

# 国内外外语教学课堂评价研究可视化分析 (2008—2020)<sup>1</sup>

杨莉芳 王婷婷

北京外国语大学 苏州工业园区星汇学校

© 2022 外语教育研究前沿 (原名《中国外语教育》) (2), 34-41 页

**提 要:** 本文以 CiteSpace 5.7.R2 为分析工具, 运用文献计量法梳理了 2008—2020 年 Web of Science 核心数据库中外语类期刊、CNKI 中 CSSCI 来源期刊和北京大学中文核心期刊中的外语教学课堂评价研究相关文献, 通过对关键词的可视化分析, 呈现近 13 年来国内外外语教学课堂评价研究的核心研究主题及研究热点变化。研究结果显示, 国内外共同关注的研究主题有反馈、学生自评与同伴互评和写作评价。国内研究关注课堂评价的效果, 国外研究侧重课堂评价的过程。在研究热点的变化上, 国内外趋势基本一致, 国内教师评价素养研究与聚焦性课堂评价实践研究略有延迟, 研究体量有待发展。基于上述发现, 本文进一步指出未来研究的发展方向。

**关键词:** 课堂评价; 形成性评价; 外语教学; 可视化分析

[中图分类号] H319 [文献标识码] A [文章编号] 2096-6105(2022)02-0034-08

## 1 引言

课堂评价 (classroom assessment) 指通过获取学生学习情况信息, 并对其进行解读, 从而提供理据为下一步教与学作出更好决策的过程。课堂评价伴随课堂内的教与学, 是检测与提高教学效果的最有效途径 (Black & Wiliam 2009), 因而成为国内外外语教学与语言测试研究的共同焦点。尽管课堂评价的研究成果丰富, 但研究话题多样, 且发展脉络复杂, 因此这一研究领域的发展趋势亟待梳理和总结。

关于“课堂评价”和“形成性评价” (formative assessment) 的具体区别, 目前学界尚未有明确定义。考虑到课堂评价以促学为目的, 倡导教、学、评三位一体, 这与以学习为目的、注重学习过程的形成性评价 (Leung & Mohan 2004) 相契合, 因

此本文的综述范围将二者均纳于其中。目前, 仅有 3 篇相关的综述类文献。其一, 吴秀兰 (2008) 探讨形成性评价在国内高校外语教学中的应用研究, 文献覆盖范围为 2000—2007 年。其二, 李清华 (2012) 简要评述国内外形成性评估研究, 主要介绍政府管理经验并指出我国形成性评估研究上的问题, 以 2008 年以前的文献为主。其三, 金艳、孙杭 (2020) 对 2007—2018 年国内外外语类权威期刊上的外语课堂评估文献进行描述性统计, 以定性研究方法探讨了研究话题的类别, 但未呈现研究热点的历时变化。鉴于此, 本文运用多元、分时、动态的第二代信息可视化技术 (Chen *et al.* 2012), 挖掘 2008—2020 年国内外外语教学课堂评价研究的核心研究主题, 分析研究热点变化, 进而指明我国外语教学课堂评价研究的发展方向。

<sup>1</sup> 本文由中央高校基本科研业务费专项资金 (编号: 2018JJ003) 资助, 同时得到北京外国语大学北京高校高精尖学科“外语教育学”建设项目 (项目编号: 2020SYLZDXM011) 的支持。

## 2 数据来源及处理

本文采用CiteSpace 5.7.R2软件,挖掘与分析近13年来国内外外语教学课堂评价研究成果。以Web of Science (WoS) 和中国知网 (CNKI) 为主要数据来源。分别以classroom assessment或formative assessment和“课堂评价”或“形成性评价”以及“课堂评估”或“形成性评估”为主题词进行检索,将期刊限定在与语言测试密切相关且发文量较高的期刊,其中包括*Language Testing*、*Language Assessment Quarterly*、*Assessing Writing*等20种外语类国际期刊以及CSSCI来源期刊、北京大学中文核心期刊及外语测试研究专业期刊《外语测试与教学》,时间跨度为2008—2020年。最终获得221篇国外文献和329篇国内文献作为统计分析数据。

## 3 国内外外语教学课堂评价核心研究主题分析

CiteSpace软件可根据文献共被引关系,自动抽取施引文献标题、摘要中的名词短语或文献关键词生成聚类(cluster),归纳研究焦点(Chen *et al.* 2012)。本文采用聚类分析方法生成了国内外外语教学课堂评价研究聚类知识图谱,得到国外研究45个聚类、国内研究52个聚类,其中与外语教学课堂评价密切相关的主要聚类如图1、图2所示。聚类模块值(Modularity Q)分别为0.64和0.75,聚类平均轮廓值(Weighted Mean Silhouette S)分别为0.90和0.95,均大于理想值0.4(马晓雷等2016),说明聚类内节点的联系紧密、主题关联性较强,结果具有参考价值。

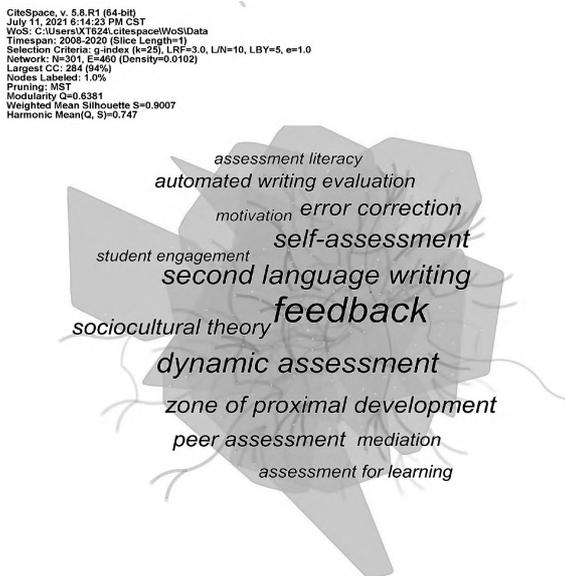


图1 国外外语教学课堂评价研究聚类知识图谱

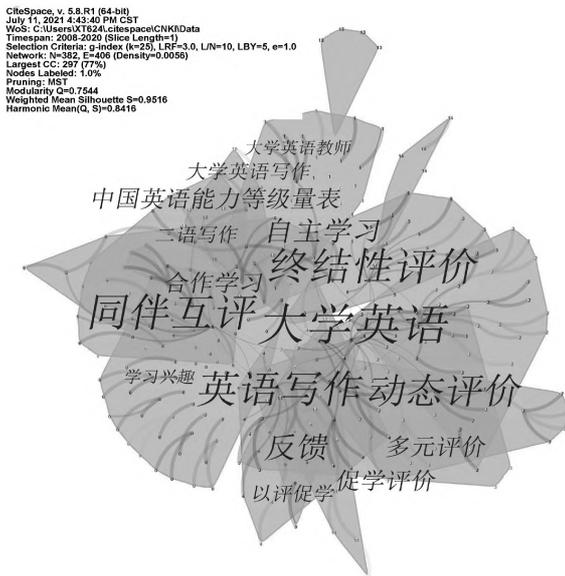


图2 国内外外语教学课堂评价研究聚类知识图谱

结合各聚类的标签描述,能更深入分析其涉及的研究话题,从而得出包含节点数最多(即囊括的文献最多)的前10个与外语教学课堂评价最密切相关的聚类及其LLR(log-likelihood ratio)标签<sup>1</sup>描述(见表1)。

1 CiteSpace中有三种算法可以进行聚类分析:TF-IDF加权算法提取出的特征词强调研究主流,对数似然率(LLR)算法与互信息(MI)算法提取出的特征词强调研究特点(徐金雷2018)。其中,LLR算法最常用,故本文列出LLR算法生成的各个聚类的标签,以对聚类进行更全面深入的解读。

表1 国内外外语教学课堂评价研究的10大主要聚类及其标签(2008—2020年)

国外研究主要聚类及其标签			国内研究主要聚类及其标签		
聚类号	节点数	标签(LLR)	聚类号	节点数	标签(LLR)
0	48	corrective feedback (纠正性反馈); error correction (纠错); teacher response (教师回应)	0	51	课程标准; 大学英语; 云端测评平台
1	36	self-assessment (自评); evaluation criteria (评价标准); peer-assessment (同伴互评)	1	43	同伴互评; 英语写作; 同伴反馈
2	27	essay writing (写作); factor analysis (因素分析); validity (效度)	2	39	同伴互评; 大学英语; 终结性评价
3	26	computer-generated feedback (计算机反馈); collaborative assessment (合作评价); language assessment literacy (语言评价素养)	3	36	终结性评价; 多元评价体系; 学习动机
4	25	zone of proximal development (最近发展区); peer feedback (同伴反馈); dynamic assessment (动态评价)	4	26	大学英语; 多元智力理论; 动态评估
5	24	automated writing evaluation (AWE, 作文自动评价); conversation analysis (话语分析); zone of proximal development (最近发展区)	5	19	以评促学; 共建的评分标准; 高中英语
6	24	AWE feedback (作文自动评价反馈); assessment criteria (评价标准); peer feedback (同伴反馈)	6	17	多元评价; 小学英语教学; 实验法
7	24	Coh-Metrix; L2 classroom experiment (二语课堂实验); second language writing (二语写作)	7	12	实证研究; 同伴互评; 大学英语
8	18	writing motivation (写作动机); assessment literacy (评价素养); student engagement (学生参与)	8	11	反馈信息; 网络环境; 学生自主学习
9	16	primary school (小学); sociocultural theory (社会文化理论); dynamic assessment (动态评价)	9	7	基础英语课程; 小学英语教学; 动态评估

根据各聚类的标签描述,按话题关键词所在聚类号先后顺序(聚类号越靠前,所包含的节点数越多,此话题的研究体量越大),本研究分别整合出国外与国内五大研究主题。国外五大研究主题为:1)反馈(聚类0、3、4、6);2)学生自评、同伴互评(聚类1、4、6);3)写作评价(聚类2、5、6、7、8);4)评价素养(聚类3、8);5)动态评价(聚类4、9)。国内五大研究主题为:1)大学英语(聚类0、2、4、7);2)同伴互评(聚类1、2、7);3)写作评价(聚类1、5),聚类5中“共建的评分标准”这一关键词所在的研究都是关于英语写作评价的,因此将其归为本主题;4)反馈(聚类1、8);5)终结性评价(聚类2、3)。国内外共同关注的研究主题有:“反馈”“学生自评与同伴互评”和“写作评价”。

在“反馈”这一主题下,国内外研究主要涉及两方面。第一,不同类型反馈的有效性及其效果,

如同伴反馈、纠正性反馈(corrective feedback)、作文自动评价反馈等。有研究提倡将同伴反馈与教师反馈、系统自动反馈相融合(如Dikli & Bleyer 2014; 何佳佳 2019),但由于研究持续时间短,各种反馈方式及其相融合的反馈方式的长期效果还有待进一步验证。第二,对评价反馈的作用机制进行探索,挖掘反馈效果的影响因素。目前国外此类研究较多,常采用混合研究,设计实验、量化分析实验组和对照组学生的前后水平差异,再通过分析学生写文本及教师、学生访谈等质性数据,解释实验结果(如Wu 2019)。

学生自评与同伴互评研究,根据侧重点不同,大致分为学生自评或同伴互评的促学效果探究以及影响评价效果的因素探索两类。国内外关于此话题的研究呈现两种趋势:其一,由早期对学生自评或同伴互评泛化的探讨转向学生自评或同伴

互评模式的建构、对某种具体评价工具有效性的验证(如May *et al.* 2020; 解冰等 2020);其二,从关注互评效果转向关注互评过程,聚焦性研究增多,从教师、学生、评价量表、评语等多角度对影响互评效果的因素进行探索。实证研究居多,常采用实验法、调查法、访谈法、文本分析或几种研究方法相结合,通过课堂观察、录像录音、教师评分等手段,探究学生自评或同伴互评的实施效果及作用机制(如Yu & Hu 2017; 周季鸣、束定芳 2019)。国内外关于自评与互评的研究成果丰富,通过对比同伴互评、学生自评、教师评价和在线评价系统等,并结合影响互评效果的因素,不断优化评价模式,提出了教师支持的同伴评价(teacher-supported peer assessment)(Zhao 2014)、“同伴互评+教师讲评”的循环反馈模式(白丽茹 2013)、“师生合作评价”(文秋芳 2016)、“线上同伴互评+线下教师面辅”(何佳佳 2019)等新型评价方式,以实现“以评促学”的目的。比较国内外研究特点,国内多聚焦于评价的效果(39篇中有17篇,占43.6%),国外研究除评价效果外也关注评价过程,通过课堂的实时评估行为及师生互动,探究互评的作用机制(如Yu & Hu 2017)。

“写作评价”这一主题具体涵盖的研究话题有某种评价工具在英语写作教学中的运用及其作用机制探索(如Zhang 2020; 隋晓冰、程璐璐 2019)、二语写作评价模型的构建(如于书林、Lee 2013)、评价量表的设计和效度验证(如Ghaffar *et al.* 2020)。“作文自动评分”和“师生共建的评分标准”近年引起了国内外学者的广泛关注(如Ghaffar *et al.* 2020; 隋晓冰、程璐璐 2019)。在此话题下,以实验法为主、调查和访谈为辅的混合研究最为常见,已成为惯用的研究范式。国外则更侧重行动研究、个案研究、课堂观察和访谈等质性研究方法(关于纯质性研究或以质性研究为主的混合研究,国外83篇中有35篇,占42.2%;国内67篇中有4篇,占6.0%)。此外,国内研究在构建二语写作评价模型时,有31.3%的研究以提出理论模型为目的(如刘敏、吴始年 2020),未

进行有效性验证;进行验证的研究中,36.4%的研究证据来源较单一,以采集学生的反馈意见为主,结论的说服力有待增强。

聚类结果所呈现的国内外不同研究主题值得关注,国外有“评价素养”与“动态评价”,国内有“大学英语”与“终结性评价”。评价素养研究涵盖语言测评素养的内涵与要素(如Lee & Butler 2020)、测评利益相关者的测评素养现状(如Yan & Fan 2020)。国内此类研究近几年开始受到关注(如江进林 2019),教师的测评素养直接影响教学中测评活动的开展与效果,其实践与研究未来有较大发展空间。有关动态评价的国外研究,已将“中介提供者”(mediator)由教师(如Zhao 2014)拓展至同伴(如Davin & Donato 2013),也挖掘了计算机辅助动态评价的潜力(如Poehner *et al.* 2015),以最大程度促进学生发展。国内动态评价研究更偏重理论基础探讨(21篇研究中有4篇实证研究)。

“大学英语”是国内特有主题。相关研究以大学外语课堂为情境,或以英语专业学生为研究对象,探讨课堂形成性评价的促学效果(如隋晓冰、程璐璐 2019)。该话题的凸显反映出国内研究在大学英语教学情境中开展得更为充分。以同伴互评研究为例,86.8%的研究对象集中在大学本科生这一群体。国外研究涵盖了处于不同教育阶段的研究对象。国内另一主要研究话题是“终结性评价”,不过该话题下的研究并非研究终结性评价本身,而是通过比较形成性评价与终结性评价,来展现形成性评价的作用或探究终结性评价与形成性评价相结合的多元评价模式(如刘森、牛子杰 2018)。

#### 4 2008—2020年外语教学课堂评价研究热点变化

关键词是对文章主题的高度概括和凝练,出现频次高的关键词通常也被视为该领域的研究热点(Chen *et al.* 2012)。本文绘制了国内外课堂评价研究关键词共现时区图,以呈现近13年来国内外研究热点变化(见图3、图4)。

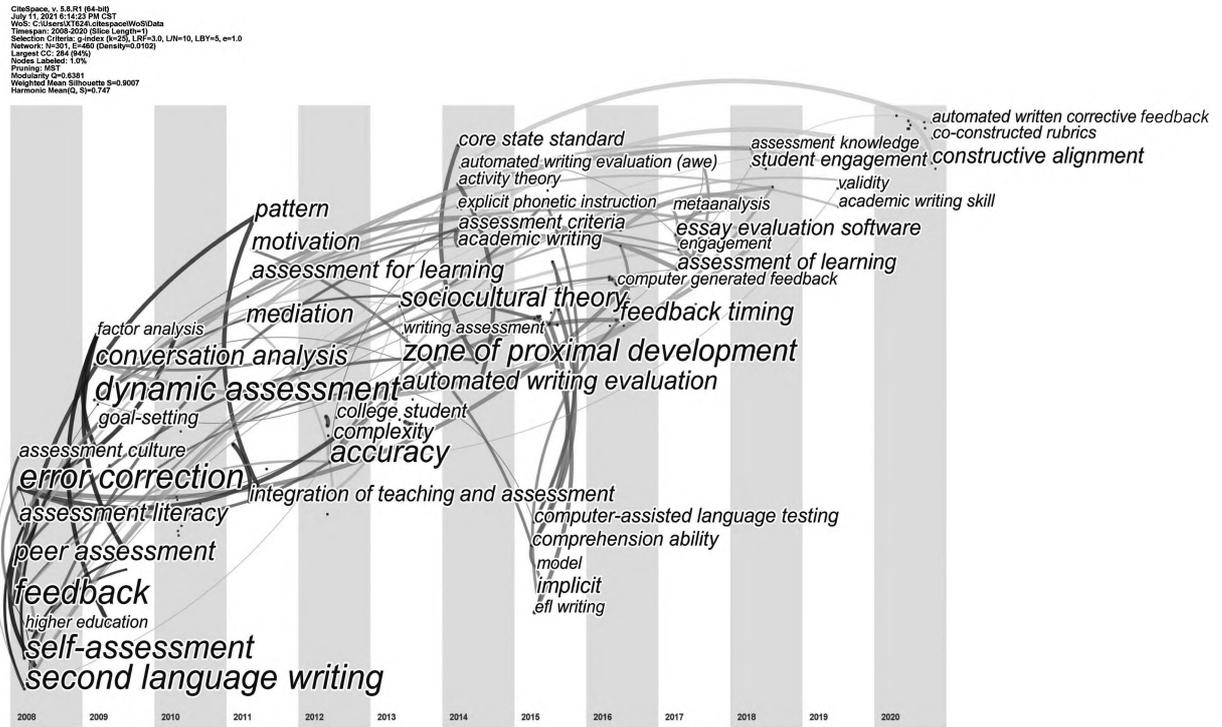


图3 国外外语教学课堂评价研究的关键词时区图

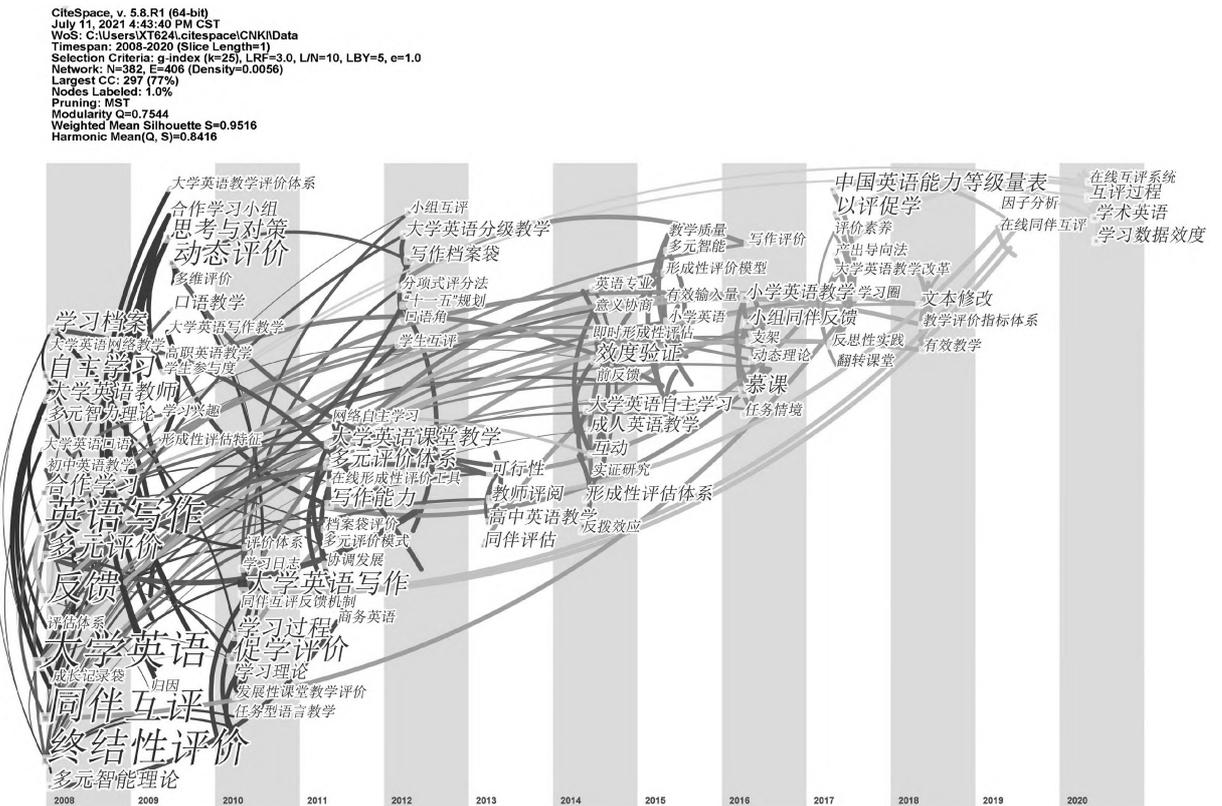


图4 国内外语教学课堂评价研究的关键词时区图

剔除与检索主题词相同或相近的关键词后，关键词列举如下(见表2)，以便进行横纵向对比。本研究将国内外研究中每年出现频次最高的3个

表2 近13年国内外外语教学课堂评价研究热点 (2008—2020年)

年份	国外研究热点	国内研究热点
2008	feedback (反馈); second language writing (二语写作); self-assessment (自评)	大学英语; 同伴互评; 终结性评价
2009	dynamic assessment (动态评价); conversation analysis (话语分析); factor analysis (因素分析)	动态评价; 思考与对策; 合作学习小组
2010	error correction (纠错); recast (重铸); language assessment literacy (语言评价素养)	大学英语写作; 促学评价; 学习过程
2011	assessment for learning (促进学习的评价); mediation (中介作用); motivation (动机)	大学英语课堂教学; 多元评价体系; 写作能力
2012	accuracy (准确性); complexity (复杂性); college student (大学生)	大学英语分组教学; 写作档案袋; 学生互评
2013	zone of proximal development (最近发展区); sociocultural theory (社会文化理论); automated writing scoring (作文自动评分)	同伴评估; 高中英语教学; 教师评阅
2014	academic writing (学术写作); assessment criterion (评价标准); automated writing evaluation (自动写作评价)	效度验证; 大学英语自主学习; 形成性评估体系
2015	computer-assisted language testing (计算机辅助语言测试); comprehension ability (理解能力); EFL writing (二语写作)	形成性评价模型; 多元智能; 教学质量
2016	feedback timing (反馈时机); computer-generated feedback (计算机生成的反馈); assessing speaking (口语评价)	小学英语教学; 小组同伴反馈; 慕课
2017	assessment of learning (对学习的评价); essay evaluation software (写作评价软件); engagement (参与)	中国英语能力等级量表; 以评促学; 评价素养
2018	student engagement (学生参与); assessment knowledge (评价知识); AWE feedback (作文自动评价反馈)	文本修改; 教学评价指标体系; 有效教学
2019	validity (效度); academic writing skill (学术写作技巧); implicit learning (隐性学习)	因子分析; 在线同伴互评; 评价量表
2020	constructive alignment (一致性建构); automated written corrective feedback (写作自动评价纠正性反馈); co-constructed rubrics (共建的评分标准)	在线互评系统; 互评过程; 学术英语

注: 按照关键词出现频次由高至低排列, 表中每一行的3个关键词从上至下依次为该年度的第一、第二和第三名。

从历时角度看, 2008—2020年, 国内外研究热点基本保持同步, 某些方面略有差异。

共性特点为: 第一, 以促学评价为导向, 对学习的评价(assessment of learning)、促进学习的评价(assessment for learning)、作为学习的评价(assessment as learning)、学习导向的评价(learning-oriented assessment)等课堂评估新范式陆续出现, 学生作为评价主体备受关注; 第二, 呈现多学科(如二语习得、教育技术学)交叉的态势, 评价研究运用社会文化理论、动态系统理论、新读写理论、教育技术理论等多种相邻领域理论; 第三, 随着“互联网+教育”的不断推进, 线上反馈以及基于平台的网络互评逐渐成为研究热点, 学者们开始运用云端测评中的数据挖掘与自动反馈功能来诊断学习和教学成效(如Zhang 2020; 刘敏、吴始年 2020)。

国内外研究的差异主要体现为: 其一, 评价

素养在2008年已成为国外研究热点, 至2020年间, 仍多次成为热点, 但该话题自2017年起在国内引起广泛关注, 这可能是由于此前国内课堂评价研究与教学实践之间的对接尚未成熟, 学界研究重心仍在评价的促学效果, 近年才开始重视教师评价素养对于评价与教学的重要作用; 其二, 国内研究的热点与外语教育政策制定、教学改革发展基本保持同步, 关键词涉及领域比较宏观, 凸显研究背景, 近年一些研究更能将背景具化为研究问题, 如《中国英语能力等级量表》的研发为国内外外语教学课堂评价研究提供了参照, 催生了量表的应用性研究(如潘鸣威、吴雪峰 2019); 国外研究更聚焦评价过程本身, 研究问题比较微观, 针对某一具体环节(如“反馈”)进行深入探究, 成果颇丰, 但孤立式的聚焦性研究, 若脱离教学评一体化的现实, 其成果对教学质量的提高作用有限。

## 5 研究结论与启示

本文运用CiteSpace对近13年来国内外外语教学课堂评价研究进行了可视化分析,对国内外的核心研究主题和研究热点变化进行了分析与比较,从而为国内相关研究的未来发展提供启示。总的来说,2008—2020年我国外语教学课堂评价研究已取得可观成果,既体现了对研究热点的敏感性,又与教学改革方向联系紧密,为未来研究提供了广阔的发展空间。

在研究内容方面,未来可重点关注以下几点。第一,在教学评一体化认识的前提下,基于实际教学情况,研究不同评价手段与评价阶段的原理与效果,逐步真正解决课堂评价的促学问题。第二,在受到国外研究热点影响的同时,结合国内外外语教学的重点问题、国内外语测评新动向以及我国学习者的主要特点,开展课堂评价的具体研究。待探讨的问题包括:通过评价促进学生思辨能力、跨文化能力和语言能力的协同发展(杨莉芳2018);将《中国英语能力等级量表》运用于设计和开展课堂评价活动;面对新出现的教育评价形式,为更好实现评价的促学效果,教师和学生应具备的素养;教材编写设计与课堂教学评价活动的关系等。第三,避免对同伴互评、学生自评等评价形式进行过于泛化的探讨,更多地开展基于具体学习目标的实践研究,如课程教学中同伴互评的设计与实现、促学作用机制等,并对其效度进行探究(顾永琦、李加义2020)。第四,在已有课堂评价模型或框架构建的基础上,探索其教学实践应用,以及基于具体案例对其进行验证与完善。第五,扩大研究对象的范围,除了聚焦大学英语课堂,也应关注形成性评价在基础英语教学阶段的应用。一线英语教师可以开展课堂评价的行动研究,通过学生数据的横向、纵向对比以及课堂观察等手段,对形成性评价在中小学英语教学中的实践进行深入探究。

在研究方法方面,由于国内外外语教学课堂评价研究起步较晚,较多采用非实证研究来讨论促学评价的可行性和重要性。近年实证研究有所增加,但多为量化研究和以量化研究为主的混合研

究,如以实验法为主,调查和访谈为辅,通过实验班、对照班成绩对比来考察学生学习成果的差异,描述发展趋势。若要深挖背后的机制和原理,我们须开展更多质性研究、以质性研究为主的混合研究及教师行动研究。语言测评研究在传统上基于大规模、高利害的标准化考试,采用项目反应理论、拉什模型等量化研究路径,描述群像,得出测量工具的效度。不同于标准化考试的实践与研究,课堂评价采用多种评价与教学手段,聚焦促学效应,关注个体学习者的体验与发展变化。因此,课堂评价研究也应广泛学习及吸纳叙事研究、案例研究、文化人类学研究、行动研究等质性研究方法,从而深描现象、挖掘多维及动态的因素与联系。

### 参考文献

- Black, P. & D. Wiliam. 2009. Developing the theory of formative assessment [J]. *Educational Assessment, Evaluation and Accountability* 21: 5-31.
- Chen, H., G. Zhao & N. Xu. 2012. The analysis of research hotspots and fronts of knowledge visualization based on CiteSpace II [A]. In S. K. S. Cheung, J. Fong, L.-F. Kwok, K. Li & R. Kwan (eds.). *Hybrid Learning* [C]. Heidelberg: Springer. 57-68.
- Davin, K. J. & R. Donato. 2013. Student collaboration and teacher-directed classroom dynamic assessment: A complementary pairing [J]. *Foreign Language Annals* 46: 5-22.
- Dikli, S. & S. Bleyele. 2014. Automated Essay Scoring feedback for second language writers: How does it compare to instructor feedback? [J]. *Assessing Writing* 22: 1-17.
- Ghaffar, M. A., M. Khairallah & S. Salloum. 2020. Co-constructed rubrics and assessment for learning: The impact on middle school students' attitudes and writing skills [J]. *Assessing Writing* 45: 100468. doi: 10.1016/j.asw.2020.100468.
- Lee, J. & Y. G. Butler. 2020. Reconceptualizing language assessment literacy: Where are language learners? [J]. *TESOL Quarterly* 54: 1098-1111.
- Leung, C. & B. Mohan. 2004. Teacher formative assessment and talk in classroom contexts:

- Assessment as discourse and assessment of discourse [J]. *Language Testing* 21: 335-359.
- May, L., F. Nakatsuhara, D. Lam & E. Galaczi. 2020. Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice [J]. *Language Testing* 37: 165-188.
- Poehner, M. E., J. Zhang & X. Lu. 2015. Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation [J]. *Language Testing* 32: 337-357.
- Wu, Z. 2019. Lower English proficiency means poorer feedback performance? A mixed-methods study [J]. *Assessing Writing* 41: 14-24.
- Yan, X. & J. Fan. 2020. "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups [J]. *Language Testing* 38: 219-246.
- Yu, S. & G. Hu. 2017. Understanding university students' peer feedback practices in EFL writing: Insights from a case study [J]. *Assessing Writing* 33: 25-35.
- Zhang, Z. 2020. Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions [J]. *Assessing Writing* 43: 100439. doi: 10.1016/j.asw.2019.100439.
- Zhao, H. 2014. Investigating teacher-supported peer assessment for EFL writing [J]. *ELT Journal* 68: 155-168.
- 白丽茹, 2013, 基础英语写作同伴互评反馈模式的可行性及有效性检验[J], 《解放军外国语学院学报》(1): 51-56。
- 顾永琦、李加义, 2020, 形成性评估的效度验证方法[J], 《外语教育研究前沿》(4): 57-63。
- 何佳佳, 2019, 基于Peerceptiv在线同伴互评系统的学术英语写作个性化辅导模式研究[J], 《外语电化教学》(2): 25-33。
- 江进林, 2019, 高校英语教师测评素养现状及其影响因素研究[J], 《外语界》(6): 18-26。
- 金艳、孙杭, 2020, 外语课堂评估研究(2007—2018): 回顾与展望[J], 《东北师大学报(哲学社会科学版)》(5): 166-173。
- 李清华, 2012, 形成性评估的现状与未来[J], 《外语测试与教学》(3): 1-7, 26。
- 刘敏、吴始年, 2020, 英语教学形成性评价云端测评模式构建[J], 《外语教学》(5): 71-75。
- 刘森、牛子杰, 2018, 优化英语语音教学多元评价模式的实证研究[J], 《外语教学理论与实践》(4): 62-68。
- 马晓雷、陈颖芳、侯俊霞、张宗波, 2016, 基于CiteSpace的二语习得研究可视化分析(1970~2014)[J], 《解放军外国语学院学报》(1): 121-129。
- 潘鸣威、吴雪峰, 2019, 中国英语能力等级量表在中小学英语形成性评价中的应用——以写作能力为例[J], 《外语界》(1): 89-96。
- 隋晓冰、程璐璐, 2019, 作文自动批改系统辅助大学英语写作慕课的交互式教学模式研究[J], 《现代教育技术》(2): 66-71。
- 文秋芳, 2016, “师生合作评价”: “产出导向法”创设的新评价形式[J], 《外语界》(5): 37-43。
- 吴秀兰, 2008, 形成性评价在国内高校外语教学中的应用研究综述[J], 《外语界》(3): 91-96。
- 解冰、高瑛、贺文婧、梁忠庶, 2020, 英语写作同伴互评感知量表的编制与探索性应用[J], 《外语教学》(3): 67-72。
- 徐金雷, 2018, 国际学科教学知识的文献计量研究[J], 《课程·教材·教法》(3): 132-138。
- 杨莉芳, 2018, 融合思辨能力与跨文化交流能力的语言测试任务设计——以“国际人才英语考试”为例[J], 《外语界》(2): 49-56。
- 于书林、I. Lee, 2013, 基于社会文化活动理论的二语写作同伴反馈系统模型构建[J], 《山东外语教学》(5): 24-29。
- 周季鸣、束定芳, 2019, 同伴互评中的教师实践与学生认识互动研究[J], 《外语界》(5): 64-71。

## 作者简介

杨莉芳, 北京外国语大学英语学院副教授。主要研究领域: 语言测评、外语教学、二语习得。电子邮箱: yanglifang@bfsu.edu.cn

王婷婷, 苏州工业园区星汇学校教师。主要研究领域: 语言测评、外语教学。电子邮箱: sandybubala@163.com

(审稿编辑: 孙曙光)

## A visual analysis of research on classroom assessment in foreign language teaching (2008—2020)

YANG Lifang & WANG Tingting..... 34

**Abstract:** By employing the data analysis tool of CiteSpace 5.7.R2 and the method of bibliometrics, this paper conducts a visual analysis of the literature on classroom assessment in foreign language teaching published in key international and domestic journals from 2008 to 2020 in WoS and CNKI database. Through investigating the core topics and the developmental trends of classroom assessment research, it is revealed that common topics shared by the international and domestic studies include feedback, self/peer assessment and writing assessment. While the domestic studies focus more on the assessment effect, the assessment process has been emphasized in overseas research. Concerning the research trends, domestic studies generally follow the pace of their international counterparts, slightly lagging behind in assessment literacy and focused classroom assessment research. Based on interpretations of such findings, this paper proposes directions for future research.

**Keywords:** classroom assessment; formative assessment; foreign language teaching; visual analysis

## Maritime ESP course unit framework and practice based on POA

YANG Cen & GUO Suihong..... 42

**Abstract:** Combining POA and the current maritime ESP teaching, the paper puts forward maritime ESP course unit framework: 1) “cross-border cooperation based” ability and task chain; 2) three-layered gradual enabling activities of “trial, multi-modal compensation, reconstruction”; 3) “cross-border cooperation based” assessing criterion. The paper also illustrates the practice of this framework via the course “Maritime Technology English”, revealing that the framework enhances students’ motivation, language application and producing abilities.

**Keywords:** POA; maritime ESP; Maritime Technology English

## A critical review of L2 learners’ emotion

QIN Lili, YAO Lan & NIU Baogui..... 51

**Abstract:** Emotion research in SLA has received increasing attention in recent years, and learners’ emotion is crucial for understanding the second language learning process. This study reviews L2 learners’ emotion research from the traditional psychological, positive psychological and sociocultural perspectives, and compares their empirical research paradigms. The study finds that: 1) the theoretical development of emotion research has been diversifying from the traditional psychological to the positive psychological perspectives, and there is also a sociocultural turn; 2) existing emotion research topics are mainly focused on foreign language enjoyment and anxiety, and there is more room for development; 3) the empirical research paradigms on emotion are developing from monolithic to diversified, and there is an urgent need to find a more scientific research channel. This study aims to provide new ideas for future related research.