

任务特征框架视阈下的大学英语四级翻译测试内容效度研究

柳娟, 胡小青

新疆大学外国语学院, 新疆 乌鲁木齐

收稿日期: 2025年6月9日; 录用日期: 2025年7月15日; 发布日期: 2025年7月25日

摘要

内容效度是语言测试领域评价测试质量的重要标准之一。翻译测试是对受试者翻译能力的测量, 贯穿翻译人才培养的全过程。本文以Bachman & Palmer的任务特征框架为理据, 以1996~2025年期间的大学英语四级段落翻译试题为样本, 从环境特征、测试规则、语篇输入、预期回答、语篇输入与预期回答之间的关系维度探究该测试的内容效度。研究表明, 大学英语四级翻译试题内容与拟测试构念趋于一致, 该测试与《大学英语四、六级考试大纲》测试目标以及《大学英语教学指南》教学要求相关性较高, 但该试题在语篇输入、预期回答两个核心维度方面存在不足, 即试题所涉题材、体裁、题型单一, 测试难度偏低。基于测试内容效度研究, 以期为大学英语四级翻译试题内容效度的进一步优化提供研究性支持。

关键词

内容效度, 大学英语四级, 翻译测试, 任务特征框架, 构念

A Study on the Content Validity of College English Test Level 4 Translation Tasks from the Perspective of the Task Characteristics Framework

Juan Liu, Xiaoqing Hu

School of Foreign Studies, Xinjiang University, Urumqi Xinjiang

Received: Jun. 9th, 2025; accepted: Jul. 15th, 2025; published: Jul. 25th, 2025

Abstract

Content validity is one of the important criteria for evaluating the quality of language tests. Translation testing measures the translation ability of test-takers and runs through the entire process of cultivating translation professionals. This research, based on the Task Characteristics Framework by Bachman & Palmer, takes the translation tasks from the College English Test Level 4 (CET-4) paragraphs between 1996 and 2025 as the sample, and explores the content validity of this test from the perspectives of environmental characteristics, test rules, input, expected responses, and the relationship between input and expected responses. The study findings indicate that the content of the CET Level 4 translation tasks is largely consistent with the construct being tested. The test shows a high correlation with the testing objectives outlined in the College English Test Band 4 and 6 Syllabus and the College English Teaching Guidelines, but there are shortcomings in the two core dimensions of input and expected responses. Specifically, the test topics, genres, and question types are too narrow, and the difficulty level of the tasks is relatively low. Based on the content validity of the test, this study aims to provide research support for the further optimization of the content validity of the College English Test Level 4 translation tasks.

Keywords

Content Validity, CET, Translation Testing, Task Characteristics Framework, Construct

Copyright © 2025 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

全国大学生英语四级考试(CET-4)是在教育部高等教育司的领导下,由全国大学英语四级考试委员会主持开发、设计并与教育部考试中心共同实施的全国性大规模标准化的水平测试。它主要用来评估中国大学生的基础英语语言能力,即接受性能力听力、阅读,以及产出性能力口语、写作、翻译。自1987年实施以来经历了多次改革,其中翻译试题顺应时势发生了最为明显的变化:1996年,翻译试题正式出现在大学英语四级考试中;2006年,英译汉的短句翻译变成汉译英;2013年,句子翻译更新为有关中国文化的段落翻译。从最初的语言能力考查,到与中国文化的融合,翻译测试是对受试者翻译能力的测量,更贯穿了培养“讲述中国故事,传播中国声音”的翻译人才的全过程[1]。现阶段,有关大学英语四级的研究聚焦于该测试对英语教学的反拨作用,鲜有研究从内容效度视角出发,探究试题内容与拟测试构念的一致性,以及与《大学英语四、六级考试大纲》测试目标、《大学英语教学指南》教学目标的相关性。翻译测试被广泛应用于外语课堂教学和大规模外语等级考试,对翻译教学与外语教学具有重要的“反拨作用”[2]。因此,评估大学生英语四级考试翻译试题内容效度至关重要。

2. 相关研究

大学英语四级作为全国性大规模标准化测试之一,涵盖了听力、阅读、写作和翻译等多个方面,旨在全面考察学生的英语综合应用能力。通过不同题型的设置,能够准确测量学生在不同语言技能上的掌握程度,为教学和人才选拔提供反馈。通过运用 Cite-space 对 CNKI 及 WOS 的文献进行聚类分析发现,现阶段围绕大学英语四级考试的研究聚焦于该考试对教学的反拨作用,鲜有研究以“大学英语四级考试

试题”为中心探究试题本身的内容效度,即试题内容是否全面覆盖了该测试的拟测试构念。

2.1. 四级考试内容效度研究

“效度”这一概念最初是由美国心理学家罗伯特·加涅在 20 世纪 60 年代提出的。1961 年,经 Lado 将这一心理测量学领域的概念引入到语言测试领域。之后,Heaton 在 *Writing English Language Testing* 中将语言测试效度分为表面效度、内容效度、构念效度和实证效度。

已有大学英语四级考试的内容效度研究聚焦于“阅读”、“写作”两个方面。穆惠峰[3]以四级考试中的完形填空题为语料对所涉考点及搭配进行频率统计和显著性检验,发现该试题整体上符合考试大纲的要求,但选项的干扰选项设计不够科学。在快速阅读试题方面,研究者[4][5]也发现试题选材与《考试大纲》要求基本一致,考点也在教学要求范围内,但语篇题材、体裁、题型类型较单一,而且阅读难易度变化幅度过大。此外,Xie [6]发现阅读试题的语篇可读性、文本长度、阅读速度也在《考试大纲》要求范围内。

在写作试题方面,谭燕保[7]从题目及测试方法两方面对写作试题设计及实施的合理性进行评估,发现试题题材和体裁单一,尚未覆盖《考试大纲》中规定的应考内容,而且评分标准局限于词汇、句法层面表达的准确性,忽略了从文体、内容表达、组织结构、策略等方面的评判标准。此外,大学英语四级写作在互动性层面欠缺,大多数题目缺乏明确的交际目的和交际对象[8]。

现阶段,有关大学英语四级考试的内容效度研究局限于阅读与写作两个方面,忽略了对考查产出性能力翻译的试题进行内容效度评估。此外,大多数研究样本选取数量较少,且时间跨度小,可能无法全面反映英语四级各部分的整体内容效度情况。

2.2. 翻译考试内容效度研究

翻译是跨文化交流的重要工具。以考试为依托的翻译测试,可以促使学习者参与语言转换活动,而且能够帮助全球化背景下的语言学习者提高对不同文化背景下的语言表达的敏感性和适应性,增强跨文化交际能力。当前专业及非专业外语课堂教学和大规模外语等级考试,翻译测试被广泛使用,诸多研究聚焦于该测试对教学的反拨作用,忽略了以试题内容与拟测试构念的一致性,以及与《考试大纲》测试目标[9]、《教学指南》教学目标[10]的相关性中为出发点,评估翻译测试的内容效度。有关翻译测试效度研究主要聚焦于内容评估、评估方法以及影响因素。

在内容效度评估方面,徐鹰[11]以 2017 年至 2021 年四、六级翻译试题及其参考译文为对象,从语篇输入和预期回答两个维度进行详细分析发现,该测试任务具有较高的内容效度,且基本符合《考试大纲》的要求。刘佳玲[12]则以英语专业八级翻译试题为研究对象,也发现翻译试题整体内容效度较高,符合考试大纲对翻译技巧的要求。然而,研究者从结构效度与内容效度层面对四、六级翻译测试进行内容评估发现,翻译试题题材选择范围广,但试题内容与《考试大纲》一致性较低,题目数量之少,考查范围之窄,试题内容上下文关联度不足[13]。此外,也有研究者以源语言为研究对象邀请具有丰富教学经验的教师对翻译内容的多元素构成进行一致性验证,发现 2014 年至 2018 年的四级翻译试题的源语文本在一定程度上符合命题大纲要求,但题材以及体裁选择单一,主要是以社会、文化为主的说明文,有关经济、传统文化以及伦理主题的记叙性、描写性、论述性等文本类型少[14]。除对特设翻译试题进行内容效度分析外,也有研究对阅读中的句子翻译试题进行内容效度进行定量分析,通过句子翻译成绩与阅读理解总成绩之间的相关性证实了此类翻译试题的有效性[15]。

在内容效度评估方法与影响因素层面,郑美玲[16]以 Rasch 模型综合翻译测试中受试者的语言能力及其评分人、任务、评分标准、评分量表等多种因素对分数进行修正补偿,发现该模式对提高翻译试题质量具有一定的指导意义,但也存在数据误差、评分一致性、研究工具的局限。此外,也有研究者基于

PACTE 翻译能力模式, 自建涵盖整体印象评分量表以及分项评分量表模型, 综合受试者及评分人员提高翻译产出与评估的信度与效度。

翻译测试是对受试者翻译能力的测量, 贯穿翻译人才培养的全过程。综上所述, 翻译测试研究数量少, 且评估维度单一, 现有研究也主要基于“语篇输入”以及“预期回答”维度进行小样本试题内容效度研究。据此, 本研究以 1996~2025 所有的语篇翻译试题为研究对象, 从任务特征框架的五个维度对大学英语四级翻译试题的内容效度进行深度评估。

2.3. 任务特征框架下的内容效度研究

在语言测试中, 可用于分析测试内容效度的主要理据有 Heaton 的测试框架及 Bachman 和 Palmer 的任务特征框架理论, 前者以测试内容与教学大纲的一致性, 忽略了测试内容的多样性与灵活性, 而且该框架尚未详细说明具体分析和评估测试内容与教学目标之间匹配程度的具体步骤。然而, Bachman 和 Palmer 的任务特征框架[17]涵盖了测试任务的多个维度, 可以基于测试任务特征的分析, 评估测试内容是否全面覆盖了目标语言能力的各个方面。现阶段, 基于任务特征框架的内容效度研究聚焦于阅读试题的研究, 忽略了对考查产出能力的“写作”、“翻译”试题进行内容效度研究。仅有徐鹰[11](2023)以任务特征框架为理据, 对连续五年的四、六级翻译试题进行“语篇输入”、“预期回答”维度的分析, 发现该试题具有较高的内容效度, 但题材、体裁、考试构念表征仍然有待提高。

综上所述, 已有测试研究聚焦于对教学的反拨作用, 并且仅有的内容效度研究也以阅读测试内容为核心, 忽略对考查产出能力的翻译试题内容进行效度评估。鉴于此, 本研究以大学英语四级 1996 年至 2025 年期间的所有翻译试题为语料, 以任务特征框架为主要理据, 探究该测试内容与拟测试构念的一致性以及与《考试大纲》测试目标、《教学指南》教学目标的相关性, 以期为大学英语四级翻译试题内容效度的优化提供研究性支持。

3. 研究设计

3.1. 理论依据

任务特征框架理论(Framework of Language Task Characteristics)是 1996 年由 Bachman 和 Palmer 在著作 *Language Testing in Practice* 中提出的一种用于评估语言测试质量的理论框架。旨在分析该框架之下语言测试任务的有效性和可靠性, 即测试任务的特征是否与测试目标相一致。该框架将测试任务的特征分为五个主要维度, 即场景特征、测试规则特征、输入特征、预期回答特征、输入与回答之间的关系。

(1) 场景特征。场景即为测试环境, 指测试实施的具体环境条件, 包括物理环境、考试参与者以及考试时间安排。

(2) 测试规则特征。测试规则是对受试者进行的测试指导和说明, 包括测试的目的、要求、答题时间以及评分标准。

(3) 输入特征。输入特征指的是测试试题中呈现给受试者的语言材料, 包括试题的呈现形式(文字、音频或视频)以及输入的语言。

(4) 预期回答特征。预期回答指的是受试者在完成测试试题后提供的回答或反应, 即考生对试题做出的预期反应, 包括期望回答的形式以及回答的语言。

(5) 输入与回答之间的关系。预期回答与语篇输入的关系涉及三个方面, 即互动关系、关系范围、关系的直接性。互动关系按照是否涉及他者反馈及反应分为互惠性关系、非互惠性关系以及适应性关系; 关系范围根据数量或范围分为“宽”范围和“窄”范围; 关系的直接性是根据语言活动的顺利完成多大程度上取决于使用者的输入、主题知识以及语境推断, 分为直接关系和间接关系。

基于任务特征框架理论, 本研究梳理出适用于该研究的核心理论框架, 其中以语篇输入以及预期回答、输入与回答之间的关系为核心维度评估大学英语四级翻译试题的内容效度(见表 1)。

Table 1. The analytical framework of task characteristics framework

表 1. 任务特征框架理论分析框架

Dimension	Elements	Detail
(1) Characteristics of the input (2) Characteristics of expected response	Format	Channel
		Form
		Language
		Length/Time
		Vehicle
	Language	Degree of speediness
		Type
		Theme
		Genre
		Text length Text difficulty
(3) Characteristics of expected response and input	Type of external interactive-ness	Reciprocal/Non-reciprocal
	Scope of relationship	Broad/Narrow
	Directness of relationship	Direct/Indirect

3.2. 研究对象

本研究以 1996~2025 年间共 70 篇段落翻译试题及其参考译文为语料, 共计 22,500 多字。该试题 1996 年、2000 年考查形式为阅读理解中的单句英译汉, 从 2012 年 12 月开始每次考次备有三套翻译试题平行卷, 2013 年 12 月份单句翻译更新为段落翻译, 考试时间为 30 分钟, 分值比例为 15%。

3.3. 研究步骤

内容效度是语言测试领域评价测试质量的重要标准之一[18]。它指的是考试的题目在多大程度上能代表它所测量的内容, 考试内容与考试大纲规定内容的一致性。本研究以《大学英语四级考试大纲》《中国英语能力等级量表》《大学英语教学指南》为参考标准, 以任务特征框架的五个维度为视角探究大学英语四级翻译试题的内容效度。具体步骤为:

(1) 以 1996 年至 2025 年大学英语四级试题为语料建立语料库。

(2) 运用内容分析法探究测试环境特征(时间安排)、测试规则特征(考试说明、试题结构、评分方法)、语篇输入特征(题材、体裁、文本长度、语篇难度)、预期回答特征(测试题型、测试语言、答题速度、考察内容)、预期回答与输入之间的关系(互动类型、关系范围、直接性)。

(3) 以量化及质性数据探究测试内容特征与该测试的拟测试构念的一致性, 以及与《考试大纲》的测试目标、《教学指南》的教学目标的相关度。

根据上述, 本研究的研究问题如下:

(1) 全国大学生英语四级考试翻译试题内容是否与《考试大纲》规定的测试目标、《教学指南》教学

目标相关?

(2) 全国大学生英语四级考试翻译试题内容是否与该考试的拟测试构念一致?

4. 分析与讨论

本研究从任务特征框架的五个维度, 即测试环境特征、测试规则特征、语篇输入特征、预期回答特征、预期回答与语篇输入之间的关系分析大学英语四级翻译试题的内容特征, 由此探究试题内容与该测试拟测试构念的一致性, 以及与《考试大纲》的测试目标、《教学指南》的教学目标的相关度。

4.1. 环境特征

自 2006 年起, 大学英语四级考试每年六月、十二月各举行一次, 具体考试时间为 9:00~11:20。该考试是由国家教育部高等教育司主办的全国性英语考试, 其具体组织和实施是由各个高校负责的。以受试者为中心, 涉及到监考人员、监考组织人员等多方人员。校际之间、校园内部各考场之间具有相同的试卷配置、考场环境设置以及考试组织设置, 因此可以保证考试的公平性, 促使四级考试能够准确反映并测量受试者的翻译综合能力。

4.2. 测试规则特征

测试规则包括考试说明、考试结构、考试持续的时间以及评分标准。大学英语四级翻译试题为笔译形式, 试题内容以中国文化、中国社会发展、中国自然地理为主, 综合考查了受试者的源语解读能力以及目的语的输出能力(见表 2)。考试说明清晰地阐述了考试的目的、考查的知识点和能力范围等, 使命题者能够精准地围绕这些目标和范围编制试题, 确保试题内容全面覆盖考试所要考查的核心内容, 避免出现考查内容的遗漏或偏颇, 从而保证试题内容效度; 考试结构规定了不同考查内容在试题中的分布和比重, 使命题者能够合理安排各类题型和考查点, 确保试题内容的全面性、代表性、科学性; 考试持续时间的合理设置能够保证考生有足够的时间完成所有试题, 使受试者的能力能够得到充分的体现; 明确的评分标准能够确保评分的客观性和准确性, 使评分者之间以及评分者内部按照统一的标准对受试者的作答进行评价, 避免因评分者的主观因素导致无法基于分数准确反映受试者的能力。

Table 2. The environment and rule characteristics of CET-4 Translation

表 2. 大学英语四级翻译考试的环境及规则特征

测试规则特征	具体内容	
(1) 考试说明	考试形式	笔试形式, 考生需要在答题卡上进行作答
	考试内容	汉语段落翻译成英语的能力, 涉及中国的文化、社会等
	考试要求	准确理解原文的意思, 并用流畅、准确的英语表达出来
(2) 考试结构	分值比例	占整套试卷的 15%, 对应的分值为 106.5 分
	题目数量	1 个段落翻译题
(3) 考试时间	答题时限	30 分钟
(4) 评分	评分标准	五个档次; 标准化评分及计分流程

4.3. 语篇输入特征

语篇输入特征涵盖考试任务输入材料的题材、体裁、文本长度、语篇难度, 本文将基于此四个方面分析 1987 年至 2025 年期间大学英语四级段落翻译试题的语篇输入特征。

4.3.1. 题材

本研究基于 70 多篇大学英语四级段落翻译试题的分析,发现该试题题材以中国文化和中国社会发展为核心(见图 1),其中涉及中国传统文化“剪纸”“舞狮”“灯笼”、中国的饮食文化“团圆饭”“北京烤鸭”以及中国社会发展的方方面面,如“交通工具”“移动支付”“农村教育”(见表 3)。2013 年至 2025 年的所有语篇是与受试者有互动性的“中国文化”、“中国社会发展”、“中国自然”、“中国地理”,这与《考试大纲》中的考核要求高度一致,但未覆盖《大学英语教学指南》中提及的“摘译与所学专业或未来所从事工作岗位相关的文献资料”。题材与受试者的社会生活互动性高,但与受试者的专业学习相关度较低。此外,该试题题材设计局限于中国文化圈,单一文化的滋养,不利于受试者跨文化能力的培养。测试题目需全面覆盖所要测量的语言能力或构念的各个方面,局限于单一题材的试题设计,无法保证翻译试题的构念效度,影响测试结果有效性和可靠性。

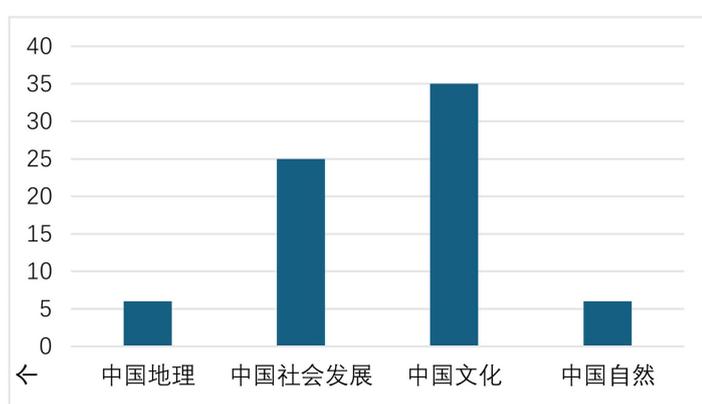


Figure 1. The theme of CET-4 translation
图 1. 大学英语四级翻译试题题材类型

Table 3. The theme, genre and text length of CET-4 translation
表 3. 大学英语四级翻译试题题材、体裁、文本长度统计

题材					
时间	卷号	题材类型	主题	体裁	文本长度
2025.06	(一)	中国社会发展	杂交水稻	说明文	153
	(二)	中国社会发展	冰雪大世界	说明文	139
	(三)	中国社会发展	便民生活圈	说明文	156
2024.6	(一)	中国文化	四合院	说明文	158
	(二)	中国文化	农历	说明文	154
	(三)	中国文化	“福”字	说明文	161
2024.12	(一)	中国社会发展	新能源汽车	说明文	156
	(二)	中国地理	敦煌莫高窟	说明文	154
	(三)	中国社会发展	环境保护	说明文	151
2023.6	(一)	中国社会发展	终身教育	说明文	148
	(二)	中国社会发展	高等教育	说明文	154
	(三)	中国社会发展	自助旅游	说明文	156

续表

2023.12	(一)	中国社会发展	健康饮食	说明文	148
	(二)	中国社会发展	饮食变化	说明文	152
	(三)	中国社会发展	饮食变化	说明文	158
2022.6	(一)	中国文化	揠苗助长	说明文	158
	(二)	中国文化	亡羊补牢	说明文	160
	(三)	中国文化	守株待兔	说明文	158
2022.12	(一)	中国文化	立秋	说明文	153
	(二)	中国文化	立春	说明文	159
	(三)	中国文化	冬至	说明文	162
2021.6	(一)	中国文化	铁观音	说明文	157
	(二)	中国文化	普洱	说明文	164
	(三)	中国文化	龙井	说明文	162
2021.12	(一)	中国地理	大运河	说明文	157
	(二)	中国地理	坎儿井	说明文	167
	(三)	中国地理	都江堰	说明文	155
2020.9	(一)	中国文化	茅台	说明文	154
	(二)	中国文化	茶	说明文	162
	(三)	中国文化	北京烤鸭	说明文	161
2020.12	(一)	中国文化	饮食差异	说明文	150
	(二)	中国文化	团圆饭	说明文	152
	(三)	中国文化	鱼	说明文	148
2019.6	(一)	中国文化	灯笼	说明文	143
	(二)	中国文化	舞狮	说明文	149
	(三)	中国文化	剪纸	说明文	141
2019.12	(一)	中国文化	家庭教育	说明文	155
	(二)	中国文化	家庭观念	说明文	159
	(三)	中国文化	姓氏	说明文	159
2018.6	(一)	中国社会发展	飞机	说明文	144
	(二)	中国社会发展	公交车	说明文	148
	(三)	中国社会发展	地铁	说明文	147
2018.12	(一)	中国社会发展	手机阅读	说明文	156
	(二)	中国社会发展	在线社交	说明文	157
	(三)	中国社会发展	手机购物	说明文	159
2017.6	(一)	中国自然	黄河	说明文	159
	(二)	中国自然	长江	说明文	160
	(三)	中国自然	珠江	说明文	147
2017.12	(一)	中国自然	泰山	说明文	142
	(二)	中国自然	华山	说明文	145
	(三)	中国自然	黄山	说明文	139

续表

2016.6	(一)	中国文化	功夫	说明文	165
	(二)	中国文化	潍坊风筝	说明文	149
	(三)	中国地理	乌镇	说明文	155
2016.12	(一)	中国文化	红色	说明文	154
	(二)	中国文化	白色	说明文	144
	(三)	中国文化	黄色	说明文	150
2015.6	(一)	中国社会发展	中国	说明文	164
	(二)	中国社会发展	快递服务	说明文	162
	(三)	中国文化	大米	说明文	157
2015.12	(一)	中国文化	家庭教育	说明文	163
	(二)	中国地理	丽江古镇	说明文	155
	(三)	中国社会发展	外国人汉语演讲比赛	说明文	155
2014.6	(一)	中国社会发展	农村教育	说明文	154
	(二)	中国社会发展	核电	说明文	169
	(三)	中国文化	全民阅读日	说明文	158
2014.12	(一)	中国社会发展	互联网	说明文	160
	(二)	中国社会发展	假日经济	说明文	149
	(三)	中国社会发展	旅游	说明文	158
2013.12	(一)	中国文化	中餐	说明文	146
	(二)	中国文化	中国结	说明文	153
	(三)	中国文化	茶	说明文	162

4.3.2. 体裁

通过人工标注以及 ChatGPT 辅助核查, 发现大学英语四级翻译试题均为说明文, 尚未涉及《考试大纲》中要求的记叙文、议论文, 也未涵盖《教学指南》中提及的“正式的议论性或不同话题的口头或书面材料”。单一体裁的段落翻译易造成负面教学反拨作用, 导致受试者在日常学习中集中于说明文的练习, 忽略多元化体裁的学习。此外, 单一体裁试题可能无法准确反映或测量受试者的翻译能力, 易导致测试效度过高或过低。

4.3.3. 文本长度

本研究运用 Microsoft Word 的字数统计功能对大学英语四级段落翻译试题进行文本长度分析, 发现该试题的源语文本段落字数平均值为 154 字。该字数在《考试大纲》所规定的文本长度范围(140~160)之内, 具有一定的效度。

4.3.4. 语篇难度

本研究运用由美国教育考试中心开发的文本难度分析器 Test-Evaluator, 从词汇、句子结构、衔接、组织四个方面对大学英语四级翻译测试的目的语文本进行难度评估。研究发现大学英语四级翻译试题整体难度偏低, 其中句法复杂度、词汇陌生度、衔接、叙事性均处于中等难度(见表 4)。该测试词汇、句法、衔接、叙事性能力都有所考察, 该测试难度与《考试大纲》中规定的“语言难度较低”要求高度一致, 但是简单试题易造成大部分受试者的分数普遍集中于高分段的“天花板效应”, 试题区分度降低, 无法进

一步区分受试者能力的细微差别。因此, 试题无法综合测量并反映受试者的真实翻译能力。此外, 简单试题可能由于其考查内容的单一性和局限性, 使得测试结果受到一些非构念因素的影响, 如受试者对试题的熟悉程度、应试技巧的运用, 而不是真正考查到被试者所具备的构念能力, 从而导致测量偏差, 影响测试构念效度。

Table 4. The difficulty level of CET translation test

表 4. 大学英语四级翻译试题难度

(1) 句法(1~100)	(2) 词汇(1~100)	(3) 衔接(1~100)	(4) 叙事性(1~100)	整体难度(100~2000)
60.1	63.9	50.3	44.5	880.5

注: 句子结构、词汇难度: 数值越高, 复杂度越高; 衔接、叙事性: 数值越低, 难度越高。

另一方面, 本研究应用 SPSS 进行 ANOVA (方差分析), 发现大学英语四级翻译三套题并无显著性差异($P > 0.05$) (见图 2), 三套平行试题在内容选择上较为均衡, 覆盖了相似的知识点和技能, 没有明显的偏重或遗漏, 整体难度设计较为一致。

ANOVA					
V2	平方和	自由度	均方	F	显著性
组间	166616.472	2	83308.236	2.488	0.090
组内	2310141.528	69	33480.312		
总计	2476758.000	71			

Figure 2. Significance analysis of CET translation

图 2. 大学英语四级翻译试题平行卷显著性分析

4.4. 预期回答特征

在任务特征框架下, 语篇输入特征与预期回答特征都具有相似的维度, 即“Format”与“Language”, 但两个维度强调不同的主体, 前者是试题材料特征, 后者则是受试者对试题的反应, 包括“渠道(Channel)”、“形式(Form)”、“语言(Language)”、“类型(Type)”“速度(Degree of Speediness)” (见表 5), 本研究的预期回答特征分析将以题型、速度为主要切入点。

Table 5. Characteristics of the input and expected response

表 5. 语篇输入与预期回答特征

维度	语篇输入	预期回答
渠道	Visual	Visual
形式	Language-based	Language-based
语言	Chinese	English
类型	Item	Item
速度	140~160/30 seconds	139~162/30 minutes

4.4.1. 翻译类型

自 2013 年起, 大学英语四级翻译试题从单句翻译转变为篇章翻译。相比于单句翻译, 篇章翻译对受试者综合能力的考查更为综合, 即语言能力、翻译策略、文化素养。然而, 该试题只设置“汉译英”, 未能涵盖《教学指南》中提及的“英汉互译”文章, 试题只能反映或测量受试者的汉译英翻译能力, 无法保证受试者综合翻译能力的测量效度。因此, 为了满足《教学指南》中对学生英汉互译能力的培养要求, 有学者建议在汉译英试题的基础上增加英译汉模块。

4.4.2. 速度

通过对大学英语四级翻译试题的字数统计发现, 该试题的源语文本字数的平均值为 154 字, 翻译试题作答时间为 30 分钟。这与《考试大纲》规定的“每 30 分钟 140~160 个汉字翻译”基本一致, 受试者基本可以在有限时间内完成对个人翻译能力的测量。

4.5. 预期回答与输入的关系

在任务特征框架下, 预期回答与语篇输入的关系涉及三个方面, 即互动关系、关系范围、关系的直接性。

4.5.1. 互动关系

Bachman & Palmer 将互动关系分为三类, 包括互惠性关系、非互惠性关系以及适应性关系(见表 6)。互惠性关系是指语言活动的完成依赖于使用者的共同反应和反馈; 非互惠性关系是指语言活动不涉及使用者之间的互动和反馈。适应性关系是指语言活动中使用者的输入受到响应的制约, 但没有反馈。大学英语四级翻译试题是受试者输入源语文本并输出目的语文本, 在此过程中只涉及单个主体受试者, 尚无其他主体的反应与反馈, 因此该试题为非互惠性任务。大学英语四级翻译为笔译形式, 而且《考试大纲》考核要求规定受试者的译文“基本准确地表达原文的意思, 语句通顺, 句式和用词较为恰当”, 因此该互惠性试题具有效度。

Table 6. The illustration of interaction

表 6. 互动关系图示

关系	反馈	反应
互惠性关系	+(present)	+(present)
非互惠性关系	-(absent)	-(absent)
适应性关系	-(absent)	+(present)

4.5.2. 关系范围

关系范围根据数量或范围分为“宽”范围和“窄”范围, 前者是指语言使用者需加工大量的输入信息后才能进行语言活动, 后者则是指语言使用者只需加工有限的信息输入便可以进行语言活动。大学英语四级翻译试题是基于既有文本进行汉译英翻译, 受试者需要对整个段落进行解读, 因此该试题为“宽”范围任务。这与《考试大纲》考核技能中的“语言能力”“策略能力”要求一致, 可以综合反映并测量受试者的翻译能力。

4.5.3. 关系的直接性

关系的直接性是指语言活动的顺利完成多大程度上依赖于语言使用者的输入、主题知识的以及语境的推断。关系的直接性可以分为两类, 即直接关系、间接关系。直接关系是指语言活动的进行或完成有

赖于语言使用者的输入内容, 而间接关系是指语言使用者依靠个人经验或既有知识积累完成语言活动。大学英语四级翻译试题需要受试者解读源语文本, 也考察受试者的语言、文化知识的积累, 这与《考试大纲》《教学指南》的要求具有一致性, 因此该试题具有良好的效度。

5. 结语

内容效度是语言测试领域评价测试质量的重要标准之一[18]。本研究以 Bachman & Palmer 的任务框架为理据, 以语篇输入、预期回答、预期回答与语篇输入的关系为核心维度探究大学英语四级翻译试题与《大学英语四、六级考试大纲》规定的测试目标、《大学英语教学指南》的教学目标的相关性以及拟测试构念的一致性。研究表明, 该测试内容与拟测试构念具有较高的一致性, 同时也与《考试大纲》的测试目标、《教学指南》的教学目标具有一定的相关度。然而, 在语篇输入、预期回答维度, 该测试内容具有一定的不足。

语篇输入维度: 该测试内容的题材与《考试大纲》中的考核要求高度一致, 但题材分布单一, 尚未覆盖《教学指南》中提及的其他主题, 尤其是外国文化主题。此外, 该试题体裁单一, 尚未涉及《考试大纲》中要求的记叙文、议论文, 也未涵盖《教学指南》中提及的“不同话题的口头或书面材料”。另一方面, 翻译试题难度偏低, 易造成“天花板效应”, 试题区分度降低, 无法综合测量并反映受试者的真实翻译能力。

预期回答维度: 该试题内容只设置“汉译英”, 未能涵盖《教学指南》中提及的“英汉互译”文章, 试题只能反映或测量受试者的汉译英翻译能力, 无法保证受试者综合翻译能力的测量效度。

基于任务特征框架对大学英语四级翻译试题进行深入剖析后, 本研究综合考虑实施的可操作性、可行性以及实施成本, 尝试建议分阶段优化大学英语四级翻译试题: 在题材方面, 采用“核心 + 拓展”的模式, 即以中国文化为核心, 逐步增加科技、商务、跨文化主题类文本, 并实施题材阶段更新机制, 实施成本可控且更加符合《教学指南》要求。在题型方面, 采用“渐进式”变革, 即保留原有“汉译英”题型的基础上, 增设选择性“英译汉”试题, 尝试兼顾测试时长与试题的多样性。

基金项目

本文系 2025 年新疆高校外语教学改革研究项目“讲好中国新疆故事”项目驱动下大学英语智慧教学创新路径研究(项目编号: XJGXWYJG2025A01)的阶段性成果。

参考文献

- [1] 肖维青, 熊凌崧. 翻译测试研究二十年: 现状与展望[J]. 外语界, 2023(4): 25-33.
- [2] 邹申. 语言测试[M]. 上海: 上海外语教育出版社, 2012.
- [3] 穆惠峰. 基于语料库的大学英语四级完型填空测试内容效度验证研究[J]. 外语电化教学, 2011(4): 66-70.
- [4] 李芳娜. 大学英语四级考试快速阅读理解测试内容效度分析——以 2010 年 CET4 为例[J]. 右江民族医学院学报, 2011, 33(6): 860-861.
- [5] 辜向东, 李志芳, 张书奎. 大学英语四级考试快速阅读部分内容效度研究[J]. 西南民族大学学报(人文社科版), 2009, 30(1): 258-263.
- [6] Xie, S.M., Singh, C.K.S. and Lun, W.W. (2025) A Research on Content Validity of CET-4 Reading Comprehension—Based on Test Task Characteristics Framework. *International Journal on Studies in English Language and Literature*, 12, 24-36. <https://doi.org/10.20431/2347-3134.1208004>
- [7] 谭燕保. 浅谈 CET4 作文测试部分的内容效度[J]. 纺织高校基础科学学报, 2004(1): 85-88.
- [8] Zhai, L.L. and Li, X.R. (2017) The Study of Content Validation on CET Writing Section after Reform in China. *Cross-Cultural Communication*, 13, 1-7.
- [9] 教育部高等学校大学外语教学指导委员会. 大学英语教学指南[M]. 北京: 高等教育出版社, 2015.

-
- [10] 全国大学英语四、六级考试委员会. 全国大学英语四、六级考试大纲(2016年修订版)[M]. 上海: 上海交通大学出版社, 2016.
- [11] 徐鹰, 邓雅玲. 大规模语言测试翻译任务内容效度研究——以 CET 为例[J]. 山东外语教学, 2023, 44(3): 26-37.
- [12] 刘佳慧. 2021 年英语专业八级考试翻译试题内容效度研究[J]. 现代英语, 2021(20): 43-45.
- [13] 余志谋. 大学英语四、六级考试翻译试题效度研究[J]. 淮南职业技术学院学报, 2019, 19(3): 93-94.
- [14] 唐友军, 马晓梅. CET-4 翻译测试中源语翻译文本内容效度研究[J]. 翻译研究与教学, 2020(2): 69-76.
- [15] 母燕芳. 阅读理解测试中句子翻译内容效度的定量分析[J]. 中北大学学报(社会科学版), 2009, 25(6): 56-59.
- [16] 郑美玲. 以多面 Rasch 模型对一堂翻译测试的效度验证[J]. 山东农业工程学院学报, 2014, 31(6): 164-166.
- [17] Bachman, L.F. and Palmer, A.S. (1996) *Language Testing in Practice*. Oxford University Press, 53.
- [18] 韩宝成, 罗凯洲. 语言测试效度及其验证模式的嬗变[J]. 外语教学与研究, 2013, 45(3): 411-425+481.