

国外近十年深度学习实证研究综述

——主题、情境、方法及结果

沈霞娟, 张宝辉, 曾 宁

(陕西师范大学 教育学院, 陕西 西安 710062)

[摘 要] 为厘清国外教育领域深度学习的研究进展,选取 Web of Science、Springer、ScienceDirect 等数据库的 85 篇文献(2008—2018 年),依据实证研究的核心要素建立编码体系进行内容分析。结果表明:(1)研究主题集中在深度学习策略、方式和评价三个方面;(2)比格斯、马顿 & 塞尔乔、恩特威斯尔 & 拉姆斯顿提出的深度学习概念框架得到较多认可;(3)深度学习在各类学科情境中广泛应用,但研究对象多为大学生;(4)实证研究数据呈现以量化分析为主、质性分析为辅的特征,混合分析日益得到重视;(5)自我报告量表、编码标准、条件化测量、眼动追踪是深度学习的主要测量方法;(6)深度学习的有效性突出表现在学业成绩、知识理解、学习体验、思维与能力发展等方面。

[关键词] 深度学习;实证研究;文献综述;研究主题;研究方法;研究情境

[中图分类号] G434

[文献标志码] A

[作者简介] 沈霞娟(1982—),女,河南商丘人。讲师,博士研究生,主要从事学习科学、教育信息资源等方面的研究。
E-mail:shenxiajuan2009@163.com。张宝辉为通讯作者,E-mail:baohui.zhang@snnu.edu.cn。

一、研究背景

面对知识更迭速度的加快,以及经济全球化、信息化带来的发展挑战,人们开始对学生的学习方式和质量进行深入反思。在基础教育领域,世界各国纷纷通过颁布核心素养的方式提升学习目标。观察这些核心素养的内容与结构,不难发现,为了成为合格的 21 世纪公民,学习者不仅要牢固掌握学科知识,还要学会学习,具有良好的问题解决、高阶思维、自主学习和实践能力。因此,机械的、被动的、简单记忆式的浅层学习已经无法满足社会发展的需要。在高等教育领域,创新人才培养成为最核心的价值追求,欧洲的《博洛尼亚宣言》指出,“成功的高等教育应该给予学生参与深度学习的机会”^[1];美国新媒体联盟发布的《2017 地平线报告(高等教育版)》认为,深度学习将成为未来 5 年以及更长时间内高等教育机构决策制定的重

要影响因素^[2]。

深度学习是技术领域和教育领域共同关注的热点问题,但具体内涵差异较大。前者是指实现机器学习的高效技术,因人工神经网络的隐层数量多而得名^[3],主要应用于图像及语音识别、专项竞赛等方面;后者是指以提升批判性思维、问题解决等高阶能力为目标的有效学习方式,旨在提高学生的学习质量。自 1976 年马顿(Marton)和塞尔乔(Saljo)根据信息加工方式的不同,明确提出教育领域中深度学习和浅层学习的概念以来,国外研究者开展了大量卓有成效的深度学习理论和实践探索。其中,实证研究因提倡“用数据说话”而备受研究者、管理者和决策者的重视^[4]。特别是近年来,随着深度学习与学科课程的有机融合,以及深度学习测量工具的不断完善,实证研究逐渐走向繁荣,呈现出研究主题丰富、方法多样、数据翔实、过程严谨、情境广泛的典型特征,积累了大量具有较

基金项目:2017 年度中央高校基本科研业务费专项资金资助项目“面向深度学习的在线课程设计与应用研究”(项目编号:2017TS068)

高借鉴价值的研究文献。

我国现有的深度学习综述文献,多采用知识图谱分析方法,在表层呈现某类研究的样本分布(国家、期刊、著者、机构等)、基本趋势和研究热点,较少针对研究的详细内容与过程进行系统描述、归类和分析,无法帮助读者在操作层面上理解深度学习的教学设计、成功途径和评价方法。因此,为了更好地推进深度学习在我国教育教学中的实践应用,本文对国外近十年深度学习的实证研究文献进行梳理和筛选,并依据实证研究的核心要素建立内容分析框架,进行多维度的分析解读。

二、研究设计

(一)研究方法

采用内容分析法开展研究,即在一定的分析框架指导下,将文献内容组织和集成到一定的主题、类别和代码中形成客观、系统的量化,并对量化结果进行分析、描述和推论^[5]。该方法主要包括四个步骤:(1)依据研究问题选择研究样本,确定分析单位;(2)设计内容分析维度和类目编码体系;(3)依据类目编码体系对研究内容进行评判和归类统计;(4)利用统计分析工具进行量化分析和描述,并通过推理得出结论。

(二)研究问题

本文希望通过内容分析厘清国外近十年深度学习实证研究的开展过程和重要发现,具体研究问题包括:深度学习实证研究的核心主题有哪些?研究情境是什么?遵循哪些概念框架?研究方法有何特征?如何对深度学习进行测量?取得了什么样的研究结果?

(三)样本选取

为有效获取国外近十年深度学习的高质量实证研究文献,在 Web of Science、EBSCO ERIC、Elsevier ScienceDirect、Springer、Wiley 五个主要的教育文献数据库中,以“deep learning”“deep approach”“deep processing”“deep strategy”“deep learner”为关键词进行标题精确检索,时间设定为 2008 年 1 月 1 日—2018 年 1 月 1 日,共获得 345 篇外文文献。导入 EndNote 中进行文献管理,通过自动和手动搜索,共找到全文文献 267 篇。

根据研究目的,通过 4 个筛选条件对初步检索到的全文文献进行精炼:(1)研究领域必须是教育领域中的深度学习,而不是技术领域中的深度学习;(2)研究类型必须是实证研究,即每篇文章需要报告定量或质性数据的搜集、处理方法及分析结果;(3)研究过程必须包含明确的研究问题和清晰的研究结论;(4)研

究语言和质量必须是同行评议的英文论文。

经过阅读摘要和概览全文,确定符合精炼条件的文献共 85 篇,其中 Web of Science 40 篇, Springer 15 篇, EBSCO ERIC 子库 16 篇, Elsevier ScienceDirect 9 篇, Wiley 5 篇,进入下一步内容分析流程。

(四)内容分析编码体系

为深入探析实证研究的开展过程,我们需要在明晰实证研究基本流程与核心要素的基础上建立内容分析框架。丁斯莫尔和亚历山大(Dinsmore & Alexander)提出,实证研究应包含概念界定、测量方式、学科情境、研究结果四个核心要素^[6]。刘选认为,从操作框架来看,一项实证研究是由“真实问题—方法运用—数据分析—结论诠释”构成的统一体^[4]。因此,本文建立了包括研究主题、概念框架、研究情境、研究方法、测量方式、研究结果六个维度的内容分析编码体系(见表 1)。

表 1 深度学习实证研究内容分析编码体系

一级类目	二级类目	内容编码
研究主题	主题类别	深度学习方式、深度学习策略、深度学习评价、深度学习资源、深度学习动机、深度学习过程、其他
概念框架	概念框架类型	明确型的、代理型的、未说明的
	概念框架提出者	根据实际情况填写
研究情境	学科领域	根据实际情况填写
	参与者类别	大学生、中学生、小学生、成人、未说明
	参与者数量	根据实际情况填写
	研究持续时间	以月为单位计算
研究方法	数据搜集方法	问卷调查、访谈、观察、测验、文档(如反思日记、交互记录、作业等)、其他
	数据分析方法	定量分析、定性分析、混合分析
测量方式	测量类型	自我报告量表、生理反应、依据特定条件、编码标准、无测量
	测量工具	R-SPQ-2F、ASSIST、ALSI、NSSE、MSLQ、自主开发量表(SDQ)、其他
研究结果	结果类型	记忆或回忆、学术/课程表现、文本理解、能力发展、情感体验、其他
	结果倾向	积极的、消极的、中立的

(五)内容分析信度

本文以篇为单位开展内容分析,每篇文献的具体内容从上述 6 个维度(共 13 个二级类目)进行统计分析,85 篇文章共计需要完成 1105 次评判。为了保证内容分析的信度,三位作者共同制定内容分析框架,

并多次交流对编码的内涵理解。由第一作者(主评)和第三作者(辅评)分别完成所有文献的全文阅读和内容编码,经过两轮独立评判后,编码结果相互同意的类目数为927个。根据内容分析信度公式 $R=(n \times K)/(1+(n-1) \times K)$, 平均相互同意度 $K=2M/(N1+N2)$, M 是彼此同意的栏目数, $N1$ 是评判员 A 分析的类目总数, $N2$ 是评判员 B 分析的类目总数^[5], 计算出平均相互同意度 K 是 0.84, 信度 R 为 0.91。针对评判结果不一致的类目, 进一步通过协商和咨询第二作者(通信作者)的方式确定最终编码。

三、数据分析结果

(一) 深度学习研究主题

国外近十年深度学习实证研究的主题非常丰富, 涉及深度学习方式、学习策略、学习过程、学习评价、学习资源、学习动机等六个方面。其中, 最受关注的三大研究主题是深度学习策略(37.65%)、深度学习方式(32.49%)和深度学习评价(15.29%), 说明国外研究者通过多视角对深度学习方式进行解读, 不断探究如何促进深度学习的具体策略, 并试图对深度学习进行科学评价。但是, 国外研究者对深度学习的发生过程(7.06%)、资源建设(4.71%)和动机激发(2.35%)关注不足, 仅有极少量文献开展了相关研究, 这可能会制约深度学习在教育教学中的有效应用。

下面进一步对三大核心研究主题进行详细阐述:

1. 深度学习方式

国外近十年的实证研究结果显示, 深度学习方式在不同概念框架下具有不同的内涵, 其影响因素远比我们想象的复杂, 除人口统计学因素外, 自我效能感、学习动机、学习投入、学习环境体验等都可能对深度学习方式产生直接或间接影响, 但哪些是主要影响因素尚有待后续研究进一步识别。珀斯塔瑞弗等(Postareff)通过访谈法, 探究了不同学习环境中导致深度学习方式发生变化的因素, 结果表明, 相比较学习环境因素而言, 学习时间和精力投入、学习兴趣、课程预期目标、自我调节能力等个体因素能够对深度学习方式的变化作出更好的解释^[6]。潘(Phan)为了揭示深度学习方式与个体因素间的复杂关系, 设计了两个调查研究, 其中研究 1 为横向调查, 路径分析揭示学业成就受深度学习方式、批判性思维的直接影响, 目标取向和自我效能感通过深度学习策略和批判性思维对学业成就产生间接影响作用; 研究 2 是纵向调查, 时隔 1 年对同一群体 2 次采样, 路径分析表明, 掌握目标取向与自我效能感、掌握目标与批

判性思维、深度学习方式与批判性思维呈现复杂的双向互促关系^[7]。

2. 深度学习策略

国外近十年深度学习的促进策略, 主要体现在学习过程可视化、学习任务结构化和学习情境真实化等方面, 并呈现出典型的技术支持特征。彭(Peng)等通过构建虚拟学习环境将复杂的编程过程可视化, 结果表明, 不仅提高了学生的学业表现、促进了学科知识的发展, 而且有效激发了学生的内部学习动机^[9]; 佩格勒姆等(Pegrum)通过恰切的结构化任务形式在课程中引入创造性播客, 对比实验结果表明, 有效促进了知识的深度理解和有效保持^[10]; 特纳和巴斯克维尔(Turner & Baskerville)则设计了面向评价的学习任务、团队协作、友好的师生关系、基于真实案例等综合干预措施, 结果表明, 大部分学习者加深了对课程内容的理解并开展了有效的合作学习^[11]。

3. 深度学习评价

评价对学习具有重要的导向与调控作用, 如果教师采纳的评价策略鼓励学生变成批判性和创新性思考者, 就会引发深度学习, 反之, 如果采纳的评价策略强调背诵和记忆则可能导致浅层学习。国外近十年的实证研究表明, 深度学习评价方式不断多元化, 非常重视过程性评价和表现性评价对深度学习的促进作用, 评价内容主要包括核心知识掌握、内容理解、高阶能力发展和学习体验。格林等(Green)在经济学课程评价中融入图形化综合口试测评, 注重学科知识之间的相关联系和实际应用, 通过运用恰当的反例和边缘案例, 有效促进了学生对市场机制的深度学习^[12]; 林奇(Lynch)在一门职前教师教育课程中引入自我评价和同伴评价及反馈系统, 结果证明, 该方法能够鼓励学生对自我和同伴的学习过程进行反思、评价和批判, 有效提高了学习者的学业成绩和批判性思维能力^[13]; 达默等(Dummer)针对地理课中的实地考察教学, 提出了一种创新、灵活的深度学习评价方法——反思性田野工作日记, 结果表明, 该方法能够使教师了解学生的学习过程, 增强学生的批判性自我反思和书面沟通技能^[14]。

(二) 深度学习概念框架

概念框架是确定深度学习实证研究整体思路的重要依据, 不仅决定实证研究采用的测量方法, 还影响对研究结果的具体解释。从概念定义的清晰程度看, 37.65%的研究明确给出了深度学习的概念定义(即明确型), 36.47%的研究引用他人观点作为深度学习内涵的界定依据(即代理型), 另有 25.88%的研究

没有提及深度学习的定义(即缺失型)。

深度学习的内涵比较丰富,涉及内部动机、信息加工、整合推理、问题解决等方面,尚未形成统一的概念框架。因此,本文进一步对深度学习的概念框架进行词频统计(见表2)。结果发现,被5个以上实证研究引用的概念框架主要是由马顿 & 塞而乔(Marton & Saljo)、恩特威斯尔 & 拉姆斯顿(Entwistle & Ramsden)和比格斯(Biggs)提出的。

马顿和塞尔乔(1976)最早提出了深度学习的基本内涵,其典型特征是追求对内容的理解和能够将新知识与先前知识、经验联系在一起;而浅层学习则是单纯的重复信息而没有任何分析^[15]。

恩特威斯尔和拉姆斯顿(1983)进一步将深度学习的内涵从单纯的信息加工扩展为复杂的认知和元认知过程,指出深度学习不仅涉及观点之间的有机联系、模式/原则探寻、证据使用和逻辑论证,还包括学习者监测自己理解的发展;而浅层学习的目的仅仅是应对任务,将课程内容视为不相关的信息碎片,这将导致更加有限的学习过程,特别是死记硬背^[16]。

比格斯(1987)进一步强调了深度学习是学习者个体与教学环境的交互方式,建立了教学过程3P(Presage-Process-Product)模型^[17],并通过实证方法验证了深度学习方式的影响因素不仅包括个体因素,还包括大量的教学环境因素(如教学目标、教学策略、教学评价等)。在该模型中,深度学习表示学生对学习内容兴趣浓厚、主动寻求对内容的深刻理解、有效将新知识与先前经验或知识联系起来;而浅层学习则是指学生对学习内容不感兴趣、简单记忆知识、不能有效地将新知识纳入已有认知结构。

表2 深度学习概念的提出者及引用频率(N=63)

概念框架提出者	引用频率	百分比
比格斯	22	34.92%
恩特威斯尔 & 拉姆斯顿	16	25.40%
马顿 & 塞尔乔	8	12.70%
纳尔逊·莱德	4	6.35%
宾特里奇	2	3.17%
赖 & 布朗	1	1.59%
作者自定义	7	11.11%
其他	3	4.76%
合计	63	100.00%

(三)深度学习研究情境

为了进一步探究深度学习实证研究是在什么情境下开展的,本文选取学科背景、参与人数、研究持续时间三个决定实证研究结论可信度和推广价值的关

键指标进行分析。

1. 学科背景

深度学习不能仅在孤立的学习任务(如文本阅读)中开展实证研究,而是要在各类学科情境中真实应用,方能凸显深度学习对高素质人才培养的重要价值。近十年深度学习已经在社会科学和自然科学中广泛应用(如图1所示),其中医学、教育学、心理学三个学科对深度学习最为关注,研究案例比较丰富。

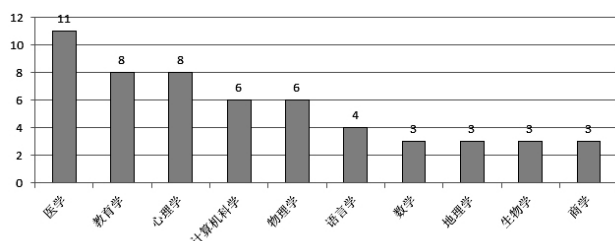


图1 深度学习实证研究的具体学科领域(N=85,前10排序)

2. 参与人员

从参与人员的类别来看,国外近十年深度学习实证研究中,80.00%以大学生为研究对象,8.24%以中学生为研究对象,3.52%以小学生为研究对象,8.24%以成年人为研究对象。可知,大学生是国外深度学习实证研究最为关注的一个群体,但是对中小学生和成人的深度学习研究不足,今后应加强不同学段中的应用研究和对比研究。

从参与人员的数量来看,1~30人的小样本研究17个,31~100人的研究27个,100~500人的研究25个,500~1000人的研究10个,多于1000人的大样本研究6个。其中,参与人数最多的研究是罗科尼(Rocconi)等通过问卷调查探究了116000名大学高年级学生的研究生计划与深度学习方式之间的关系^[18];而参与人数最少的则是戈莫尔(Gomoll)等通过观察和访谈4名学生的课堂表现,提出协作问题解决空间,探索了促进深度学习和问题解决的话语特征和动作特征^[19]。

3. 研究持续时间

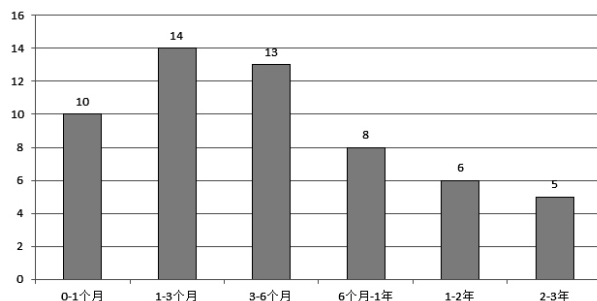


图2 深度学习实证研究周期统计图(N=56)

针对56个明确提出持续时间的实证研究进行统计分析(如图2所示),可知超过60%的深度学习实证研究是半年之内的一次性干预研究,长时间、多轮次的持续研究较少。因此,为了实现深度学习远期效果的跟踪研究,并通过反馈数据实现干预的优化和完善,建议研究者今后尝试开展多轮迭代的设计研究。

(四)深度学习实证研究方法

为了更好地透视国外近十年深度学习实证研究的开展过程,本文将研究方法细分为数据搜集方法和数据分析方法两个子类。

1. 数据搜集方法

一个研究可能存在多种数据搜集方法,因此,观察个案百分比更能看出某一方法在整体中的应用情况(见表3)。可知,问卷法的应用最为广泛(76.47%),该方法能够在短时间内、面向大量研究对象搜集信息,一般用于深度学习方式或能力的调查。排在第二位的是测试法(41.18%),主要用于深度学习干预策略实施前后学生在知识、思维、能力等方面变化情况的数据搜集,如提苏束(Tsaushu)通过多选题和高阶思维训练开放题测试,探究了同伴学习和个性化智能导师对深度学习的促进作用^[20];排在第三位的是访谈法(31.76%),由于在人员、时间方面访谈法的成本代价相对较大,一般用于小规模质性研究,实现深度学习过程和结果的深入了解或原因探寻,如阿德勒等(Adler)利用半结构访谈,探究了18名学生对“深度理解数学”和“深度数学学科知识”的理解^[21]。此外,反思日记、学生作业、在线交互记录等文档类资料(25.88%)和课堂观察结果(14.12%)也被不少研究纳入分析范围,以便从多维度开展深度学习过程或结果的测量。值得注意的是,为了消除每种单一方法的局限性,统计显示,52.94%的研究应用多种数据收集方法,从而为更好地研究推论奠定基础。

表3 实证文献数据搜集方法统计表(N=85)

数据搜集方法	频率	百分比	个案百分比
问卷	65	39.16%	76.47%
测试	35	21.08%	41.18%
访谈	27	16.27%	31.76%
文档	22	13.25%	25.88%
观察	12	7.23%	14.12%
其他	5	3.01%	5.88%
合计	166	100.0%	195.29%

2. 数据分析方法

从数据分析方法的统计结果看,量化分析占60.00%,质性分析占17.65%,另有22.35%的研究采

用了定量和定性相结合的混合分析方法。因此,鉴于多种数据搜集方法的逐步应用,混合分析方法将是未来深度学习实证研究的主流发展趋势。

(五)深度学习测量方式

“深度学习的典型特征是什么”“深度学习真的发生了吗”“深度学习的效果如何”是深度学习实证研究无法回避的三个关键问题。因此,需要在概念框架的指导下,根据具体的研究问题确定恰切的测量类型和工具。

1. 测量类型

根据丁斯莫尔和亚历山大提出的测量方式编码类目,深度学习主要采用自我报告量表、编码标准、依据特定条件、生理反应四类方式进行测量^[6]。观察个案百分比可知(见表4),自我报告量表(67.06%)的应用最为广泛,常见的有修订版两因素学习过程问卷(R-SPQ-2F)、学习方式调查问卷(ALSI)等;使用特定编码标准测量深度学习(29.41%)也占较高比例,如皮尤等(Pugh)自主开发了编码标准,从概念理解和转化视角探究高中生科学概念的深度学习^[22];依据特定条件测量深度学习(15.29%)具有较高的灵活性,如斯托得和哈特林(Stott & Hattingh)根据梅耶提出的三种工作记忆认知加工来测量学习者利用概念辅导软件进行密度概念学习过程中的深度处理和参与^[23];最后,少数研究采用生理反应(5.88%)测量深度学习,其中眼动作为一种可直接观测的物理特征,在眼动仪等测量工具的支持下正在成为深度学习一种有效的新型测量方式,如萨摩荣(Salmeron)等使用眼动追踪和出声思维,探究了学生利用维基百科回答问题时快速扫描、深度加工策略、阅读能力与文本理解之间的关系^[24]。

表4 深度学习的测量类型(N=85)

测量类型	频率	百分比	个案百分比
自我报告量表	57	54.81%	67.06%
编码标准	25	24.04%	29.41%
依据特定条件	13	12.50%	15.29%
生理反应	5	4.81%	5.88%
未说明	4	3.84%	4.71%
合计	104	100%	122.35%

2. 测量工具

针对采用自我报告量表的实证研究(共57篇),进一步探究了具体的测量工具。近十年国外研究者使用较多的五类深度学习量表(见表5)主要包括:(1)R-SPQ-2F,比格斯等(Biggs, et al.)开发的修订版两因素学习过程量表,由早期的SPQ量表修订而来,包括深

度动机、深度策略、浅层动机、浅层策略四个指标^[25];(2)NSSE,美国教育部发布的高校学生学习投入量表,深度学习是其中的一个测量维度,包括高阶学习、整合学习、反思学习三个指标^[26];(3)ALSI,恩特威斯尔和麦丘恩(Entwistle & McCune)开发的学习方式量表,由早期的ASL量表修订而来,包括深度方式、浅层方式和组织方式三个指标^[27];(4)ASSIST是泰特等(Tait, et al.)开发的学生学习技能量表,也是由早期的ASL量表修订而来,包括深度方式、浅层方式和策略方式三个指标,其中深度学习方式的典型表现是寻求意义、关联证据、应用证据和对观点感兴趣^[28];(5)MSLQ,宾特里奇(Pintrich)等开发的学习动机与策略量表,其中深度学习策略重点关注阐述、组织化、批判性思维,浅层学习策略则关注内容复述^[29]。此外,统计发现,不少研究者通过自主开发或改编量表开展深度学习的测量。该类量表能够与研究目的紧密结合,但往往没有经过严格的心理学检测和规模化试用,信效度可能无法有效保障,研究结果也不易与同类研究进行比较。

(六)深度学习研究结果

实证研究的结果一般建立在数据分析的基础之上并经过推理得出,是对研究假设和研究问题的明确回答。从研究结果的具体类型来看(见表6),学业成就(41.18%)、思维与能力发展(37.65%)和情感体验(17.65%)是国外实证研究评价深度学习成效的三个核心指标。此外,也有一些研究通过文本/内容理解(7.06%)和记忆/回忆成绩(5.88%)评价深度学习。

从结果倾向来看,82.35%的实证研究取得了较为积极的结果,突出表现在学业成绩的提升^[9]、更好的内容理解和保持^[10]、更加愉悦的学习体验^[11]、更好的高阶思维能力和专业技能发展^[30]等方面。但是,17.65%的

研究认为,深度学习没有带来具有显著差异的学习结果改变。因此,在具体的实证研究中应加强干预设计的科学性和先进性,并对学科教师进行必要的指导,以取得预期的研究效果。

表6 深度学习实证研究结果类型(N=85)

结果类型	频率	百分比	个案百分比
学术/课程表现	35	30.43%	41.18%
思维/能力发展	32	27.83%	37.65%
情感体验	16	13.91%	18.82%
关系探究	12	10.43%	17.65%
其他	9	7.83%	10.59%
文本理解	6	5.22%	7.06%
记忆/回忆	5	4.38%	5.88%
合计	115	100%	138.83%

四、结论与启示

(一)研究结论

本文以国外近十年深度学习领域的85篇英文文献为例,根据实证研究的核心要素建立分析框架,开展了研究主题、概念框架、研究情境、研究方法、测量方式以及研究结果六个方面的系统化内容分析,以揭示深度学习实证研究的内容与过程,得出以下结论:

(1)从研究主题看,主要集中在深度学习方式、策略和评价三个方面,深度学习的发生过程、资源建设和动机激发等方面的研究较为薄弱。(2)从概念框架看,深度学习实证研究十分重视核心概念的内涵界定,其中比格斯、马顿 & 塞尔乔、恩特威斯尔 & 拉姆斯顿提出的概念框架得到较多认可。与丁斯莫尔和亚历山大(2012)的深度学习综述文献^[6]相比,本文进一步指出了深度学习的概念类别与特征,可以为研究者提供更加确切的指导。(3)从研究情境看,深度学习已

表5 深度学习常用量表统计(N=57)

量表简称	量表英文全称	开发者及时间	主要指标	频率	百分比
R-SPQ-2F	R-Study Process Question-2F	Biggs, Kember & Leung (2001)	深度动机、深度策略、浅层动机、浅层策略	15	26.32%
NSSE	National Survey of Student Engagement	NSSE Institute(2000)	高阶学习、整合性学习、反思性学习	6	10.53%
ASSIST	Approaches and Study Skills Inventory for Students	Tait et al. (1998)	深度方式、浅层方式、策略方式	2	3.51%
ALSI	Approaches to Learning and Studying Inventory	Entwistle and McCune (2004)	深度方式、浅层方式、组织方式	2	3.51%
MSLQ	Motivated Strategies for Learning Questionnaire	Pintrich et al. (1993)	批判性思维、阐述、组织化	2	3.51%
SDQ	Self-Development Questionnaire	——	——	23	40.35%
Others	——	——	——	7	12.27%

经在社会科学和自然科学中广泛应用,其中医学、教育学、心理学三个学科中的研究案例最为丰富,但是从学段来看,研究对象多为大学生,对中小学生和成人关注不足。该结论与张思琦等(2017)^[31]的研究结论基本一致,并进一步分析了深度学习的具体学科背景和学段分布情况。(4)从研究方法看,问卷、访谈、测验是应用较多的数据搜集方法,数据分析则呈现以量化为主、质性为辅的特征,同时,混合分析日益得到研究者的重视。(5)从测量方式看,自我报告量表、编码标准、依据特定条件是最常用的深度学习测量类型,眼动作为一种可观察的生理特征正在成为一种有效的新型测量方式。同时,该研究进一步通过统计得出了五类常用的深度学习方式调查问卷,为后续实证研究的工具选择提供参考。(6)从研究结果看,深度学习的有效性和先进性得到了实证研究的充分支持,突出表现在学业成绩的提升、更好的知识理解与保持、更加愉悦的情感体验、更好的高阶思维和能力发展等方面。

(二)研究启示

我国教育领域的深度学习研究近几年得到了较多研究者的关注,从发文量看,目前已经进入呈指数增长的快速发展阶段^[32],主流研究范式正从理论探究转向实践应用。因此,亟须以国外已有研究为鉴,厘清深度学习实证研究的主题、情境、方法与结果,为后续研究者提供参考和依据。本文在分析深度学习实证研究的过程中,得到以下几点启示:

首先,重视深度学习的内涵界定。技术领域中的深度学习作为一种高效的神经网络算法,内涵比较明确,应用效果显著。然而,教育领域中的深度学习内涵比较复杂,而且由于人类学习的内隐性,目前深度学习的发生机制尚未完全揭示,学者们往往从信息加工方式、认知过程、学习结果等视角给出内涵各异的界定。因此,为了确保实证研究中概念、测量、结果解读之间保持良好的一致性,应该重视深度学习的领域特征和内涵界定,否则就容易出现泛化和异化的现象。

其次,关注深度学习与学科融合的具体路径。近十年国外实证研究表明,不同学科背景下深度学习的具体表征、促进策略和应用成效不尽相同。因此,为了探寻有效的深度学习实践路径,研究者应以批判性思维、问题解决能力、交流协作能力等通用标准为参照,

明晰不同学科中深度学习能力的具体表征,如在数学领域可结合数学运算能力、数学建模能力、空间想象能力等进行详细描述,而医学领域的深度学习能力,则要结合临床技能、评鉴技能、病理分析技能等进行详细描述。在此基础上,形成具有鲜明学科特色的深度学习策略或模式。

再次,加强深度学习的科学测量。客观、公正的测量数据能够为深度学习方式的确定或深度学习能力的变化提供有效证据。与国外相比,我国深度学习的实证研究,特别是调查与测量研究仍处于小范围探索阶段,深度学习的关键影响因素与作用机制有待进一步揭示,广泛认可的本土化测量工具较少。因此,我国研究者应加强深度学习测量工具和方法的专项研究^[33],尝试开展大规模横向对比研究和纵向跟踪研究,有效提高研究结果的推广价值。同时,鉴于自我报告量表受限于学习者的忠诚度和自我判断准确度,建议研究者综合使用量表、编码标准、生理反应(如眼动)等多种测量工具,以准确把握深度学习的真实状态。

最后,完善深度学习的评价体系。应用成效作为衡量深度学习的重要指标,依赖于科学的评价体系。评价是国外深度学习一个非常重要的研究主题,近十年的典型特征是重视多元化、过程性和表现性评价,但尚未形成较为完善的评价体系。依据比格斯和唐提出的“教与学一致性建构”原则^[34],以及威廉与弗洛拉·休利特基金会提出的深度学习框架^[35],深度学习的评价应至少围绕学习结果和学习过程两个视角,个体、小组和群体三个层面,知识掌握、内容理解、能力发展、迁移应用、学习体验、学习投入等具体维度展开,并明确各维度的评价方式和评价标准,从而为深度学习的有效评价提供依据。

深度学习以注重批判性理解、强调信息整合、着意迁移运用为典型特征,指向高阶思维和问题解决能力培养的复杂问题,尝试给出解决方案,已经成为我国变革学习方式^[36]、落实核心素养和培养创新人才^[37]的重要抓手。现阶段面临的关键问题是解决“如何做”的问题,即寻找深度学习与学科融合的实践路径。因此,本文旨在通过系统分析国外近十年实证研究中的有效策略和成功案例呈现他山之石,为我国深度学习概念框架的确立、实践情境的创建、测量工具的选择、评价方案的制定等提供有益参考。

[参考文献]

- [1] 亚当斯·贝克尔,卡明斯,戴维斯,等. 新媒体联盟地平线报告:2017 高等教育版[J]. 北京广播电视大学学报, 2017(2):1-20.
- [2] LUEG R, LUEG K, LAURIDSEN O. Aligning seminars with Bologna requirements: reciprocal peer tutoring, the solo taxonomy and

- deep learning[J]. *Studies in higher education*, 2016, 41(9):1674-1691.
- [3] 祝智庭, 彭红超. 深度学习:智慧教育的核心支柱[J]. *中国教育学报*, 2017(5):36-45.
- [4] 刘选, 田党瑞, 汪燕. 我们需要什么样的实证研究:类型与应用模式——以《现代远程教育研究》2010—2017年195篇实证研究论文为例[J]. *现代远程教育研究*, 2018(4):49-58.
- [5] 谢幼如, 李克东. 教育技术学研究方法基础[M]. 北京:高等教育出版社, 2006.
- [6] DINSMORE D L, ALEXANDER P A. A critical discussion of deep and surface processing: what it means, how it is measured, the role of context, and model specification[J]. *Educational psychology review*, 2012, 24(4): 499-567.
- [7] POSTAREFF L, PARPALA A, SARI LINDBLOM-YLÄNNE. Factors contributing to changes in a deep approach to learning in different learning environments[J]. *Learning environments research*, 2015, 18(3): 315-333.
- [8] PHAN H P. Relations between goals, self-efficacy, critical thinking and deep processing strategies: a path analysis[J]. *Educational psychology*, 2009, 29(7): 777-799.
- [9] PENG J, WANG M, SAMPSON D. Visualizing the complex process for deep learning with an authentic programming project[J]. *Educational technology & society*, 2017, 20(4): 275-287.
- [10] PEGRUM M, BARTLE E, LONGNECKER N. Can creative podcasting promote deep learning? The use of podcasting for learning content in an undergraduate science unit[J]. *British journal of educational technology*, 2015, 46(1): 142-152.
- [11] TURNER M, BASKERVILLE R. The experience of deep learning by accounting students [J]. *Accounting education*, 2013, 22(6): 582-604.
- [12] GREEN G P, KELLY B D, PETERSON D J, et al. Using integrative graphic assignments to promote deep learning of the market mechanism[J]. *Journal of economic education*, 2015, 46(1): 28-44.
- [13] LYNCH R, MCNAMARA P M, SEERY N. Promoting deep learning in a teacher education programme through self- and peer-assessment and feedback[J]. *European journal of teacher education*, 2012, 35(2): 179-197.
- [14] DUMMER T J B, COOK I G, PARKER S L, et al. Promoting and assessing 'deep learning' in geography fieldwork: an evaluation of reflective field diaries[J]. *Journal of geography in higher education*, 2008, 32(3):459-479.
- [15] MARTON F, SALJO R. On qualitative differences in learning: i-outcome and process [J]. *British journal of educational psychology*, 1976, 46(1):4-11.
- [16] ENTWISTLE N, RAMSDEN P. Understanding student learning[M]. New York: Croom Helm, 1982.
- [17] BIGGS J. Student approaches to learning and studying. Research monograph [M]. Australian Council for Educational Research Ltd. Radford House, Frederick St. Hawthorn 3122, Australia. 1987.
- [18] ROCCONI L M, RIBERA A K, LAIRD T F N. College seniors' plans for graduate school: do deep approaches learning and holland academic environments matter?[J]. *Research in higher education*, 2015, 56(2):178-201.
- [19] GOMOLL A, HMELO-SILVER C, SELMA ŠABANOVIĆ, et al. Moving apart and coming together: discourse, engagement, and deep learning[J]. *Educational technology & society*, 2017, 20(4):219-232.
- [20] TSAUSHU M, TAL T, SAGY O, et al. Peer learning and support of technology in an undergraduate biology course to enhance deep learning[J]. *Cell biology education*, 2012, 11(4):402-412.
- [21] ADLER J, HOSSAIN S, STEVENSON M, et al. Mathematics for teaching and deep subject knowledge: voices of mathematics enhancement course students in England[J]. *Journal of mathematics teacher education*, 2014, 17(2):129-148.
- [22] PUGH K J, LINNENBRINK-GARCIA L, KOSKEY K L K, et al. Motivation, learning, and transformative experience: a study of deep engagement in science[J]. *Science education*, 2010, 94(1): 1-28.
- [23] STOTT A, HATTINGH A. Conceptual tutoring software for promoting deep learning: a case study [J]. *Journal of educational technology & society*, 2015, 18(2):179-194.
- [24] SALMERÓN L, NAUMANN J, GARCÍA V, et al. Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study[J]. *Journal of computer assisted learning*, 2017, 33(3): 222-233.
- [25] BIGGS J, KEMBER D, LEUNG D Y P. The revised two-factor study process questionnaire: R-SPQ-2F [J]. *British journal of educational psychology*, 2001, 71(1), 133-149.

- [26] National Survey of Student Engagement Institute. The NSSE 2000 report: national benchmarks of effective educational practice[R]. Bloomington, IN: Indiana University Center for Postsecondary Research, 2000.
- [27] ENTWISTLE N, MCCUNE V. The conceptual base of study strategies inventories[J]. Educational psychology review, 2004, 16:325-345.
- [28] TAIT H, ENTWISTLE N, MCCUNE V. ASSIST: a reconceptualisation of the approaches to studying inventory. In C. Rust (Ed.), Improving student learning: improving students as learners[R]. Oxford: Oxford Centre for Staff and Learning Development, 1998:262-271.
- [29] PINTRICH P R, SMITH D A F, GARCIA T, et al. Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ)[J]. Educational and psychological measurement, 1993, 53(3):801-813.
- [30] ENGEL S, PALLAS J, LAMBERT S. Model United Nations and deep learning: theoretical and professional learning [J]. Journal of political science education, 2017, 13(2): 171-184.
- [31] 张思琦, 张文兰, 李宝. 国外近十年深度学习的研究现状与发展趋势——基于引文分析及共词矩阵的知识图谱分析[J]. 远程教育杂志, 2016, 35(2): 64-72.
- [32] 卜彩丽. 深度学习视域下翻转课堂教学理论与实践研究[D]. 西安: 陕西师范大学, 2018.
- [33] 李玉斌, 苏丹蕊, 李秋雨, 等. 面向混合学习环境的大学生深度学习量表编制[J]. 电化教育研究, 2018(12): 94-101.
- [34] 约翰·比格斯, 凯瑟琳·唐. 卓越的大学教学——建构教与学的一致性[M]. 王颖, 丁妍, 高洁, 译. 上海: 复旦大学出版社, 2015.
- [35] William and Flora Hewlett Foundation. Deeper learning strategic plan summary education program [EB/OL]. [2018-11-30]. https://hewlett.org/wp-content/uploads/2016/09/Education_Deeper_Learning_Strategy.pdf.
- [36] 何克抗. 深度学习: 网络时代学习方式的变革[J]. 教育研究, 2018(5): 111-115.
- [37] 杨玉琴, 倪娟. 深度学习: 指向核心素养的教学变革[J]. 当代教育科学, 2017(8): 43-47.

A Literature Review of 10-year Empirical Studies of Deep Learning Abroad: Themes, Context, Methods and Outcomes

SHEN Xiajuan, ZHANG Baohui, ZENG Ning

(School of Education, Shaanxi Normal University, Xi'an Shaanxi 710062)

[Abstract] In order to clarify the research progress of deep learning in education abroad, 85 papers are selected from Web of Science, Springer, ScienceDirect and other databases (2008-2018). Then a coding system is established based on the core elements of empirical research for content analysis. The results indicate that: (1) the research themes mainly focus on three aspects of deep learning strategies, approaches and evaluation; (2) The conceptual frameworks of deep learning proposed by Biggs, Marton & Saljo and Entwistle are widely recognized; (3) Deep learning is closely used in various subject contexts, but the research objects are mostly college students; (4) Research data is mainly analyzed by quantitative methods, which is supplemented by qualitative methods, while mixed analysis is increasingly valued; (5) Self-report scale, coding standard, conditional measurement and eye movement tracking are the main measuring methods of deep learning; (6) The effectiveness of deep learning is highlighted in academic performance, knowledge understanding, learning experience, the development of thinking and competence.

[Keywords] Deep Learning; Empirical Study; Literature Review; Research Themes; Research Method; Research Context