

# 英语写作测试效度研究

姜蕾 著

高等教育出版社

# 英语写作测试效度研究

姜蕾 著

高等教育出版社·北京

## 图书在版编目 (CIP) 数据

英语写作测试效度研究 / 姜蕾著. -- 北京: 高等教育出版社, 2018.12

ISBN 978-7-04-044369-1

I. ①英… II. ①姜… III. ①英语 - 写作 - 研究

IV. ①H315

中国版本图书馆 CIP 数据核字 (2018) 第 294333 号

策划编辑 谢 森 王春玲 责任编辑 刘维伟 封面设计 王 洋 版式设计 魏 亮  
责任校对 汪于祺 责任印制 田 甜

出版发行 高等教育出版社  
社 址 北京市西城区德外大街 4 号  
邮政编码 100120  
印 刷 三河市吉祥印务有限公司  
开 本 787mm × 1092mm 1/16  
印 张 12  
字 数 196 千字  
购书热线 010-58581118  
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.hepmall.com.cn>  
<http://www.hepmall.com>  
<http://www.hepmall.cn>  
版 次 2018 年 12 月第 1 版  
印 次 2018 年 12 月第 1 次印刷  
定 价 35.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换  
版权所有 侵权必究  
物 料 号 44369-00

## 前 言<sup>1</sup>

写作能力是一项十分重要的语言输出能力,国内外各种语言测试都十分重视对写作能力的考量。本研究有两个目标:一是提出一个可应用于国内大学英语教学的读写结合写作任务的设计框架;二是通过实验验证读写结合写作任务的可操作性及其信度和效度。验证研究包括发现读写结合写作过程中可能出现的问题,应用于不同层次及进行大规模应用时的信度和效度等,并通过收集数据并采用定量分析的方法来提供解释依据。

90名非英语专业学生参与了实验,他们分层抽样于大学一年级学生、二年级学生和研究生一年级学生,每个年级各30人。研究工具是独立写作测试任务和读写结合写作测试任务以及问卷调查。信度检验和相关检验结果显示,该写作测试任务既适用于测试不同层次的学生,也适用于大规模的课堂写作能力测试。本研究还对学生在读写结合写作任务和独立写作任务中的作文进行了对比分析,包括评分员从内容、结构、语言准确性和语言复杂性几个方面给出的分数对比,也包括使用计算机软件提取的语言复杂性和流利性相关数据的对比。对比分析结果与问卷调查结果基本一致,阅读材料对写作有一定的促进作用,学生在读写结合写作测试任务中表现更加出色。本研究对大学英语课堂测试和大学英语学期测试中的写作任务设计具有一定的启示和参考作用。

本书是在笔者博士论文的基础上修改而成的,博士论文的完成得到了很多人的热心帮助。首先,衷心地感谢笔者导师邹申教授在论文选题、资料收集、实验方案设计、论文撰写及修改的各个环节给予的专业指导,是邹老师精湛的专业知识、丰富的测试项目经验、严谨的治学态度,引领笔者走进了语言测试研究。能够师从邹申老师,是笔者一生的荣幸。求学期间,有幸聆听梅德明、陈坚林、许余龙等几位教授的课程和教诲,这为笔者的语言学知识积累和后来的语言学研究打下了坚实的基础。

感谢东北大学的赵春曦、姜雪和赵睿三位老师以及他们可爱的学生,他们的热情帮助和积极配合保障了本研究实验数据的提取。感谢东北大学

---

1 本书为国家哲学社会科学基金语言类重点项目“基于语料库的大学英语语言能力标准特征参数研究”(项目编号13AYY005)资助研究成果。

的李碧玉、刘迪、张月和王婷等老师，他们热心慷慨地参与了评分工作，使得本研究得以顺利完成。

特别感谢东北大学赵雯教授，与她的讨论常常激发笔者论文写作的灵感，她的鼓励和支持帮助笔者克服了很多困难。感谢东北大学刘武教授，当笔者在统计方面遇到问题的时候是他为笔者指点迷津。感谢刘芹教授、刘宝权教授和徐永教授，他们的博士论文为笔者的论文选题和实验设计提供了重要的启示，他们也为笔者的论文修改和完善提出了很多宝贵建议。感谢高等教育出版社的编辑们，他们在审校期间一丝不苟的工作态度令笔者十分感动。

特别感谢家人在笔者攻读博士学位期间所给予的支持、鼓励和牺牲，他们是笔者即使遇到困难也不言放弃、不断前行的动力。

本书对大学英语课堂写作测试任务设计和效度进行了系统、科学的研究，也是针对英语读写结合写作测试任务的综合性研究，具有一定的理论价值和实践意义。但由于作者水平有限，虽经多次修改，其中谬误仍然难以避免，敬请广大读者提出宝贵意见和建议。

姜 蕾

2018年11月

## Abstract

Where there is language instruction, there is language testing. Effective language testing measures how much teaching goals have been achieved and how much students have learned, and has positive washback effects on instruction as well. Writing is regarded as important language output. Writing test is valued as one of the most scientific and effective ways to evaluate learners' English language proficiency, and thus has become an important and necessary component of many large-scale English tests. Both TOEFL iBT (Internet-based Test) and academic writing assessment in American universities include integrated writing tasks, which gives inspiration to writing assessment in Chinese universities.

Two goals are expected to achieve in this study. On the one hand, this study focuses on review and overview of theories on writing and writing assessment, after which a framework for designing reading-to-write tasks is proposed to be used in college English classrooms in China. Theoretical foundations of the reading-to-write task include the exploration of nature of writing and writing process, explanation of constructs of writing ability and their measurable variables, classification of writing tasks and their merits and demerits, examination of teaching requirements and testing syllabuses of CETs, and interpretation of Bachman and Palmer's authenticity and interactiveness. Essence of classroom assessment and test specifications are described, on the basis of which the reading-to-write task is designed. On the other hand, experimental tests are carried out to verify practicality, reliability and validity of the reading-to-write task. The validation study is composed of finding possible problems of the reading-to-write task, testing its reliability and validity when applied to different levels and on a large scale, and identifying the difference of students' performance in the reading-to-write task and the independent writing task through the comparison of discourse. Quantitative data is collected and analyzed to provide interpretation.

In nature, writing is a linguistic, social-cultural and cognitive activity. Written language keeps the features of formality, accuracy and complexity in contrast to that of speaking. Writing is a means of communication

and is influenced by social and cultural conventions. Writing is also a cognitive process in the writer's mind, recursive rather than linear. Expert writers tend to involve knowledge transforming, while novice writers are likely to generate texts in the way of knowledge telling. The writing process is mainly the interaction between the individual writer and the task environment, influenced by the individual's working memory, motivation and affect, cognitive processes, and long-term memory. Hymes (1972), Canale and Swain (1980), and Bachman (1990) divide language knowledge into three types: linguistic knowledge, discourse knowledge and sociolinguistic knowledge. Later, Bachman and Palmer (1996) introduce strategic competence into communicative language ability. Built on their work, writing ability is summarized by the author as linguistic knowledge, discourse knowledge and strategic competence. Writing ability is multi-componential and its dimensions can be measured by complexity, accuracy and fluency of the language (Skehan, 1998; Ellis, 2008). Backed by their study, the author summarizes ten quantitative measures, which are average word length, type/token ratio of words, average number of words per T-unit, percentage of clauses to T-units, percentage of dependent clauses of total clauses, percentage of error-free T-units, percentage of error-free clauses, average number of words per text, average number of T-units per text, and average number of clauses per text. A review of large-scale writing assessment home and abroad finds that the writing task is generally divided into the independent task and the integrated task. And the integrated writing task is broken down into the reading-writing integrated task and the listening-reading-writing integrated task. Merits and demerits of each type are discussed and the reading-to-write task is proposed to be used for English classroom assessment in universities in China (the reading-to-write task is more feasible than the listening-reading-to-write task for English teachers in the classroom).

The reading-to-write task aims to provide theoretical guide and practical reference for college teachers to design writing tasks for classroom assessment, and therefore, it should be driven by teaching requirements and testing syllabuses of CETs. Teaching requirements describe writing in three

levels, while testing syllabuses define topics, genres, words, time and marking criteria. Bachman and Palmer's (1996) authenticity and interactiveness lay theoretical foundation for reading-to-write tasks in that the integrated writing task mirrors the academic or professional task in the real life, and interacts with students' interest, topic knowledge, and linguistic competence and strategic competence, and therefore, facilitates students' performance in writing at an optimal level.

Designing the reading-to-write task involves discussion of the essence of classroom writing assessment and description of test specifications. It's an achievement test and a criterion-reference test as well, evaluating the effectiveness of teaching and learning. The holistic scoring is widely used and quick; while the analytic scoring is able to provide detailed reports but time-consuming. On the basis of Weigle's (2007) summary of writing dimensions and Bachman and Palmer's (1996) account of characteristics of input, the author sets forth a detailed framework, according to which sample tasks are designed.

The reading-to-write task undergoes two stages of experimental tests, which are conducted to test its practicality, reliability and validity.

The pilot study mainly concerns the possible problems of the reading-to-write task application and its reliability and validity. 30 candidates (freshmen, non-English majors) in Northeastern University participated in the study. They were asked to take part in a writing test of a reading-to-write task and finish a candidate's questionnaire immediately after the writing test. Their writings were analytically marked by two experienced raters. The raters were asked to finish a rater's questionnaire immediately after rating. The average score of two analytic ratings was the core parameter in the study. The candidates' scores in CET4 (including the total score and the score of writing in CET4) and that of writing exercises in English class were used as external measures. Parameters such as the reliability quotient alpha and correlation coefficients were tested and satisfactory results were gained. First of all, the reliability quotient alpha of the reading-to-write task is 0.704, above the theoretical requirement of writing tests. Secondly, the sub-total correlation coefficients are 0.781–0.860 and those of sub-sub correlation are



0.473–0.639, which demonstrate acceptable construct validity of the reading-to-write task. Thirdly, the correlation coefficients of RA-CET4W (0.726) and RA-SWE (0.710) indicate desirable criterion-related validity of the reading-to-write task. What's more, the results of the questionnaires show most students and all teachers hold positive attitude towards the reading-to-write task, believing that the reading-to-write task is able to evaluate Chinese college students' English ability of writing. Therefore, it can be concluded that the reading-to-write task exhibits reasonable reliability and validity in application and it is applicable for classroom use.

The field study focuses on the reliability and validity of the reading-to-write task when it is applied to different levels and on a large scale, and the comparative study with the traditional independent writing task. 90 candidates (non-English majors) took part in the study, including 30 freshmen, 30 sophomores and 30 graduates. The stratified random selection was intended to represent all levels of students who are taking English as a compulsory course in most Chinese universities. All candidates were required to take an independent writing task first, a reading-to-write task of the same topic one month later, and a questionnaire immediately after the second task. Their writings in two tasks were all analytically marked by two experienced raters. The average score of two analytic ratings were taken as the core component of the study. Similar to the pilot study, tests of reliability and correlation were carried out to demonstrate the following findings. On the one hand, when applied to different levels of candidates, the reading-to-write task exhibits reasonable reliability quotient alphas (0.718, 0.753 and 0.701 respectively) and acceptable construct validity (with satisfactory sub-total correlation coefficients at 0.736–0.764, 0.841–0.906 and 0.778–0.873 respectively and generally acceptable sub-sub correlation coefficients at 0.278–0.560, 0.535–0.817 and 0.467–0.677 respectively). On the other hand, when applied on a large scale, the reading-to-write task also displays the satisfactory reliability quotient alpha (0.713) and acceptable construct validity (with sub-total correlation coefficients at 0.790–0.843 and sub-sub correlation coefficients at 0.441–0.676). It can thus be concluded that the

reading-to-write task would be applied in classroom assessment to different levels of students, as well as on a large scale, to evaluate Chinese college students' English ability of writing with acceptable reliability and validity.

Furthermore, discourse-based comparative studies on 180 essays by 90 candidates are made to look for similarities and differences of candidates' performance in these two writing tasks. For one thing, the subjective analytic measurement was analyzed from aspects of content, organization, language accuracy and language complexity; for another, the objective discourse analysis was examined in terms of language complexity and fluency. Both the stratified sample comparison and the entire sample comparison were under discussion. Paired sample t-tests show that candidates tend to perform better in the reading-to-write task. On the one hand, the entire sample generally obtains higher scores (including the total score and those of sub-categories) in the reading-to-write task; the freshman candidates tend to perform better in language complexity and the total score; the sophomore candidates tend to perform better in language accuracy, language complexity and the total score; while the graduate candidates tend to perform better in all four aspects and the total score. On the other hand, both the stratified sample and the entire sample are likely to use longer words, make longer sentences (except sophomore candidates) and produce more percentage of complex sentences (CS) and compound-complex sentences (CCS). The percentages of usage of passive voice and nominalization are both doubled. On the whole, the fluency is enhanced by using more words, more sentences, more percentage of CSs and more CCSs. However, both the stratified sample and the entire sample demonstrate lower word type/token ratio in the reading-to-write task, indicating that candidates are likely to exhibit same amount of vocabulary in these two tasks. And to some extent, findings of the comparative study conform to those of questionnaires. Results of the questionnaires show that students mostly hold that the reading material in the reading-to-write task help them produce better writing. For freshman candidates, facilitation would occur to all four aspects, including content, organization, language accuracy and complexity. For candidates of higher levels, ideas for content and organization may be highly promoted. Therefore, it can be concluded

that the reading-to-write task might well help college students produce better essays.

In conclusion, this study proposes designing the reading-to-write task for classroom assessment on the basis of theoretical reviews and verifies its reliability and validity by experimental tests. This empirical study attempts to promote a change in writing assessment in college English classrooms and is intended to provide reference and implication to English teachers. It is hoped that this framework will have positive washback effects on writing instruction in English in Chinese universities.

**Key words:** reading-to-write task, validation study, reliability, construct validity, criterion-related validity

## 摘 要

有语言教学的地方就有语言测试,有效的语言测试不仅可以测量教学效果和学生成就,对相关教学也起到积极的反拨作用。写作能力是一项十分重要的语言输出能力,被认为是检验学习者英语语言水平的最科学、最有效的途径之一,是许多大型语言测试的重要和必要组成部分。观察国外写作测试研究,作者发现有影响力的托福iBT考试和美国大学的学术写作考试都采用综合写作测试的形式,其成功经验对国内高校的写作测试具有一定的启示作用。

本研究有两个目标:一是对写作相关理论进行全面的回顾和综述,之后提出一个可应用于国内大学英语课堂的读写结合写作任务设计框架。读写结合写作任务的理论基础包括写作本质和写作过程的探讨,写作能力的构念和可测量变量的分析,写作任务类型和相应的优缺点,也包括对大学英语教学要求和大学英语考试大纲的重新解读和对Bachman&Palmer的真实性和互动性概念的重新认知。框架的设计基于对课堂测试本质和测试规范的描述。二是通过实验验证读写结合写作任务的可操作性及其信度和效度。验证研究包括发现读写结合写作过程中可能出现的问题,测试读写结合写作任务应用于不同层次及进行大规模应用时的信度和效度等,同时也对学生在读写结合写作任务和独立写作任务中的作文语料进行对比分析,试图发现他们在两种任务中的表现差异。主要通过收集数据并采用定量分析的方法来提供解释依据。

从本质上讲,写作是语言行为,是社会文化行为,也是认知行为。书面语言和口语相比有正式、准确和复杂等特点;作为一种交际手段,写作的表现形式受社会、文化规约的限制;写作过程不仅是个体表达思想的过程,更是思想在大脑内进行认知努力和认知加工的过程,这个过程是非线性的循环过程。有经验的作者运用知识转述模式,而没有经验的新手则采用知识直述模式,写作过程是作者个体与写作环境进行互动的结果,受工作记忆、动机和情感、认知过程和长期记忆的影响。Hymes (1972)、Canale&Swain (1980)和Bachman (1990)把语言知识分为语言知识、语篇知识和社会语言知识三种。后来Bachman&Palmer (1996)又引入策略能力来解释交际语言能力。在前人理论研究的基础上,作者把写作能力概括为语言知识、语篇知识和策略能力。写作能力是多元的,其维度可通过

语言的复杂性、准确性和流利性来衡量 ( Skehan, 1998; Ellis, 2008 )。根据前人的研究, 作者总结出十个量化指标, 即平均词长、类符/形符比、T单位词比、分句/T单位比、从属分句/分句比、无错误T单位比、无错误分句比、平均词数、T单位数、分句数。回顾国内外的大规模测试发现, 写作任务可大致分为独立写作和综合写作两种。综合写作又分为读写综合写作和听读写综合写作。作者举例分析了每种类型的优缺点, 并提出尝试在大学英语课堂环境下实施读写结合写作测试的设想 ( 相较于听读写任务, 读写任务的设计对普通教师更具可操作性, 因此其研究结果可能更有推广性 )。

本文提出读写结合写作测试任务的设计框架, 旨在为大学英语教师在课堂范围内设计写作任务提出理论指导和实践参考, 因此, 应在《大学英语课程教学要求》和《全国大学英语四、六级考试大纲》的指导下进行。《教学要求》从三个层面描述了写作能力。《考试大纲》对主题、题材、字数、时间和评分要求做了说明。Bachman&Palmer (1996) 的真实性和互动性理论为读写结合写作任务奠定了基础, 因为它反映了现实生活中的学术或专业任务, 与学生的兴趣、主题知识、语言能力和策略能力相互作用, 促进学生在测试任务中有最优表现。

读写结合写作任务的设计包括对课堂写作测试本质的分析和测试规范的描述。课堂测试属于成就测试, 同时也是标准参照测试, 是对教学有效性的检验。综合评分法应用广泛且使用快捷, 分析评分法能够提供详细的诊断信息但耗时费力。结合Weigle (2007) 对写作测试任务维度的总结以及Bachman&Palmer (1996) 对测试任务输入材料特点的描述, 作者提出了详细的框架并据此设计了相关示例任务。

实证研究的目的是验证该写作测试任务在应用时的可操作性及其信度和效度, 分两个阶段进行。

第一阶段是小范围的试点研究, 旨在发现读写结合写作任务在应用时可能存在的问题以及其信度和效度。东北大学的30名非英语专业本科一年级学生参加了本次研究, 他们需要参加读写结合写作测试并在测试后立即完成调查问卷。参加测试的学生的作文由两位经验丰富的教师采用分析评定法评分, 教师们在评分结束之后, 随即完成了教师调查问卷。分析评定平均分是主要研究参数。这些学生同期参加的大学英语四级考试成绩 ( 包括总成绩和写作成绩 ) 和他们的英语课堂写作练习成绩为标准关联效标。效度验证主要包括信度检验和相关检验, 得出结论如下。首先, 本任务的

信度系数为0.704, 高于写作测试的理论标准。其次, 部分—整体相关系数为0.781—0.860, 部分—部分相关系数是0.473—0.639, 显示可接受的构念效度。此外, 四级考试写作成绩分析评定平均分和英语课堂写作成绩的相关系数分别为0.726和0.710, 表明理想的效标关联效度。同时, 问卷结果显示参加本研究的大多数学生和所有教师对该任务持肯定态度, 普遍认为其能够合理地测评中国大学生的英语写作能力。因此, 读写结合写作测试显示了符合理论要求的信度和效度, 可在大学英语课堂中应用。

第二阶段的研究涉及读写结合写作任务在不同层次及大规模应用时的信度和效度以及传统的独立写作任务的对比研究。90名非英语专业的学生参加了本次研究, 他们分层抽样于一年级学生, 二年级学生和研究生一年级学生, 每个年级各30人。分层随机抽样是为了能代表大多数中国大学中参加英语必修课的所有水平的学生。研究工具是独立写作任务和读写结合写作任务以及调查问卷。实验分两个步骤进行; 第一步, 参加独立写作测试; 第二步, 一个月后, 参加读写结合写作测试, 题目与独立写作测试相同, 同时完成调查问卷。参加测试的学生的两份作文同时由两位有经验的教师采用分析评定法评分, 平均分是主要研究参数。同初步研究一样采用信度检验和相关检验并得出以下结论。首先, 当该写作测试任务应用于不同层次的学生时, 显示合理的信度系数(分别是0.718, 0.753和0.701)和可接受的构念效度[部分—整体相关系数十分理想(分别为0.736—0.764, 0.841—0.906和0.778—0.873), 部分—部分相关系数基本可接受(分别为0.278—0.560, 0.535—0.817和0.467—0.677)]。其次, 当写作测试任务进行大规模应用时, 信度系数(0.713), 构念效度[部分—整体相关系数(0.790—0.843), 部分—部分相关系数(0.445—0.676)]总体符合理论要求。因此, 该写作测试任务既适用于测试不同层次的学生, 也适用于大规模地应用在课堂上的写作能力测试, 以可接受的信度和效度测评中国大学生的英语写作能力。

此外, 本文也对学生在读写结合写作任务和独立写作任务中完成的作文进行了对比分析(90人, 180篇), 试图发现他们在两种任务中的表现差异。一方面, 从内容、结构、语言准确性和语言复杂性几个方面进行了主观分析的对比; 另一方面, 从语言的复杂性和流利性等方面进行了客观语料对比。对比的样本包括分层对比和整体对比。配对样本t检验显示学生们倾向于在读写结合写作任务中有更好的表现。其一, 整体均在读写结合任务中获得更高的分数(包括总分和分项小分); 分层对比时一年级学

生倾向于在语言复杂性和总分上表现更好，二年级学生倾向于在语言准确性、语言复杂性和总分上表现更好，而研究生在总分和四个分项中均表现很好；其二，分层和整体均在读写结合任务中倾向于使用更长的单词、造更长的句子（二年级学生除外），更多地运用复合句和复杂句，更高比例地使用被动语态和动词名词化（大约2倍），同时通过使用更多的词汇、句子、更大比例的复合句和复杂句，流利性也在整体上进一步增强。然而，分层和整体均显示出更低的类符/形符比，说明在两种写作任务中学生使用的词汇量是相似的。对比分析结果与调查问卷结果基本一致。调查问卷的结果显示，学生普遍认为读写结合任务中的阅读材料对写好作文有一定帮助。一年级学生反映这些帮助体现在作文内容、结构、语言准确性和语言复杂性四个方面，而高年级学生认为帮助可能更多地表现在作文的内容、结构上。因此，可以断定读写结合写作对大学生写出更好的文章应该有一定帮助。

总之，本文基于理论综述提出读写结合写作测试任务的设计框架并通过实证研究验证了其信度和效度，期望这个尝试对大学英语课堂写作测试带来一定改变并对英语教师提供一定的参考和提示，并对大学英语写作教学产生积极的反拨作用。

**关键词：**读写结合写作测试任务；效度研究；信度；构念效度；效标关联效度

## List of Abbreviations

BEC	Cambridge Business English Certificate
CAF	complexity, accuracy and fluency
CCS	compound-complex sentence
CCS/S	compound-complex sentences to total sentences
CET	College English Test
CET4	Band 4 of CET
CET4T	total score in CET4
CET4W	score in CET4 writing component
CET6	Band 6 of CET
CET6T	total score in CET6
CRT	criterion-referenced test
CS	complex sentence
EFC/C	error-free clauses of all clauses
EFT/T	percentage of error-free T-units in a text
ELTS	English Language Testing Service
ESL	English as a Second Language
GT	general training
IELTS	International English language Testing System
NEU	Northeastern University
PETS	Public English Test System
R1	score given by the author or the first analytic rater
R2	score given by the other analytic rater
RA	average score of R1 and R2
SE	score in the English term examination
SPSS	Statistical Package of Social Sciences
SWE	score in writing exercises in English class
TEM	Test for English Majors
TOEFL	Test of English as a Foreign Language
TOEFL iBT	TOEFL Internet-based Test
UCLES	University of Cambridge Local Examinations Syndicate
UK	The United Kingdom of Great Britain and Northern Ireland



