

基于文献计量的国内语言测试研究现状分析

◆徐腊梅

(四川工商学院, 四川 眉山 620000)

【摘要】文献计量分析(Bibliometric review)作为新兴的二手文献研究方法,近年来在应用语言学领域内获得了极大的关注。然而,国内对于其相关介绍还十分稀少。本文应用文献计量分析法,对国内语言测试研究现状进行了分析。以语言测试与评价专业委员会第二届候选名单为具体研究样本,总结了委员会中研究者的区域背景、发表成果与项目研究的主题。本文发现,国内过半的语言测试研究者群居在广东、北京和上海一带,研究主题主要集中于测试体系的开发和量表相关研究,对于诊断反馈、能力构建和考试的社会性影响等也有大量关注。本文介绍的文献计量分析法可以为研究者提供二手文献分析的新思路,也助于他们快速了解国内语言测试研究现状。

【关键词】语言测试;文献计量分析;研究现状

1 语言测试研究现状分析背景

近年来,应用语言学界对研究综述(research synthesis)的关注度在不断的提升。国际知名 SSCI 期刊 Applied Linguistics Review 在 2023 年 1 月发行了专刊,专门介绍了研究综述的发展现状。在专刊的 5 篇文章中,以 Chong and Plonsky(2023)的论文尤其突出,他们总结了研究综述的 11 种类型,并从研究内容、综述过程、文章结构和结果呈现四个角度对这 11 种类型的综述研究方法进行了评价。其中最引人瞩目的是文献计量分析(Bibliometric review),它是指对研究的引用数、作者的特质(数量,性别,语言和区域背景),文章的标题和关键词等内容进行数据分析的方法。然而,国内对于该方法的介绍与应用相对短缺。因此,本文选取语言测试与评价专业委员会第二届候选名单为例,使用文献计量分析的方法来分析研究者的区域背景、发表成果和研究项目,总结并概括了近年来国内语言测试的研究重心。

2 基于文献计量的国内语言测试研究现状分析方法

根据官网的介绍,语言测试与评价专业委员会(CACSEC Association for Language Testing and Assessment,简称 CALTA)是中国英汉语比较研究会的分支机构,该组织由从事语言测试与评价研究者组成,主要致力于组织语言测评研究相关活动。为了更好地展开工作,该组织会每两年投票选举理事会成员。委员会候选名单上的研究者需要满足以下一个条件:至少发表一篇 CSSCI 期刊或国际期刊;具有硕士或博士研究生导师头衔;主持省部级以上科研项目一项以上。考虑到语言测试与评价专业委员会的权威性以及名单上榜的要求,选择该名单做为分析对象具有典型性。然而,由于名单中的限制条件(最多两项发表的成果,最多一个主持的项目),该名单仅仅展示了语言测试研究者最有代表性的成果而非他们的全部研究。2023 年 6 月 11 日公示

的语言测试与评价专业委员会第二届候选名单提名了 71 名候选人,该名单包括研究者姓名、所属院校、职称、近五年发表的成果和主持的项目。

3 基于文献计量的国内语言测试研究现状分析结果

3.1 研究者的区域背景

经过数据统计发现,71 位语言测试研究者来自 49 所高校,其中广东外语外贸大学最多(7 人),排名第二是北京语言大学(4 人),并列第三的是华南师范大学和上海外国语大学(均为 3 人),其他院校的人数都不超过 2 人。从省份区域来看,排名前三的分别是广东(17 人),北京(14 人),上海(11 人),其他的省份人数均不超过 5 人。区域背景分析显示过半的语言测试研究者呈现出群居态势,而且往往集中在经济发达、资源充沛的区域,例如广东、北京和上海等地。

3.2 发表成果

3.2.1 成果数量和质量分析

候选名单中发表成果包含论文、专著和研究报告。就发表论文而言,研究者们报告了 48 篇英文文献,其中 SSCI 级别 43 篇,ESCI 级别 2 篇,普通期刊 3 篇。国际期刊发表的整体质量非常优秀,SSCI 期刊的占比达到了 89.58%,而且均为一区和二区的高质量期刊。其中最受青睐的是 Frontiers in Psychology(6 篇),其次是 Assessing writing, Language Assessment Quarterly, Studies in Educational Evaluation, Language Testing(均为 5 篇),再次是 The Asia-Pacific Education Researcher 和 Modern Language Journal(均为 3 篇),其他的期刊数量不超过 2 篇。近五年内国内研究者报告发表了 65 篇中文论文,其中核心 49 篇,国家级 8 篇,省级 6 篇,普刊 2 篇。整体的发文质量也很优秀,核心占比达到了 75.38%,其中发文数最多的期刊为《外语界》(12 篇),其次为《外语教学》和《现代外语》(均为 8 篇),再次

为《语言教学与研究》和《中国考试》(均为4篇),其余期刊发文均不超过3篇。

此外,研究者们报告发行了专著共21部,其中来自一级出版社15部,国际出版社1部,其他出版社21部。一级出版社的比例达到了71.43%,但是国外出版社仅有1部,出版专著的国际化程度不够高。同时还发现,语言测试研究者涉及的研究报告只有一项,为国家留学基金管理委员会资助的中国政府奖学金本科来华留学生预科结业“中医汉语考试”首次考试分析报告。因此,语言测试领域内研究报告的撰写还有较大的空间可以挖掘。

3.2.2 成果主题分析

通过统计各个主题的出现频率发现,第一梯队的研究主题是测评体系开发和量表研究。测评体系的开发主要分为综合性研究、传统技能研究和应用技能研究三类。综合性研究涵盖面较广,比如中考命题、高考效度、大学英语测试设计、四六级英语考试和托福雅思的效度对比,还有效度历史等。传统技能研究涵盖了听说读写四项技能,例如阅读考试效度、听力考试设计、口语测试设计(例如口音对打分的影响)、写作测试设计(例如写作者的声音)、综合阅读以及读后续写等。应用技能研究则更加偏重测试使用者对语言的综合使用能力,例如语用能力、口译能力、笔译能力、手语测评和内容语言融合测评等。除了英语综合力量表效度研究外,本文发现语言测试研究者们对传统力量表也做了相关研究,例如阅读力量表的研究中的阅读策略、口语力量表研究中的口语语言特质、英语写作力量表研究中的写作语言特质等。发表的研究主题还包括应用技能的力量表研究,例如语用量表、口译量表、自评量表等。此外研究还包括量表的对接,比如四六级和量表对接、口语和写作量表对接的效度。但是对于听力量表的研究却没有发现。

第二梯队的研究主题包含诊断和反馈、能力构建、考试的社会性影响以及动态评估。诊断研究主要集中在阅读认知诊断、听力诊断、写作诊断以及比较新颖的自适应性认知诊断。反馈研究主要集中在听力和写作的反馈,反馈的作用和同伴反馈也有涉及。然而,诊断和反馈研究主要集中在阅读、听力和写作,对于口语的研究则没有发现。能力构建主要研究使用者的阅读能力素养和学术听力能力,也包含了综合能力的构建,例如,测试素养、语用能力、法律翻译能力、职业英语能力、批判思维的能力以及很新颖的数字多维创作能力。其中语言测试素养是一个研究的重点,得到了很多研究者的关注,但主要集中在理论综述部分。考试的社会性影响研究主要包含考试的社会性研究,例如它的历史、测试价值观、真实性、学科背景差异以及反拨作用。其中反拨作用是一个热点趋势,主要研究考试对学生学习动力、学习方法、学习过程和结果、考试准备以及对于教师教

学的影响。动态评估也是最近新兴的研究热点,相关研究主要涉及听力和口语的动态评估,也涉及考察语言使用者的学术能力和元认知能力。

第三梯队的研究则更为零散,主要包含课堂内的评估,例如以学习为导向的评估、同伴测评和语文课堂评价等,其中还涉及了课堂参与评分,它的研究结果对课堂表现的评价具有很大的现实意义。信息技术在测试中的运用也日益凸显,例如自动评分在翻译和写作考试中的运用、线上评估,以及口语测试语料库研究等。研究者们还研究了形成性评价在实践中的困境和在翻译测试中的应用,以及表现性评价的实践操作和自评工具等。值得一提的是,除了英语之外,发表成果也涉及对外汉语和法语研究,研究的主题分布与英语类似,也主要聚焦于汉语力量表的开发、汉语考试的设计(例如写作考试的设计和分析),以及法语口语和教学的评估。

3.3 研究项目

从项目数据统计中发现,71位语言测试研究者在近五年内立项共39项,拥有项目的人数占到总人数的54.93%。其中国家级项目15项,省部级20项,市厅局级别项目3项,其他级别项目1项。尽管候选名单明确要求填写省部级及以上项目,但仍有研究者填写了市厅局级别项目。

研究项目的主题也可以分为三个梯队。第一梯队涉及的是测评体系的开发和量表研究。测评体系的开发涉及面比较广,主要包含的是综合能力的测评系统开发,例如外语综合能力测评、沟通能力测评,以学习为导向的评价、表现性评价、学术英语学习体系、义务教育试题研究、普高英语课程评价、对外文化传播语言人才培养体系评价和STEAM教育的课程评价等。当然,研究者也涉及了传统技能的测评,例如口语语音评估、读后续写、高考写作效度研究以及阅读自评系统等。最后,项目也研究了英语的应用能力,例如英汉互译策略评价,交替传译策略能力评估。可是研究立项中关于量表相关的只有一项,开发的是青少年英语阅读力量表。

第二梯队的项目研究主题包含三个方面,分别是能力构建、诊断和反馈以及考试的社会影响。其中能力构建主要聚焦传统技能,比如口语能力构建、英语写作过程、英语写作的元认知能力、青少年的英语阅读能力等。此外,能力构建研究也包含了对语言测试者的测评素养研究。关于诊断和反馈研究主要涉及大数据成绩反馈、诊断评估与干预、诊断测评与个性化反馈,其中,阅读认知诊断测试的开发和设计是一个研究重点。针对考试的社会性影响研究主要包括语言测试公平性、考试价值和口语考试对教学的反拨效应。

第三梯队的研究主题与测试相关性不高,例如翻译项



目、翻转课堂教学、来华留学政策研究、考试招生专项、高校外语教师发展以及老年人互动能力评估。和发表成果主题类似,大部分项目都以英语为主,非英语项目十分稀缺。其中,非英语语种的测试项目以对外汉语为主,且仅有三项,分别是汉语水平考试效度、汉语水平快速测试以及中文学习历时研究。

4 基于文献计量的国内语言测试研究现状分析结果讨论

对比发表成果和项目的研究主题发现,第一梯队和第二梯队的研究主题大致相同,其中量表相关研究和测试体系开发研究毫无疑问是近五年的研究热点。但是,在项目研究中关于量表的研究却十分稀少,这似乎意味着英语能力等级量表(CSE)的开发已经进入尾声,而测试体系开发工作还在继续。第二梯队的研究主要涉及诊断反馈、能力构建和考试的社会性影响方面。在研究发表主题中,动态评估也是很热的研究话题。然而,对比第三梯队的研究,发表成果和项目主题展示出巨大差别,发表成果聚焦更加细化的测试主题,而且项目则更加发散,涉及众多非测试相关的主题。这说明语言测试研究者的立项范围更加广泛,不限于语言测试学的范畴。

通过研究对比,总结出一些未来的研究方向。第一,关于听力量表的相关研究在近五年内还是一个空白,需要更多的关注。第二,虽然诊断反馈的研究在国内逐渐得到了关注,但是它往往聚焦的是阅读、听力和写作,对于口语的诊断反馈研究还是很稀缺。第三,随着人工智能的发展,科技在语言测试中的地位将不断升高,然而,国内的相关研究还是十分缺乏,形成的研究成果也不多。第四,国内的语言测试研究主要关注的是英语,非英语语种的研究十分缺乏,这一点也值得引起大家的关注。

5 结束语

本文采用了文献计量分析方法,对语言测试与评价专业

委员会第二届候选名单进行了分析,总结了语言测试研究者的区域背景以及他们的发表成果和项目研究主题。研究发现,国内的研究群体主要集中在广东、北京和上海一带,研究主题最集中于测试体系的开发和量表相关研究,对于诊断反馈、能力构建和考试的社会性影响等也有大量的关注。本文为相关研究者提供了一种新的二手文献研究思路,也有助于他们快速了解国内近些年的研究现状。

参考文献:

- [1]肖杨田.中学英语教师测评素养自评量表开发与验证[D].上海:上海外国语大学,2020.
- [2]张露露.我国教师测评素养研究述评[J].中国考试,2018(08):69-73.
- [3]肖杨田,何琼.英语国家语言教师测评素养量表:回顾与比较[J].外语测试与教学,2021(02):38-49.
- [4]周海晨,章成志,胡志刚,等.全文计量分析的实践与展望:理论、方法与应用——2022全文本文献计量分析学术沙龙综述[J].信息资源管理学报,2023,13(02):135-142.
- [5]张法连,蒋毓婧.国内外法律语言研究现状对比分析(1998—2021)——基于可视化技术和文献计量分析方法[J].当代修辞学,2023(02):29-42.
- [6]徐新民,李艳莉.近十年国际成人教育研究方法的分析与反思——基于三本国际成人教育期刊的文献计量分析(2012—2022年)[J].中国成人教育,2022(24):3-8.

作者简介:

徐腊梅(1990—),女,汉族,辽宁丹东人,硕士,助教,研究方向:外语教学、外语测试、教师教育。