新 HSK 3.0 设计逻辑提供多维理论视角。第二,分析层面:整合前人研究成果,将测评使用论证框架(Peng et al., 2021)、中文国标与 CEFR 比较研究 (Shi, 2025) 反拨效应研究 (Zhao, 2025) 及词汇等级研究 (Li, 2023) 纳入统一分析框架。第三,政策与实践层面:将新 HSK 3.0 设计原则转化为可操作的对外汉语课程开发与课堂教学建议。

## 文献综述

### 1. 语言能力测试理论:从语言能力到交际能力

本研究的理论基础建立于三个相互补充的概念框架之上:交际语言能力理论、互动能力框架以及接受性—产出性能力发展理论。这三个理论视角共同反映了当代语言测试研究从结构主义语言能力测量向多维交际能力评估转型的发展趋势。首先,(Bachman, L. F., & Palmer, 2010) 提出的交际语言能力 (communicative language ability, CLA) 模型为现代语言测试提供了重要理论基础。该模型将语言能力划分为组织性知识与语用知识两个核心维度,并强调策略性能力在语言使用过程中的调节作用。其中,组织性知识包括语法知识与语篇知识,而语用知识

则涵盖功能性知识与社会语言学知识。这一理论框架突破了传统语言测试仅关注语法与词汇掌握的局限，强调语言能力应当被理解为在具体社会情境中有效使用语言的综合能力。

(Shi, 2025)指出，当前具有国际影响力的语言能力标准体系，如CEFR与《国际中文教育中文水平等级标准》(CEFCIL)，均在不同程度上建立在交际语言能力理论基础之上。其次，(Young, 2011)提出的互动能力框架进一步扩展了交际语言能力理论的分析视角。与传统语言能力模型侧重个体语言知识不同，互动能力强调语言能力是在具体互动情境中通过参与者之间的协商与合作共同建构的。这一理论强调语言能力不仅体现在语言结构知识的掌握，还体现在参与真实交际活动时的互动策略与语境理解能力。(Peng et al., 2021)基于巴赫曼—帕默尔评估使用论证框架(Assessment Use Argument, AUA)对2009年版HSK进行评估时指出，由于缺乏强制性口语测试，该版本HSK在交际能力构念的代表性方面存在明显不足。

(Nation, 2001)提出的接受性—产出性能力不对称发展理论为理解语言能力的习得过程提供了重要心理语言学视角。该理论指出，接受性能力(如听力与阅读理解)通常早于产出性能力(如口语与书写)发展，且两者在认知负担与习得路径方面存在显著差异。这一理论对于解释汉语学习中识读能力与书写能力之间的发展差异具有重要意义，特别是在汉字学习中，书写往往涉及更高的认知与操作复杂度。此外，(Canagarajah, 2013)提出的汉语国际语视角以及(Ellis, 2003)的任务型语言评估方法也对现代语言测试理念产生了重要影响。前者强调语言使用的跨文化交际属性，而后者则主张通过真实任务情境评估学习者的语言能力。这些理论共同推动了语言评估从以母语者规范为中心的结构测试模式向更加注重实际语言使用能力的评估模式转型。综合来看，这些理论框架共同表明，当代语言能力评估正在从以离散语言要素为中心的测量模式转向以多维交际能力为核心的综合评估模式。这一理论转向为理解新HSK 3.0的设计逻辑提供了重要的理论基础。

## 2. HSK 的历史演变与旧版评估

在国际中文教育领域，现有多种汉语能力测评体系并存，例如面向青少年的中小學生汉语考试(YCT)、面向商务场景的商务汉语考试(BCT)，以及部分地区或机构开发的本土化汉语测评项目。然而，在这些测评体系中，汉语水平考试(HSK)凭借其覆盖范围最广、考生规模最大以及国际认可度最高，逐渐发展为全球最具代表性的标准

化汉语能力测评工具。汉语水平考试（HSK）自1984年创立以来，经历了多个发展阶段。早期的HSK体系主要以母语者语言规范为参照标准，侧重于对语法结构与词汇掌握情况的测量，并主要发挥选拔性评估工具的功能（Nie 聂丹, Zhao 赵琪凤, 2022）。这种设计理念在一定程度上反映了早期结构主义语言测试理论的影响。2009年改革后的HSK（通常被称为HSK 2.0）建立了六级考试体系，分别考查听力、阅读与写作三项语言技能。该体系在全球范围内得到广泛推广，并成为国际中文教育中最重要的标准化考试之一。然而，随着语言测试理论的发展，学界逐渐开始对该体系的构念效度提出批评。（Peng et al, 2021）运用巴赫曼—帕默尔评估使用论证框架对2009年版HSK进行系统评估，发现该考试在信度方面表现良好，但在构念效度方面存在若干关键问题。首先，第五级写作部分中的句子重组任务与阅读能力之间的相关度显著高于与写作能力之间的相关度，这表明该任务在测量目标能力方面存在偏差。

其次，由于缺乏强制性的口语测试，考试难以充分反映学习者的交际能力。第三，考试难度与CEFR以及CLPS等级描述符之间存在一定程度的不匹配。从反拨效应的视角来看，（Zhao, 2025）的系统综述表明，HSK对学习者的学习动机、学习策略与学习目标具有明显影响。总体而言，HSK对学习者的正向反拨效应大于负向效应。然而，对教师教学实践的影响相对有限。Wang 与 Zhang（2023）的研究进一步指出，2010年版HSK在测试设计上过于强调狭义语言能力，导致语用能力与互动能力在考试构念中的代表性不足。

这些研究结果表明，尽管HSK在国际中文教育中发挥了重要作用，但其传统设计在一定程度上仍然受到结构主义语言测试范式的影响。因此，对考试构念进行系统性改革成为语言测试领域的重要议题。

**3. 新HSK 3.0: 构念、架构与规范基础**新HSK 3.0改革建立在《国际中文教育中文水平等级标准》（CEFCIL, 2021）这一规范性文件的基础之上。该标准从语言知识、语言技能与语言运用三个维度对国际中文学习者的语言能力进行了系统界定，构建了一个更加综合的语言能力描述体系。与旧版HSK相比，新HSK 3.0在多个方面进行了结构性调整。其中最显著的变化之一是考试级别从六级扩展至九级，并按照基础（一级至三级）、中级（四级至六级）与高级（七级至九级）三个阶段进行划分。此外，新体系还引入了HSK（笔试）、PTH（口语）与CHBC（机构认证）三种考试路径，从而构建更加多元化的能力评估体系。

在词汇规模方面，研究表明新HSK 3.0词汇总量达到约11,000词，相较于旧版HSK 2.0的5,000词，实现了显著扩展。(Li, 2023)的比较研究指出，新体系不仅扩大了词汇数量，还将词汇范畴从传统单词扩展至词缀化表达、固定搭配以及功能性语言单位。这一变化反映出语言评估从词汇知识测试向语用功能单位评估的转变。与此同时，(Shi, 2025)通过比较CEFCIL与CEFR两个语言能力标准发现，CEFCIL在量化评估指标以及汉语特征描述方面具有明显优势，而CEFR则在跨语言能力比较方面更具普适性。因此，在国际中文教育实践中，学者普遍建议将CEFCIL与CEFR结合使用，以提高评估体系的国际可比性。

#### 4. 语言测试改革的教学影响

语言测试改革通常会对课程设计、教材编写以及课堂教学实践产生深远影响。Cheng (2004)指出，当考试构念能够较好地反映真实语言使用目标时，测试更有可能产生积极的反拨效应，从而推动教学改进。然而，这一理想状态在实际教学情境中并不总能实现，测试改革所带来的教学变化往往呈现出复杂性与不均衡性。在国际中文教育情境中，HSK 长期以来被广泛用作课程设计与学习目标制定的重要参照标准。这里所指的课程主要包括面向非母语学习者的对外汉语课程体系，如海外高校中文课程、孔子学院及各类国际中文教育机构的教學项目，同时也涵盖中国境内高校针对国际学生开设的汉语课程。因此，考试结构的变化不仅会影响教学内容的选择，还可能在更深层面重塑课程目标与能力培养取向。相关研究表明，在具体教学实践中，HSK 往往与 CEFR 等国际语言能力框架并行使用。例如，在印度尼西亚的中文教育体系中，高校课程目标通常以 CEFR 作为能力标准，而语言能力评估则依然主要依赖 HSK 体系 (Trihardini et al., 2020)。这一现象表明，HSK 不仅是一种测评工具，同时也在实际教学中发挥着“隐性课程标准”的作用，其结构调整将不可避免地影响课程体系的设计与实施。然而，需要指出的是，将考试标准直接转化为课程标准并非没有问题。一方面，HSK 3.0 所强调的综合语言能力与交际能力，对教师的教学能力以及教学资源提出了更高要求；另一方面，在现实教学情境中，部分教学机构仍可能出于应试需求而强化可测量的语言知识与考试技巧，从而削弱交际能力培养的目标。这种“构念理想”与“教学现实”之间的张力，正是语言测试反拨效应研究中的关键议题之一。

(Teng, 2017)的研究也表明，尽管教师与学习者总体上认可 HSK 3.0 的改革方向，但在具体实施过程中，教学机构仍面临诸多挑战，尤

其体现在课程体系重构、教材更新以及教师专业发展之间的协调问题。这进一步说明，测试改革并不会自动转化为教学改革，其实际效果在很大程度上取决于具体教学情境中的制度支持与资源配置。

综上所述，现有研究虽然从不同角度探讨了HSK改革，但整体上仍呈现出相对分散的研究格局。部分研究关注词汇规模变化，部分研究关注语言能力标准比较，而另一些研究则聚焦考试反拨效应。然而，目前仍缺乏能够系统整合语言测试理论、考试设计原则以及教学影响之间关系的综合性研究。因此，从理论整合与政策分析的角度对新HSK 3.0进行系统研究具有重要的学术价值。

### 研究方法

本研究采用解释性研究范式，通过质性研究方法分析新HSK 3.0改革的设计原则及其教育意义。研究设计结合文件分析与多源文献综述，以系统探讨考试改革的理论基础与政策逻辑。研究主要分析三类核心政策文件：《国际中文教育中文水平等级标准》（CEFCIL, 2021）、《中文水平考试（HSK）考试大纲》（2025年发布，2026年实施）以及中国国际中文教育基金会发布的新旧HSK大纲对比文件。同时，通过检索相关学术文献，筛选与HSK改革、语言能力标准比较以及语言测试理论相关的研究成果，以构建本研究的理论与文献基础。

在数据分析方面，本研究采用 (Braun et al., 2008)提出的主题分析方法 (Thematic Analysis, TAM)，并结合演绎与归纳的混合分析路径，对政策文本与文献进行系统编码与主题归纳。分析过程遵循六个阶段：熟悉数据、生成编码、搜索主题、审查主题、定义与命名主题以及撰写报告。通过反复阅读、持续比较与多轮编码修订，本研究识别出新 HSK 3.0 的关键设计特征，并进一步归纳其核心设计原则。为提升研究的可信度，本研究采用数据三角验证，并通过增强分析过程的透明性、一致性与解释力，确保研究结果的可靠性。

对政策文件与相关研究文献进行系统编码与主题归纳。分析过程通过反复阅读文本资料、开放式编码与持续比较，识别新HSK 3.0改革中的关键设计特征(曹贤文, 冯丽萍, 吴中伟, 2025)，包括多维语言能力构成、词汇规模扩展、词性标注体系以及识读与书写能力区分。在此基础上，进一步归纳出新HSK 3.0的三项核心设计原则，并结合语言测试理论进行解释性分析。为提高研究的可信度，本研究采用数据三角验证的方法，将政策文件与学术文献进行交叉比对，并通过透明的编码与分析过程增强研究结果的可靠性与解释力。

## 研究结果及探讨

### HSK 考试体系改革：从结构调整到能力构念转型

通过对《中文水平考试（HSK）考试大纲》（2025年版）及 CEFCIL 标准文件的系统分析，表 1 呈现了三版 HSK 考试体系在结构与能力描述方面的整体差异。从总体结构来看，新 HSK 3.0 将原有六级体系扩展为九级体系，同时词汇总量由 HSK 2.0 的 5,000 词显著增加至约 11,000 词。图 1 和图 2 进一步直观展示了三版 HSK 在各级词汇规模与核心统计数据上的变化趋势。整体而言，新体系在词汇规模、能力等级以及汉字数量等方面均呈现出明显的扩展趋势。(Peng et al., 2021) 对 2009 年版 HSK 的评估研究为理解这一改革提供了重要背景。该研究指出，旧版 HSK 虽然在信度方面表现稳定，但在构念效度方面存在明显挑战，例如写作任务与阅读能力之间的高度相关性以及口语能力代表性不足。这些问题表明，旧版 HSK 在一定程度上仍然受到结构主义语言测试范式的影响，其测试重点主要集中于语法和词汇知识的识别，而对真实交际能力的反映相对有限。

(Li, 2023)对新旧词汇体系的比较研究进一步揭示了改革的规模。(梁波, 2025)研究发现，新 HSK 3.0 词汇总量已超过旧版 HSK 2.0 的两倍，并且词汇范畴也得到显著扩展。图 1 所示的各级词汇分布变化表明，新体系不仅扩大了整体词汇规模，还对不同阶段的词汇负担进行了重新分配。这一变化反映出考试设计理念的转变，即从以离散语言要素为中心的测试模式逐渐转向以综合语言能力为核心的评估体系。

表 1. HSK 综合对比

| 对比项目    | HSK 2.0 | HSK 3.0 | 新 HSK 3.0 |
|---------|---------|---------|-----------|
| 一、总体结构  |         |         |           |
| 级别总数    | 6       | 9       | 9         |
| 词汇总量    | 5,000   | 11,092  | 11,000    |
| 汉字总量    | 2,663   | 3,000   | 3,088     |
| 二、各级词汇量 |         |         |           |
| 第 1 级   | 150     | 500     | 300       |
| 第 2 级   | 300     | 1,272   | 500       |
| 第 3 级   | 600     | 2,245   | 1,000     |
| 第 4 级   | 1,200   | 3,245   | 2,000     |
| 第 5 级   | 2,500   | 4,316   | 3,600     |

|               |        |         |        |
|---------------|--------|---------|--------|
| 第 6 级         | 5,000  | 5,456   | 5,400  |
| 第 7—9 级       | N/A    | 11,092  | 11,000 |
| 三、设计特征        |        |         |        |
| 书写要求          | 级别均须书写 | 全级别均须书写 | 仅五级及以上 |
| 词性标注<br>(POS) | 未提供    | 未提供     | 完整提供   |



图 1. 三版 HSK 各级别词汇量对比

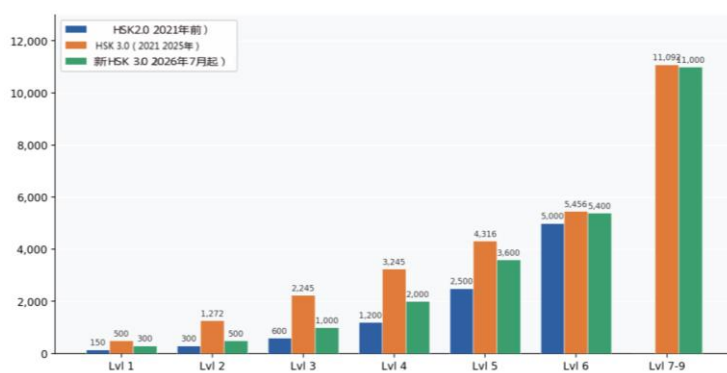


图 2. 三版 HSK 核心统计数据对比

基于表 1 所呈现的数据，可以进一步归纳出新 HSK 3.0 改革的三个核心设计原则；

第一，多维能力建构原则（principle of multidimensional competence construction）。从六级体系扩展至九级体系，以及词汇规模从 5,000 词增加至 11,000 词，体现了语言能力评估从单一语言知识测量向多维交际能力建构的范式转型。这一转变直接回应了(Peng et al., 2021) 所指出的旧版 HSK 构念代表性不足的问题。从理论角度来看，该原则与(Bachman, L. F., & Palmer, 2010) 提出的交际语言能力模型相一致，该模型强调语言能力应同时包括组织性知识与语用知识等多个维度。(Li,2023) 的研究进一步表明，新 HSK 3.0 通过词缀化词汇、惯用语表达以及语言点词汇化等方式，将更多语用功能单位纳入测试构念，从而使评估对象从单一词汇识别扩展至真实语境中的语言运用能力。

第二，交际真实性原则（principle of communicative authenticity）。《新 HSK 大纲》首次引入完整的词性标注（POS）体系，同时兼类词组数量由旧大纲的 20 组显著增加至 CEFCIL 中的 178 组（中外语言交流合作中心，2025 (Li, 2023)）。这一变化不仅提升了词汇系统的语言学准确性，也使考试内容更加贴近真实语言使用情境。从理论层面来看，该设计与(Young, 2011)提出的互动能力理论高度契合。互动能力理论强调，语言能力不仅体现在掌握抽象语法规则上，还体现在真实互动情境中的意义协商能力上。(Peng et al., 2021) 对旧版 HSK 情境真实性不足的批判，为这一设计创新提供了实证支持。

第三，渐进式降低汉字书写负担原则（principle of graduated graphemic load reduction）。根据表 1 所示的新考试设计，书写要求仅从第五级开始实施，而一级至四级主要考查汉字识读能力。这一设计调整被视为对数字化时代真实语言使用环境的适应性回应（中外语言交流合作中心，2025）。其心理语言学理论基础来源于(Nation, 2001)提出的接受性—产出性不对称发展理论，即识别词汇的认知负担显著低于产出词汇。因此，在语言学习初期阶段过度强调书写能力可能增加学习者的认知负担。(Peng et al., 2021)的研究亦提供了间接证据：机考成绩系统性高于纸笔考试成绩，其中一个重要原因在于拼音输入法降低了手写汉字的认知负担。

### 各级分布：CEFR 对应关系与词汇负担变化

表 2 展示了新 HSK 3.0 各级别在汉字数量、词汇规模以及 CEFR 对应关系方面的整体分布情况。整体来看，新体系在不同阶段呈现出明显的词汇负担再分配趋势。图 4 进一步展示了新 HSK 3.0 与 HSK 2.0 及 HSK 3.0 在词汇负担方面的变化幅度。(Peng et al., 2021)此前指出，旧

版 HSK 词汇规模与其能力描述之间存在一定程度的不匹配，即部分等级的词汇量低于其对应能力水平。新 HSK 3.0 在一定程度上对这一问题进行了系统性修正。(Li, 2023)的分析表明，在新体系中，原有基础至中级词汇整体下调一至两个等级，而高级阶段词汇则得到显著扩展。

表 2. 新 HSK 3.0 与 CEFR 对应及词汇变化比较

| 阶段          | 级别  | 汉字量   | 词汇量    | CEFR  | 较HSK 3.0 (2021) | 较HSK 2.0 |
|-------------|-----|-------|--------|-------|-----------------|----------|
| 初等<br>(1-3) | 1   | 246   | 300    | A1    | -40%            | +100%    |
|             | 2   | 371   | 500    | A1    | -61%            | +67%     |
|             | 3   | 655   | 1,000  | A2    | -55%            | +67%     |
| 中等<br>(4-6) | 4   | 1,096 | 2,000  | B1    | -38%            | +67%     |
|             | 5   | 1,527 | 3,600  | B2    | -17%            | +44%     |
|             | 6   | 1,940 | 5,400  | C1    | -1%             | +8%      |
| 高等<br>(7-9) | 7-9 | 3,088 | 11,000 | C2    | -1%             | N/A      |
| 合计          |     | 3,088 | 11,000 | A1-C2 | —               | —        |

注：CEFR 为欧洲语言共同参考框架。负值表示相较于 2021 版 HSK 3.0 词汇量下降，正值表示相较于 HSK 2.0 词汇量增加。

从整体变化趋势来看，新 HSK 3.0 在不同等级间呈现出明显的非均衡调整特征。初等阶段（1—3 级）词汇量相较于 2021 版 HSK 3.0 平均下降约 52%，显示出显著的减负趋势；而中等阶段（4—6 级）降幅明显收窄，平均下降约 19%，至第 6 级基本趋于稳定（-1%）。这一分布表明，新体系通过在低等级阶段降低词汇负担，同时在中高等级维持或微调词汇规模，从而实现整体结构的再平衡。与 HSK 2.0 相比，各等级词汇量均呈现正增长，初等及中等阶段平均增幅约为 69%，其中第 1 级增长幅度最大（+100%），而高等级增幅明显减弱（第 6 级仅为+8%）。这一变化说明，新 HSK 3.0 在保持高等级稳定性的同时，显著提高了低中等级的区分度。综合来看，该词汇调整呈现出“低等级压缩—中等级过渡—高等级稳定”的结构特征，这种分布有助于优化学习路径并增强等级之间的区分效度，但同时也意味着中级阶段的

学习负担集中化趋势，需要在教学实践中加以关注。

表 3. 新 HSK 等级、CEFR 对应及能力描述

| HSK 等级 | CEFR | 能力描述                                     |
|--------|------|--|
| 1-2 级  | A1   | 能够在日常情境中进行基本沟通，如自我介绍、理解简单信息及完成基本生活需求。    |
| 3 级    | A2   | 能够在较广泛的日常情境中进行交流，描述个人经历，并在基础社会情境中实现基本互动。 |
| 4 级    | B1   | 能够在初级学术与工作情境中使用语言，理解信息性文本并参与基础职业交流。      |
| 5 级    | B2   | 能够在学术与职业情境中有效沟通，表达较复杂观点并参与较深入讨论。         |
| 6 级    | C1   | 能够在多种正式情境中流利运用语言，理解复杂语篇并进行较高层次交流。        |
| 7-9 级  | C2   | 能够在高阶学术与专业情境中准确、灵活地使用语言，分析复杂语篇并进行跨文化沟通。  |

注：该能力描述基于新 HSK 3.0 大纲任务要求进行概括性归纳。

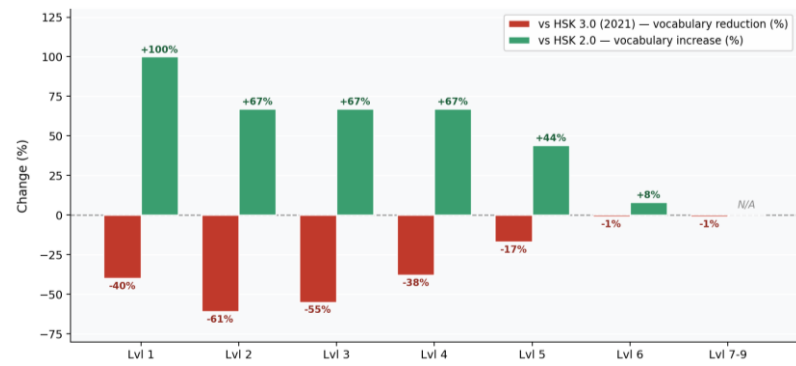


图 3. 新 HSK 3.0 各级别认读字与书写字对比

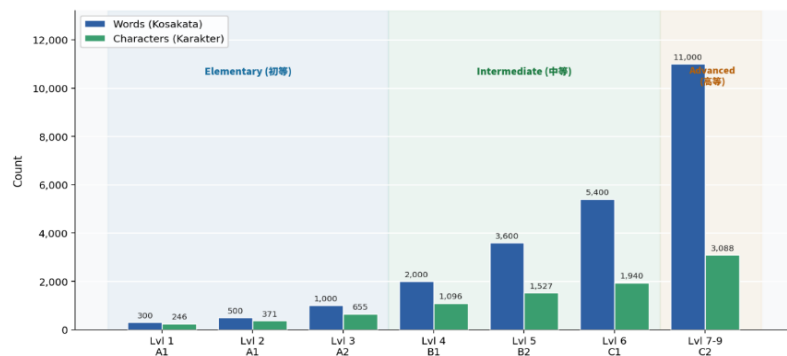


图 4. 新 HSK 的对比

表 2 所呈现的变化规律对研究问题 RQ2 具有重要启示。首先，在基础阶段（一级至三级），词汇负担相较 HSK 3.0（2021）版本出现明显下降，其中第二级降幅达到约-61%。这一变化与(Zhao, 2025)的研究结论相呼应。该研究发现，HSK 考试在低级别阶段对学习者的反拨效应最为明显，因此降低基础词汇门槛有助于提高学习的可及性与学习动机，从而产生更为持久的正向反拨效应。相反，在中高级阶段，与 HSK 2.0 相比，词汇规模出现明显扩展（约+44%至+100%）。这一变化意味着中高级阶段的教学与课程设计需要相应提高语言能力要求。(Shi, 2025)指出，在课程规划与能力评估过程中，同时参考 CEFCIL 与 CEFR 能力标准，有助于教学机构建立更加科学且具有国际可比性的课程体系。

### 识读汉字与书写汉字之区分：接受性—产出性原则的实践体现

通过对《新 HSK 大纲》（2025 年版）汉字部分的深入分析，可以识别出新 HSK 3.0 最具辨识度的设计创新之一，即明确区分识读汉字能力与书写汉字能力。表 3 详细展示了各级别识读汉字数量与书写汉字要求的具体分布情况，而图 5 与图 6 则直观呈现了不同等级之间书写要求的逐步增加趋势。根据大纲规定，一级至四级仅要求识读汉字，而从第五级开始正式要求书写 150 个汉字，第六级累计书写 300 字，第七至九级累计书写 800 字（中外语言交流合作中心，2025）。这一设计创新可以被视为当代心理语言学中接受性—产出性能力分层原则在 HSK 考试体系中的系统体现(Nation, 2001)。

表 4. 新 HSK 3.0 各级识读汉字与书写汉字区分情况

| 级别 | 识读汉字量 | 书写汉字量 | 书写状态       | 教学建议                    |
|----|-------|-------|------------|-------------------------|
| 一  | 246   | —     | 暂不作要求      | 优先听读理解；练习基本笔画，不作强制书写训练  |
| 二  | 125   | —     | 暂不作要求      | 维持识读重心；适当强化常用字的视觉识别     |
| 三  | 284   | —     | 暂不作要求      | 逐步渗透书写启蒙；引导学习笔顺规则       |
| 四  | 441   | —     | 暂不作要求      | 有限度引入产出性书写任务；培养句子书写能力   |
| 五  | 431   | 150   | 须书写（150 字） | 正式启动汉字书写训练；掌握 150 个书写汉字 |
| 六  | 413   | 150   | 须书写（       | 新增 150 个                |

|      |     |              |                |                 |
|------|-----|--------------|----------------|-----------------|
|      |     |              | 累计 300 字)      | 书写汉字 (累计 300 字) |
|      |     | 500          |                | 强化书面            |
| 七一九  | 998 |              | 须书写 (累计 800 字) | 表达; 书写负担显著加重    |
| 累计书写 | —   | 800 字 (五—九级) | —              | —               |

注: 表1 表明, 新 HSK 3.0 在扩大词汇规模的同时优化分级结构, 体现由知识导向向能力导向的转变。

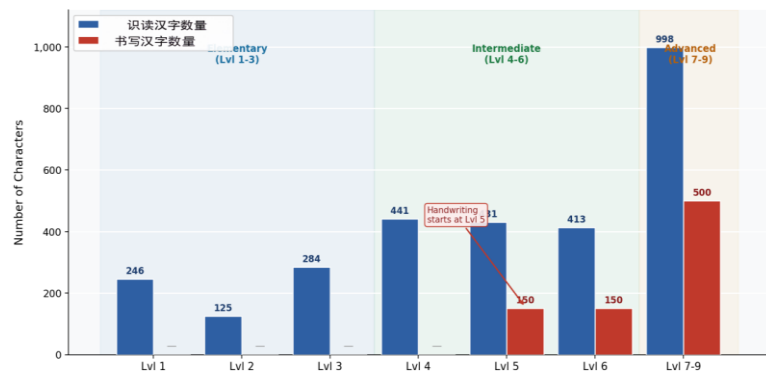


图 5. 新 HSK 3.0 各级别认读字与书写字对比 (书写要求自第 5 级起正式启动)

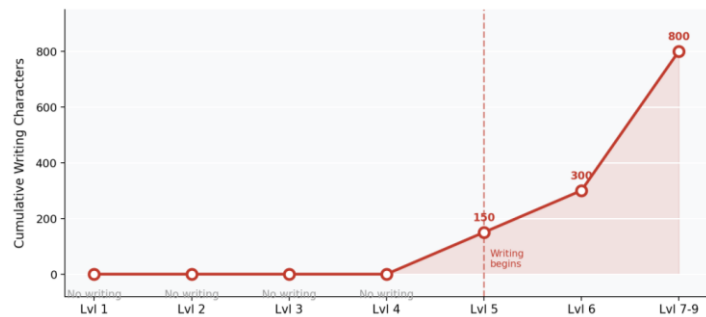


图 6. 新 HSK 3.0 各级别书写字累计数量

表 4.所呈现的分级结构对研究问题 RQ2 具有直接的教学启示。首先, 将书写训练延后至第五级, 使教师能够在基础阶段更加集中于听

说能力与阅读理解能力的发展。这一教学重心调整与(Young, 2011)的互动能力理论高度一致。其次,明确设定 800 字的累计书写目标,为中高级课程设计提供了具体且可操作的教学标准,从而有助于应对(Li, 2023)所指出的课程适应挑战。此外,随着数字技术的发展,拼音输入法已经成为现代汉语书写的重要方式之一。因此,在低级别阶段减少手写要求不仅符合语言习得规律,也反映了数字化语言使用环境的变化。(Shi, 2025)进一步指出,在教学实践中结合 CEFCIL 与 CEFR 能力标准,有助于在保持汉语语言特性的同时实现国际能力标准的可比性。

### 结论

本研究通过对《新 HSK 大纲》(2025)及相关政策文件的系统分析,探讨了新 HSK 3.0 考试体系的设计逻辑及其对国际中文教育的潜在影响。研究表明,新 HSK 3.0 在考试结构、词汇规模以及能力描述体系方面进行了系统性重构,并通过建立九级能力体系重新界定了汉语能力评估框架。综合分析显示,新 HSK 3.0 改革主要体现出三个核心设计原则:多维能力建构原则、交际真实性原则以及接受性—产出性能力分层原则。这些原则在一定程度上提升了语言能力构念的代表性,使考试评估从以离散语言知识为中心的测量模式逐步转向更加注重综合语言运用能力的评估体系。

从教育实践角度来看,新 HSK 3.0 在基础阶段适度降低词汇负担,并通过识读与书写能力的分层设计,为学习者提供了更加渐进的语言发展路径。同时,中高级阶段词汇规模的显著扩展以及语言任务复杂度的提升,也对课程设计与课堂教学提出了更高要求。这表明,新 HSK 3.0 不仅在评价体系层面进行调整,也在一定程度上重塑了教学目标与学习路径。

在政策层面,新 HSK 3.0 的实施不仅是考试体系的结构调整,更体现了国际中文教育评价标准的整体升级。与旧版 HSK 相比,新体系在词汇规模与分布方面发生了显著变化:总体词汇量明显增加,尤其是在中高级阶段扩展更为突出,同时词汇内容更加贴近学术、专业及真实社会语境。这一变化不仅提升了测试的区分度,也使语言能力评估更加贴近实际语言使用。然而,词汇规模的扩大及其复杂性的提升,也意味着学习难度的增加,学习者需要在更高层次上整合词汇知识与语用能力,从而对学习的深度与广度提出了更高要求。

这一转变对国际中文教育具有重要启示。首先,在课程标准方面,需要从以词汇与语法知识为核心的教学目标,转向以语言运用能力为

导向的能力体系。其次，在教材建设方面，应强化情境化与任务型设计，使词汇学习嵌入真实交际语境之中。再次，在教师发展方面，教师不仅需要具备词汇教学能力，还需具备引导学习者进行跨情境语言迁移与综合运用的能力。因此，为实现考试改革与教学实践之间的有效衔接，各教育机构有必要在课程规划、教学资源开发及师资培养方面建立更加系统化的协同机制，从而推动国际中文教育质量的整体提升。

需要指出的是，本研究主要基于政策文件与文献资料的分析，尚未结合具体课堂情境与学习者数据进行实证检验。因此，未来研究有必要通过课堂观察、教师访谈以及学习者表现分析等方法，对新 HSK 3.0 在实际教学情境中的实施效果进行进一步评估。同时，从长期视角追踪考试改革对课程体系、教学实践以及学习成果的影响，也将有助于更加全面地理解新 HSK 3.0 在国际中文教育发展中的作用。

### 参考文献

- [1] Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- [2] Braun, V., Clarke, V., Braun, V., & Clarke, V. (2008). Using thematic analysis in psychology *Using thematic analysis in psychology*. 0887(2006).
- [3] Canagarajah. (2013). *Translingual Practice Global Englishes and Cosmopolitan Relations*. In Routledge. <https://doi.org/Translingual Practice Global Englishes and Cosmopolitan Relations>
- [4] Cheng, L. (2004). *Washback or Backwash: A Review of Testing on Teaching and Learning*.
- [5] Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- [6] Hu, B. (2025). Unpacking the washback of the HSK on international Chinese language education: a study of the influencing factors. <https://doi.org/10.1186/s40468-025-00370-z>
- [7] Jin, Y. (2017). Construct and content in context: implications for language learning, teaching and assessment in China. 1–18. <https://doi.org/10.1186/s40468-017-0044-1>
- [8] Kong, F. (2024). Investigating the Washback of the Chinese Language Proficiency Test ( HSK ): From the Perspective of CSL Students. March, 1–15. <https://doi.org/10.1177/21582440231224599>

- [9] Li, R. (2023). Comparative analysis of Chinese proficiency grading vocabulary between HSK 2.0 and HSK 3.0. *Journal of Linguistics AndCommunication Studies*, 2. <https://doi.org/10.56397/JLCS.2023.03.01>
- [10] McNamara, T., & Roever, C. (2006). *Language testing: the social dimension*. 16(2).
- [11] Nation. (2001). *Learning Vocabulary in Another Language*. <https://doi.org/https://doi.org/10.1017/CBO9781139524759>
- [12] Nie 聂丹, D., Zhao 赵琪凤, Q. (2022). The History and Development of Assessment of Chinese as a Second Language. [https://doi.org/https://doi.org/10.1007/978-981-16-0924-4\\_4](https://doi.org/https://doi.org/10.1007/978-981-16-0924-4_4)
- [13] Peng, Y., Yan, W., & Cheng, L. (2021). Hanyu Shuiping Kaoshi ( HSK ): A multi-level , multi-purpose proficiency test. <https://doi.org/10.1177/0265532220957298>
- [14] Shi, C. (2025). A Comparative Study of the Chinese Proficiency Grading Standards for International Chinese Language Education and the Common European Framework of Reference for Languages. 23(8), 194–204.
- [15] Trihardini, A., Wikarti, M.A., A. R., Anggrini, R. N., & Abdurrahman, R. (2020). KESETARAAN HANYU SHUIPING KAOSHI LEVEL I-IV DENGAN CEFR PADA KETERAMPILAN BERBICARA BAHASA MANDARIN. *Jurnal Cakrawala Mandarin*, 4(1), 11. <https://doi.org/10.36279/apsmi.v4i1.94>
- [16] Teng, Y. (2017). Hanyu Shuiping Kaoshi ( HSK ): Past , Present , and Future. <https://doi.org/10.1007/978-981-10-4089-4>
- [17] Xinsheng, G. (2020). New Chinese Proficiency Standards and Its Impacts on Chinese Proficiency Test. June 2020, 24–26.
- [18] Young, R. F. (2011). *Interactional Competence in Language Learning, Teaching, and Testing*. Routledge, 18. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203836507-30/interactional-competence-language-learning-teaching-testing-richard-young>
- [19] Zhao, M. (2025). The Washback Effect of HSK on Students ’ Learning and Teachers ’ Teaching : Implications for Chinese Language Education. 18(3).
- [20] 曹贤文,等. (2025). “HSK3.0 新考纲及考试推广”大家谈. *云南师范大学学报 ( 对外汉语教学与研究版 )*, 23(05), 1–8. <https://doi.org/10.16802/j.cnki.ynsddw.20251029.001>
- [21] 梁波, 孙明材. (2025). 汉语国际教育本科用文学教材词表与 HSK 词汇适配情况研究. *品位·经典*, (24), 40–43.
-

- [22] <https://doi.org/10.14014/j.cnki.cn11-2597/g2.2025.24.043>
- [23] <https://www.chinesetest.cn/>
- [24] <https://www.ecnu.edu.cn/info/1426/71479.htm>
- [25] <https://www.mandarinzone.com/new-hsk-test/>
- [26] Chinese Proficiency Levels Explained: HSK, CEFR & Real-World Ability | HSKLord
- [27] New HSK levels for learning Chinese (Update 2025) - GoEast Mandarin