

人工智能赋能的可信同伴互评模型构建与验证

□孔维梁 于晓利 韩淑云 邓敏杰

摘要: 同伴互评是培养高阶思维、提升学习绩效的重要学习策略。然而在教学实践中,同伴互评存在严重的低信任问题,即学习者对同伴的评价存在质疑或漠视,整体接受度不高。究其缘由,学习者的认知差异性导致低信任问题的关键诱因:客观上,学习者的认知差异性导致同伴评价的不一致现象;主观上,学习者面对评价不一致现象时会产生确认偏误。人工智能赋能的可信同伴互评模型,使用可信系数标明评价的有效性,引导学习者建立正确的认同与信任,可以降低学习者的确认偏误。该模型的智能化实现过程分为4个阶段:评价及可信系数的表征、评分关系加权图构建、基于加权随机游走算法的同伴间认知水平关系挖掘、可信系数计算及评价反馈。基于该模型的教学实践表明:其能够依据评价者与被评者的相对认知水平为评价计算合理的可信系数;有助于提高学习者对同伴评价的接受度,在感知有用性、行为意愿2个维度上显著优于传统同伴互评模型;对学习者的批判性思维倾向的培养具有显著的正向影响,在分析性、系统性、求知欲和思想开放性4个子维度上均有显著提升。

关键词: 人工智能;同伴互评;认知差异;低信任度;可信系数

中图分类号:G434 **文献标识码:**A **文章编号:**1009-5195(2023)03-0093-10 doi:10.3969/j.issn.1009-5195.2023.03.011

基金项目: 河南省哲学社会科学规划项目“数据驱动的在线协同知识建构干预策略研究”(2021BJY021);河南省高等学校重点科研项目“自我调节学习视角下在线学习支持服务研究”(21A880009)。

作者简介: 孔维梁,博士,讲师,硕士生导师,河南师范大学教育学部(河南新乡 453007);于晓利,硕士研究生,河南师范大学教育学部(河南新乡 453007);韩淑云,博士研究生,华中师范大学人工智能教育学部(湖北武汉 430079);邓敏杰,博士,副教授,硕士生导师,河南师范大学教育学部(河南新乡 453007)。

一、引言

适时、适度的评价能为教学提供重要的导向,也对学习者认识自我、激发动机起着良好的促进作用(黄涛等,2021)。同伴互评作为一种新兴评价方式,通过组织学习者对同伴或其作品进行评价,为学习者观点的交互创新、知识的迁移应用创造了良好的环境(张红英等,2019;Gamage et al.,2021)。同伴互评是实现“以学习者为中心”的有效教学策略,在培养高阶思维能力(钟启泉,2020;Lin et al.,2021)、提高知识建构水平(许玮等,2022)和激发学习动机(Rico-Juan et al.,2019)等方面均有显著效果。

虽然同伴互评具有诸多优势,但其在教学实践中仍面临严重的低信任度问题,即学习者对同伴的评价存在质疑或漠视,整体接受程度不高(Luo et al.,2014)。信任度是影响同伴互评成功与否的核心因素。低信任度容易将潜在的积极学习体验转化为消极学习体验,降低学习者参与同伴互评的意愿,从而阻碍同伴互评有效性的发挥(Pond et al.,1997;

Admiraal et al.,2014)。目前,由学习者的认知差异性导致的低信任度问题仍普遍存在,依然是教育教学中亟待解决的关键问题。

本文聚焦由学习者的认知差异性导致的低信任度问题,引入人工智能技术(Artificial Intelligence, AI)构建可信同伴互评模型,使用可信系数标明评价的有效性,帮助学习者建立更合理的注意与认同,以提升对同伴互评的接受程度。此外,本研究还将对上述模型进行实践验证,从学习者的同伴互评接受度水平和批判性思维倾向两个方面对模型的有效性进行检验。

二、相关研究综述

1. 同伴互评促进学生的发展

同伴互评也被称为同伴反馈或互动评价,指学习者在相同的学习环境中,对该环境中的其他学习者或其作品进行的等级评分或者评论,是学习者互相评价的过程(Zundert et al.,2010)。在同伴互评中,学习者同时扮演评价者和被评价者的双重角

色。一方面,作为评价者,学习者在已有知识和经验基础上,经过内化、迁移等认知加工过程,对相同环境中同伴或其作品进行评价;另一方面,作为被评价者,学习者在同伴的评论或提问下,将隐藏的思维过程外化,清晰地观察到自己的推理和分析过程,能更好调动元认知参与和自我调节(毕景刚等,2020; Chien et al., 2020)。

研究表明,同伴互评在激发动机和提高学习绩效方面具有良好的效果。Rotsaert等(2018)指出同伴互评可以作为形成性评价,能有效改善教学中的薄弱环节,有助于提高学习者的学习成绩。Heyman等(2011)认为同伴互评是以学生为中心的教学策略,能够给予学生主人翁意识,对激发学习动机、促进学习参与度具有良好的效果。白清玉等(2016)基于活动理论对同伴互评的成效进行了研究,发现互评活动对学生的成绩有显著的积极影响。此外,黄国祯等(Hwang et al., 2014)通过教育游戏应用证实了同伴互评可以有效提高学习者的学习成绩和学习动机。

同时,同伴互评也是培养学习者高阶思维发展的有效策略之一。高阶思维是产生在较高层次上的心智活动或认知能力,主要包括问题解决能力、批判性思维(倾向和能力)、决策力以及评价能力等(潘星竹等,2019)。Yastibas等(2015)指出在同伴互评过程中,不同观点的碰撞能够强化学习者的反思和逻辑推理意识,促进认知结构不断重组,推动学习者问题解决能力的发展。汪琼等(2019)对MOOC课程中同伴互评数据进行深入分析后发现,即使在无干预的状态下,大部分学习者的推理、综合、解释等批判性思维能力都有一定程度的提升。Li等(2019)指出相较于教师评价,处于同等地位的同伴之间进行的互评能够削弱评价者的权威地位,减轻被评价者的心理压力,有利于促进评价双方进行深入辩论,对学习者的求知欲、自信心等批判性思维倾向有积极影响。张生等(2021)认为评他能力是学评融合新理念的核心内容之一,同伴互评是培养学习者辨别好坏的过程,能有效提升学习者的评他能力。

2. 同伴互评面临的低信任困境

理论与实践证明,同伴互评可以促进学习者的知识建构和高阶思维培养。然而,在实际教学应用中,学习者对同伴的评价往往持质疑或不信任态度,同伴互评面临严峻的低信任困境(吴遐等,2020)。

为明晰触发低信任的诱因,学者们从不同的视角

开展归因探究。柏宏权(2018)发现友谊和情感冲突等人际关系会影响学习者对评价公平性的感知,而对公平性的负面感知会降低评价质量,增加同伴对互评的不信任度。李红霞等(2020)指出评价不一致现象是引起低信任问题的关键诱因之一,其中,评价标准的不统一会降低对评价者的约束力,增加互评的随意性,造成不同评价者对相同被评者的评价不同甚至是完全相反,即评价不一致现象。李菲茗等(2018)也证实了评价不一致现象是低信任问题的重要根源,同时指出评价者认知水平的差异性会造成同伴互评不一致现象的主要原因之一。

在归因研究的基础上,学者们设计了相应的模型与方案以解决同伴互评的低信任问题。为解决人际关系渗入造成的低信任问题,学者们提出匿名评价模型,使评价者能够更自由地表达自身意见,提供更坦率和更具有针对性的反馈,而被评价者也可以更多地关注评价内容,而不是评价者的身份(Van Gennip et al., 2010; Chang et al., 2011)。然而,Yu等(2016)指出,匿名评价会促使学习者的自我意识减弱和对自我社会行为管制的放松,降低社会责任感,导致评价随意性的增加。为解决随意性以及评价标准不统一造成的低信任问题,梁云真(2018)在匿名的基础上提出基于量规的同伴互评模型,认为评价量规能够为评价者的评价活动提供支架辅助,能有效规范评价者行为,帮助评价者明确评价要求,减少评价的不一致性和主观随意性。

综观上述研究,学习者人际关系的渗入、评价标准不统一和认知差异性等因素是导致同伴互评低信任问题的关键诱因;同时,匿名、量规等互评模型的提出也有效解决了人际关系和评价标准不统一导致的低信任问题。然而,由同伴认知差异性造成的低信任问题却少有研究关注,仍是同伴互评实践中亟待解决的关键问题。

3. 认知差异性与低信任问题

认知差异性是指不同的个体在不同的认知维度上有着各自独特的智能优势。在同伴互评中,认知差异性会导致相同评价者在不同评价维度或者不同评价者在相同评价维度上的评价存在差异,即评价不一致现象,进而造成低信任问题。刘兴华等(2018)指出,受自身认知水平或偏好的影响,同伴互评结果很难做到较大程度上的合理统一。

评价不一致现象是引发学习者对同伴评价低信任的重要根源,确认偏误理论则对这一问题给予了

机理性的解释。确认偏误(又称证实偏差)是指无论合乎事实与否,人们总是趋向于看见期望看见的,相信愿意相信的,当认定了一个观点,大脑会持续、有选择地去寻找证据来证明该观点是正确的,而对那些证明该观点错误的证据,则会有选择地忽略或解读(Marsh et al., 2007)。

在同伴互评中,学习者会为自身覆盖一层自我认识的强化滤镜。当面对诸多不一致的同伴评价时,学习者更认同对自己有利的描述或与自身观点一致的同伴评价,而对于不一致观点的评价则本能地认为对方是能力不足或恶意评价,从而导致其所认同的评价被高估和不认同的评价被低估。研究者指出在评价不一致现象中,学习者难以甄别有效性评价,乐于接受认同自己或支持自身观点的评价,而对持有不一致观点的评价持怀疑态度,继而否定并忽视该评价(丛玺梦, 2020; 李红霞等, 2020)。由此可见,学习者在无外界干扰的情况下,易于按照支持自我的角度选择、解读评价反馈,即引发确认偏误,导致学习者对评价的整体信任度不高。

综上,认知差异性引起的低信任问题主要体现在两方面:(1)学习者的认知差异性导致同伴评价的不一致现象,这是引起低信任问题的客观因素;(2)学习者面对评价不一致现象时产生确认偏误,这是引起低信任问题的主观因素。由此,为解决认知差异性带来的低信任问题,可以从上述两方面入手。

三、人工智能赋能的可信同伴互评模型构建

1. 赋予评价有效性的可信系数

个体差异理论指出认知差异性普遍存在且不可避免,评价不一致这一客观现象难以消除,故而,从主观因素上降低学习者的确认偏误成为解决低信任问题的可行方案。为此,本研究提出为同伴评价提供可信系数,即为每条评价标明其客观有效性,值越大则表明评价越有效。可信系数为评价的有效性赋予一个权威性声明,使有效的评价引起学习者的注意与认同,即使评价与自身观点冲突,被评价者也不会轻易否定该评价,进而促使学习者对评价进行深层次的反思或与评价者进行交流,从而降低或消除学习者对评价的确认偏误。

可信系数为解决由确认偏误引起的低信任问题提供了一种可行方案,而如何科学、客观地计算可信系数成为该方案的核心。有研究发现认知水平对评价有显著影响,认知水平高的学生对作品的评价

更为准确,评价的质量也就更高(李晓菲, 2018; Meek et al., 2017)。此外,成绩优秀者的评语更为具体全面,认知广度更大,也更能同时兼顾课程知识的多个维度(李梅等, 2016; 刘迎春等, 2019)。

基于此,本研究提出将评价者和被评价者的认知水平关系作为可信系数的计算依据,即相对于被评价者,如果评价者在某方面的认知水平越高,则其评价在该方面的可信系数也就越高,反之亦然。可信系数与评价者和被评价者的相对认知水平关系密切相关,因此,相同评价者对不同被评价者的评价,其可信系数是不同的。例如,学习者A、B、C的认知水平分别为0.8、0.6、0.9,则A对B评价的可信系数为 $0.8/0.6 \approx 1.33$;而A对C评价的可信系数则为 $0.8/0.9 \approx 0.89$ 。

2. 人工智能赋能的可信同伴互评模型

可信系数的计算需要依据评价者和被评价者的认知水平关系,面对大量学习者和多维度认知水平,仅依靠个体的智慧与能力难以发现及厘清两两学习者在多维认知水平上的相对关系,也就难以为评价计算精准、客观的可信系数。人工智能技术的发展为可信系数的计算提供了技术支撑,其通过对学习者间的评分信息进行深度挖掘,以统一的尺度标明所有学习者的认知水平,从而极大推进了可信系数的智能化计算。

本研究在匿名和量规评价的基础上,以可信系数为着力点,以智能算法为支撑,提出了人工智能赋能的可信同伴互评模型(如图1所示),以解决认知差异造成的低信任问题。模型主要包括四大功能模块:

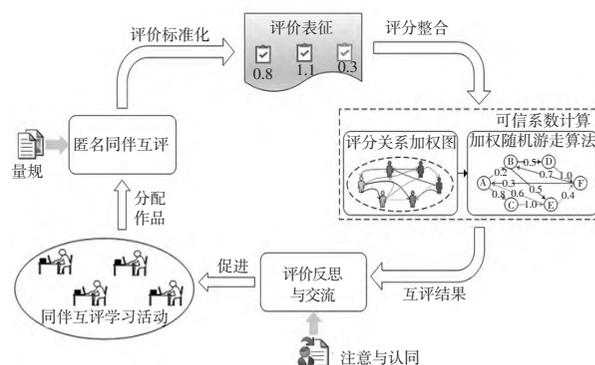


图1 AI赋能的可信同伴互评模型

匿名同伴互评模块。系统随机为每个作品分配固定数量的评价者,评价者则需要在规定时间内按照给定量规对作品进行在线匿名评价。匿名评价使得同伴互评的公正性得以保证,而量规则消除了同伴评价的随意性。

评价表征模块。该模块将同伴的每条评价分为评分与评语两部分，并对评价进行形式化表征。其中，评分是对作品的量化认同程度，而评语是质性的说明或建议。

可信系数计算模块。该模块主要包含2个步骤：首先，将所有的评价关系和评分数据转换成有向图，生成评分关系加权图；其次，利用人工智能技术中的加权随机游走算法，从评分关系加权图中挖掘出学习者间的认知水平关系，并根据评价者与被评价者的认知水平关系，为评价计算可信系数。

评价反思与交流模块。系统将每条评价及其可信系数反馈给被评价者，被评价者则在可信系数的基础上修正自身的认知注意力，并参照评语有选择地批判或接受相关评价，对有疑惑的地方与相应评价者进行在线匿名交流，在交流中修正认知不足并促进知识体系的深层次建构。

四、可信同伴互评模型的智能化实现

可信同伴互评模型的实现核心与难点在于可信系数的计算，而可信系数的计算本质上可以转换为挖掘评价者与被评价者间的相对认知水平关系。

在众多人工智能算法中，加权随机游走算法是一种用来在加权图上挖掘所有节点相对重要性的有效算法。该算法具有统一尺度、并行性好、鲁棒性高、不易陷入局部极小值等优点，在网页重要性排行、全局最优化等领域得到了广泛的应用（马慧芳等，2018），尤为适合认知水平关系的挖掘。为此，研究引入加权随机游走算法进行可信系数的计算，从而完成可信同伴互评模型的智能化实现。其核心实现过程分为4个阶段：评价及可信系数的表征、评分关系加权图构建、同伴间认知水平关系挖掘和可信系数计算及评价反馈。

1. 评价及可信系数的表征

评价及可信系数的表征是定义评价数据和可信系数的结构化形式，该步骤是数据处理的前提与基础。研究所关注的同伴评价包含两部分内容：评分与评语。其中，评分是一个多维度的打分列表，其维数由所使用的评价量规决定，表示评价者在多个量规指标上对被评作品的认同程度；评语则是一段文字说明，是对评分的解释说明或修改建议等，评语是被评价者与评价者进行交流反思的基础。如表1所示，依据3个指标进行评分，每个指标满分为10分，最后给出了整体评语。

表1 同伴评价示例

评价	指标1	指标2	指标3	评语
r(B,A)	7	5	6	选题尚可,整体上格式基本规范,但图片过小不够清晰,数据稍显老旧
r(A,F)	8	5	6	选题较新颖,色彩较清晰,但数据陈旧且语言表达水平有待提高
.....
r(G,F)	8	5	7	选题较新颖,格式一般,内容过多地借鉴了网络资源且不够完整

同伴评价的形式化表示如公式1和公式2所示。其中， $r(u,v)$ 表示学习者u对学习者的v的作品的评价， $rs(u,v)$ 和 $rc(u,v)$ 分别表示评价的评分和评语，k为评价的指标数。例如，学习者B对学习者的A作品的评价可形式化表示为 $rs(B,A) = \{7, 5, 6\}$ ， $rc(B,A) =$ “选题尚可，整体上格式基本规范，但图片过小不够清晰，数据稍显老旧”。

$$r(u,v) = \{rs(u,v), rc(u,v)\} \quad (1)$$

$$rs(u,v) = \{rs_1(u,v), rs_2(u,v), \dots, rs_k(u,v)\} \quad (2)$$

2. 评分关系加权图构建

评分关系加权图是同伴间评价关系的形式化表示，是可信系数计算的结构化数据支撑。评分关系加权图是一个加权有向图，例如表1对应的评分关系加权图如图2所示。图中节点（“人”形符号）表示学习者或其作品，当表示出链时为评价者，表示入链时为作品；有向链接则表示学习者间的评价关系，链接上的权值为评价的评分。

图2中每个作品均被2个评价者评价，如学习者A分别对学习者的B和学习者的F的作品进行了评价（即A的出链），评分分别为 $\{7, 5, 7\}$ 和 $\{8, 5, 6\}$ ，而

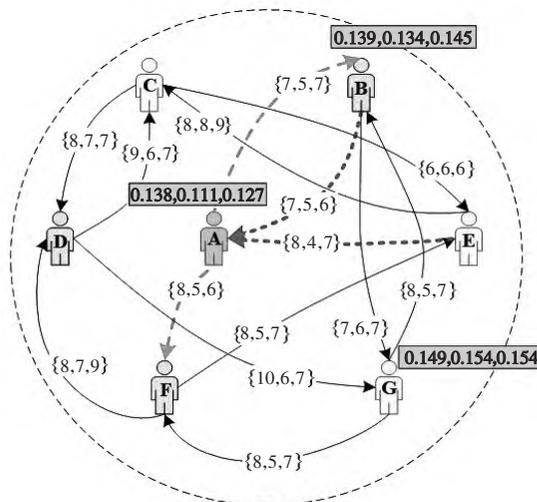


图2 评分关系加权图示例

学习者A的作品又被学习者B和学习者E评价(即A的入链)。

3. 同伴间认知水平关系挖掘

在评分关系加权图的基础上,为计算可信系数,需要挖掘同伴间的认知水平关系。加权随机游走算法能够以统一的尺度标识出所有节点的相对重要性,适用于同伴间认知水平关系的挖掘。

加权随机游走算法的基本原理为:节点被链接(图2中入链)的分值越大,该节点的重要性就越高;重要性越高的节点对别人的链接(图2中出链)也越重要。应用于评分关系挖掘,加权随机游走算法会使用统一的尺度标识出所有同伴的认知水平,该值为一个向量,其维数与评分保持一致,研究中以PR表示(图2中节点上方框内的值)。例如,图2中学习者A的认知水平 $PR(A) = \{0.138, 0.111, 0.127\}$ 。基于加权随机游走算法的同伴间认知水平关系挖掘的步骤如下:

(1) 节点初始化。为每个节点的认知水平PR赋初始值,如公式(3)所示。其中, $PR^0(u)$ 表示节点u的初始值, $pr^0(u_k)$ 表示节点u在指标k上的初始值,N表示所有节点的个数,初始时每个学习者在每个评价指标上的初始化值均为 $1/N$ 。例如,图2中有7个学习者,则学习者的初始认知水平 $PR^0(u) = \{1/7, 1/7, 1/7\}$ 。

$$PR^0(u) = \{pr^0(u_1), pr^0(u_2), \dots, pr^0(u_k)\} \\ = \{1/N, 1/N, \dots, 1/N\} \quad (3)$$

(2) 更新节点值。根据评分关系加权图为每个节点更新PR值,方法是:将每个节点当前的PR值以加权的方式分配到其所有出链上,而每个节点的新PR值计算方法如公式(4)所示,公式包含两部分,前一部分是其自身入链的PR值总和,后一部分是随机转移值。其中, $pr^{t+1}(u_k)$ 表示节点u在指标k上、第t+1次迭代时的PR值,I表示节点u的入链节点集合(即节点u的所有评价者集合), α 表示随机跳转概率,一般取经验值0.85(彭珠,2009)。 $w_k(v,u)$ 表示在指标k上,节点v对节点u评分的权重,其计算方法如公式(5)所示。其中, $r_k(v,u)$ 表示在指标k上节点v对节点u的评分,O表示节点v的出链节点集合(即被节点v评价的所有节点集合)。

$$pr^{t+1}(u_k) = \alpha \sum_{v \in I} w_k(v, u) \cdot pr^t(v_k) + (1 - \alpha) / N \quad (4)$$

$$w_k(v, u) = r_k(v, u) / \sum_{m \in O} r_k(v, m) \quad (5)$$

以图2中节点A为例,2个评价者B和E的评分为

B: {7,5,6}, E: {8,4,7}。在第一轮更新时,节点A入链的PR值分别为B→A: (0.061,0.055,0.056), E→A: (0.061,0.041,0.053),根据公式(4)可计算节点A的第一轮PR值为{0.143,0.117,0.131}。

(3) 节点的认知水平计算。迭代执行步骤(2),直至所有节点的PR值逐渐收敛或达到迭代次数为止,此时获得的每个节点的PR值即为其认知水平。

4. 可信系数计算及评价反馈

挖掘出所有同伴间的认知水平关系之后,即可以认知水平关系为基础计算评价的可信系数。评价者的相对认知水平越高,其评价越有效,可信系数也越高。为此,评价的可信系数被定义为该评价的评价者与被评价者的认知水平的比值,计算方式如公式(6)所示。其中, $C(u,v)$ 表示学习者u对学习评价者v评价的可信系数, $PR(u)$ 表示学习者u的认知水平。例如,图2中节点B对节点A评价的可信系数为(0.139/0.138, 0.134/0.111, 0.145/0.127) = (1.007, 1.207, 1.142),其中,可信系数大于1表明评价者在该方面的认知水平高于被评价者,低于1则相反。

$$C(u, v) = \left\{ \frac{pr(u_1)}{pr(v_1)}, \frac{pr(u_2)}{pr(v_2)}, \dots, \frac{pr(u_k)}{pr(v_k)} \right\} \quad (6)$$

及时有效的反馈可以触发学习者在学习过程中积极的反应,发现与弥补现有认知错误与不足,促进知识体系全面、深层次的建构。为了促进同伴互评的有效性,降低学习者对评价的确认偏误,本研究将评价与可信系数绑定并同时反馈给学习者,以帮助学习者对评价建立正确的认知,即反馈包含评分、评语以及评价可信系数3个部分,例如,图2中节点A的一条评价反馈为{(7,5,6),“选题尚可……”,(1.007,1.207,1.142)}。

五、可信同伴互评模型的实践验证

1. 实验设计

研究以某师范大学选修“现代教育技术”课程的20级物理学1班和物理学2班的本科生作为研究对象。1班的54人作为实验组,采用可信同伴互评,2班的55人作为控制组,采用传统同伴互评,两组学习者都有至少一学期的在线学习平台使用经验。课程内容以主题形式组织,并采用“线下学习+线上评价交流”的混合式教学模式。其中,线下学习是指学习者在课堂学习主题知识并完成作

品;线上评价交流是学习者对同伴上传作品进行在线匿名评价并与自身作品的评价者进行匿名交流。

为探究本文提出模型的有效性,研究将从可信系数计算的合理性、学生的同伴互评接受度、学生的批判性思维倾向3个方面进行实践验证。实验中量规满分100分,包含内容性(40分)、教学性(30分)、呈现度(30分)3个一级指标,一级指标又分别细分为2、3、3个二级指标。学生的同伴互评接受度和批判性思维倾向均使用量表测量。其中,同伴互评接受度量表是在胡畔等人量表(胡畔等,2019)的基础上改进得到,包括感知有用性等3个测量维度和18个测量题项,采用5级李克特量表测量。批判性思维倾向量表选择Facione等(1992)开发的加利福尼亚批判性思维倾向量表,选用了思想开放性、分析性等5个维度,共50个测量题项,采用6级李克特量表测量。研究使用Cronbach's α 系数和KMO系数对两个量表的信效度进行检验,结果均高于0.75,表明量表具有较高的信效度。

研究的实验流程如图3所示。课程开始之前,使用量表对两组学生进行同伴互评接受度和批判性思维倾向的前测,并向学习者讲解学习方式、评价量规等相关知识。课程学习过程中,要求学生每学习一个主题后,需完成一个作品并提交至在线互评平台,随后互评系统为每个作品随机分配3个评价者(每个评价者也评3个作品),开展同伴互评活动。实验组在互评后会得到系统反馈的评分、评语和可信系数。而控制组则得到评分和评语。得到反馈后,两组学生均可通过在线留言工具与匿名评价者开展交流,并对作品进行修改后再次上传。学期末,再次使用量表对学

习者进行同伴互评接受度和批判性思维倾向的后测,并总结反思课程的实施效果。

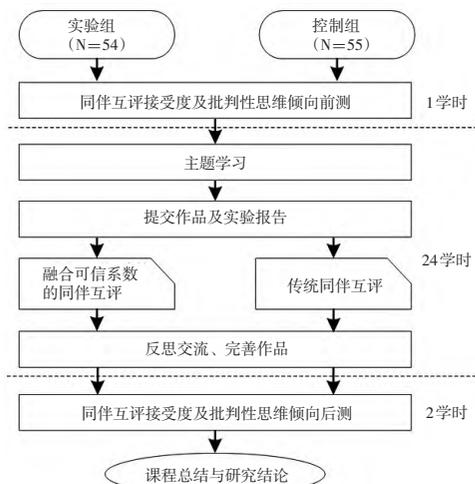


图3 实验流程图

2. 研究结果与分析

(1) 可信系数计算结果

为验证模型可信系数计算的合理性,研究随机选取了第6周S27学生及其3位评价者(S8, S38, S47)的原始评分和可信系数,如表2所示。从表中数据可知:其一,同一评价在不同评价指标上的可信系数存在差异。例如,S8对S27的评价可信系数分别是{(1.01, 1.14), (1.00, 1.09, 0.99), (0.96, 1.02, 1.17)},这一结果表明评价者在每个评价维度上的认知水平并不相同,符合学习者个体内的认知差异性原则。其二,不同评价在相同评价指标上的可信系数并不相同。例如,S27在一级指标1的第2子指标上的可信系数分别为1.14、0.99、0.98,这一结果表明不同的评价者在相同维度上的认知水平并不相同,符合学习者个体间的认知差异性原则。

表2 同伴互评的原始评分及可信系数示例

被评者	评价者	原始评分	评价的可信系数	加权得分
S27	S8	{(17, 14), (8, 6, 7), (8, 8, 6)}	{(1.01, 1.14), (1.00, 1.09, 0.99), (0.96, 1.02, 1.17)}	{(16.66, 14.95), (8.65, 6.00, 7.31), (7.66, 8.32, 6.00)}
	S38	{(17, 16), (9, 6, 8), (8, 9, 5)}	{(1.00, 0.99), (0.95, 1.01, 0.90), (1.07, 0.97, 1.07)}	
	S47	{(16, 15), (9, 6, 7), (7, 8, 7)}	{(1.04, 0.98), (0.91, 1.01, 1.00), (1.04, 1.02, 1.09)}	
S8	S5	{(17, 19), (9, 5, 9), (9, 8, 7)}	{(0.99, 0.91), (0.91, 1.07, 0.88), (0.91, 1.02, 0.96)}	{(16.66, 16.59), (8.66, 5.92, 7.96), (8.64, 8.00, 6.64)}
	S35	{(16, 16), (8, 7, 8), (9, 8, 7)}	{(1.00, 1.00), (0.99, 0.84, 0.94), (1.03, 1.07, 1.03)}	
	S44	{(17, 15), (9, 6, 7), (8, 8, 6)}	{(0.98, 1.00), (0.99, 1.03, 0.99), (1.09, 0.99, 1.10)}	
S38	S10	{(16, 15), (7, 8, 8), (8, 7, 9)}	{(1.03, 0.99), (1.10, 0.87, 0.89), (1.01, 1.20, 0.90)}	{(16.33, 14.28), (7.32, 6.22, 6.91), (7.96, 7.31, 8.64)}
	S21	{(16, 15), (7, 6, 6), (7, 7, 9)}	{(1.02, 1.03), (1.04, 0.97, 1.17), (1.04, 1.12, 0.96)}	
	S26	{(17, 13), (8, 5, 7), (9, 8, 8)}	{(0.99, 1.14), (1.00, 1.10, 1.01), (0.91, 1.02, 1.04)}	
S47	S15	{(17, 15), (8, 5, 7), (9, 8, 6)}	{(0.96, 1.04), (1.10, 1.09, 1.08), (0.98, 0.99, 1.13)}	{(17.99, 15.63), (8.00, 5.94, 7.91), (8.65, 7.65, 6.30)}
	S17	{(19, 15), (8, 6, 8), (8, 8, 6)}	{(0.94, 1.10), (1.03, 0.96, 0.95), (1.02, 1.03, 1.08)}	
	S29	{(18, 17), (8, 7, 9), (9, 7, 7)}	{(0.96, 0.98), (1.06, 0.92, 0.82), (0.92, 1.10, 0.94)}	

其三,学习者的作品在某评价指标上取得的评分越高,其对他人的评价中在该指标上的可信系数也越大。例如,S8的作品在一级指标1的第2子指标上得分最高(16.59),S27在该指标上的所有评价中,S8的可信系数最高(1.14),这一结果表明学习者的认知水平越高,则其对他人评价中的可信系数也越大,计算符合研究的预期。

由此可见,认知差异性客观存在,并且本研究提出的可信同伴互评模型能够依据评价者与被评价者的相对认知水平为评价计算合理的可信系数。

(2) 可信同伴互评模型对同伴互评接受度的影响

接受度是同伴互评有效性以及学生是否会持续使用该学习方式的重要衡量指标。为探究可信系数对同伴互评接受度的影响,研究从感知有用性、感知易用性和行为意愿3个维度对学生学习前后的接受度进行对比检验。首先,使用独立样本t检验对两组学生的前测结果进行分析,结果如表3所示。可以看出,两组学生在3个维度上的均值接近,显著性值p均大于0.05(分别是0.970、0.929和0.951),未达到显著水平,说明两组学生在学习前具有一致的接受度。

表3 同伴互评接受度前测的独立样本t检验

	N	感知有用性		感知易用性		行为意愿	
		Mean	t	Mean	t	Mean	t
实验组	54	4.01	-0.037	4.02	-0.075	3.89	-0.062
控制组	55	4.03		4.00		3.88	

然后,使用独立样本t检验方法对两组的后测结果进行分析,结果如表4所示。由表可知:在感知易用性维度,两组无明显差异($p=0.906>0.05$);而在其他2个维度上,两组差异性显著(p 分别为0.030和0.031)。结合前后测均值不难发现,学习后实验组在感知有用性和行为意愿2个维度上有所提升,而控制组在2个维度上基本保持稳定或下降。

表4 同伴互评接受度后测的独立样本t检验

检验变量	组别	N	均值 M	标准差 SD	标准误差 SE	t
感知有用性	实验组	54	4.31	0.848	0.115	2.196*
	控制组	55	3.96	0.815	0.109	
感知易用性	实验组	54	4.01	0.812	0.111	-0.118
	控制组	55	4.03	0.828	0.112	
行为意愿	实验组	54	4.07	0.854	0.116	2.187*
	控制组	55	3.72	0.788	0.106	

(注:*表示 $p<0.05$ 。)

实验结果反映出实验组学生在可信系数的引导下,能有效利用同伴评价改进作品,认为同伴评价策略对学习有较大促进作用;而控制组学生则认为同伴评价的作用不大,与评价者进行交流的意愿也较低。

(3) 可信同伴互评模型对批判性思维倾向的影响

为探究模型对学习者的批判性思维倾向的影响,研究对学习者的批判性思维倾向进行了对比。首先,使用独立样本t检验对两组学生的前测结果进行分析,结果显示,两组学生在5个维度上的均值及标准差都接近,p均大于0.05,未达到显著水平,说明两组学生的前测结果一致,具有相同的批判性思维倾向。

然后,研究使用配对样本t检验方法对两组学生的“后测—前测”成绩进行分析,结果如表5所示。由表5可知,两组学生在分析性、系统性和求知欲3个维度上均比学习前有显著提升(差值均值 >0 ,且 $p<0.05$);此外,实验组在思想开放性维度也有显著提升,而控制组没有;并且实验组的总体提升度高于控制组($24.019>16.891$)。结合后测的统计结果可知,实验组在总体结果上得分更高($Mean_{实验}=223.24$, $Mean_{控制}=214.33$)。实验结果表明,可信同伴互评模型相较于传统同伴互评,能够更有效地激发学习者的思想开放性;对学习者的批判性思维倾向发展具有

表5 批判性思维倾向的配对样本t检验

批判性思维倾向	实验组配对差分(后测—前测)				控制组配对差分(后测—前测)			
	N	M±SD	标准误差	t	N	M±SD	标准误差	t
思想开放性	54	3.611±9.659	1.314	2.747**	55	0.055±9.548	1.287	0.042
分析性	54	6.222±10.742	1.462	4.257**	55	5.636±11.720	1.580	3.567**
系统性	54	6.593±11.065	1.506	4.378**	55	5.945±10.219	1.378	4.315**
自信心	54	1.148±9.244	1.258	0.913	55	1.309±8.320	1.122	1.167
求知欲	54	6.444±9.121	1.241	5.192**	55	3.945±12.584	1.697	2.325*
整体	54	24.019±23.159	3.152	7.621**	55	16.891±24.715	3.333	5.068**

(注:*表示 $p<0.05$,**表示 $p<0.01$ 。)

更加显著的正向影响。

六、总结与思考

同伴互评是促进学习者深度知识建构和高阶思维培养的有效策略,然而,低信任问题会削弱学习者参与反思活动的积极性,引发评价焦虑,严重阻碍策略有效性的发挥。针对由同伴认知差异性导致的低信任问题,研究引入人工智能技术,构建可信同伴互评模型,通过可信系数帮助学习者降低确认偏误,更大程度包容与自己观点相互冲突的评价。通过实践证明,相较于传统同伴互评模型,本文提出的模型对学习者的接受度和批判性思维倾向均有更优的正向影响。本研究对同伴互评的教育应用实践有以下启示:

一是注重同伴互评的形成性评价。形成性评价通过诊断教学过程中存在的问题,为正在进行的学习活动提供反馈信息,在改善未来学习方面,比总结性评价具有更好的效果。在同伴互评教学中,评价不应以区分评价对象的优良程度为重心,而是要充分利用评价反馈和讨论,引导学习者通过个人的反思和彼此间的交流来理解所收到的评价,将反馈信息融入自身认知策略调整和知识建构中,最终完成问题的解决、知识的迁移和高阶思维的培养。因此,要想同伴互评有效,它必须是对话的,而不是单向的信息传递过程(Filius et al., 2018)。

二是关注学生对评价的确认偏误。评价的不一致性使得学习者需要直面质量参差不齐,甚至是冲突矛盾的评价。学习者往往会不自觉地以自身情绪或喜好判定同伴评价的好坏,产生确认偏误。确认偏误会导致学习者关注的评价不全面,易使学习者沉浸在主体构造的一个支持自我的世界里。为此,除却技术方面的信息补足,还可以引导学习者养成换位思考的习惯,使用逆向思维来思考决策的对立面甚至其他面的利弊,然后一一进行对比分析,保持从理性和客观的视角去做学习决策。

研究在一定程度上解决了由同伴的认知差异性导致的低信任问题,为充分发挥同伴互评在教学中的独特优势提供了有效的支持。但是,除研究所涉及的认知水平外,评价者的评价能力、评价态度等也是影响评价质量的重要因素,如何整合多方面影响因素,构建更全面完善的可信指标,以解决低信任困境对同伴互评造成的不良影响是研究接下来的关注重点。

参考文献:

- [1]白清玉,张屹,沈爱华等(2016).基于同伴互评的移动学习对小学生学习成效的影响研究——以科学课程为例[J].中国电化教育,(12):121-128.
- [2]柏宏权(2018).基于移动学习环境的大学生同伴互评的公平性研究[J].教育发展研究,38(7):75-81.
- [3]毕景刚,韩颖,董玉琦(2020).技术促进学生批判性思维发展教学机理的实践探究[J].中国远程教育,41(7):41-49,76-77.
- [4]丛玺梦(2020).在线作业同伴互评与教师评分一致性的影响因素研究[D].沈阳:沈阳师范大学:2-8.
- [5]胡畔,蒋家傅(2019).中小学生对数字教材的技术接受度及其影响因素研究[J].现代远程教育,(4):77-83.
- [6]黄涛,赵媛,耿晶等(2021).数据驱动的精准化学习评价机制与方法[J].现代远程教育研究,33(1):3-12.
- [7]李菲茗,李晓菲,黄亚平等(2018).训练对同伴互评评分准确性的影响——以“三维动画设计与建模”课程为例[J].中国远程教育,(5):63-67,78.
- [8]李红霞,赵呈领,蒋志辉(2020).匿名与量规对基于SPOC的混合式同伴互评投入度的影响[J].现代教育技术,30(10):20-27.
- [9]李梅,刘英群,周潜(2016).同伴评价的可信度与特点分析[J].电化教育研究,37(9):48-54.
- [10]李晓菲(2018).同伴互评评分准确性影响因素研究[D].杭州:浙江工业大学:25-27.
- [11]梁云真(2018).基于量规的同伴互评对在线学习认知、情感投入度及学习成效的影响研究[J].电化教育研究,39(9):66-74.
- [12]刘兴华,纪小凌(2018).大学英语写作同伴评分的可行性和有效性研究[J].外语界,(5):63-70.
- [13]刘迎春,朱旭,陈乐(2019).精准教学中基于同伴互评的评价者认知网络分析[J].远程教育杂志,37(1):85-93.
- [14]马慧芳,刘芳,夏琴等(2018).基于加权超图随机游走的文献关键词提取算法[J].电子学报,46(6):1410-1414.
- [15]潘星竹,姜强,黄丽等(2019).“支架+”STEM教学模式设计及实践研究——面向高阶思维能力培养[J].现代远程教育,(3):56-64.
- [16]彭珠(2009).一种求解PageRank问题的修正乘幂法(英文)[J].徐州师范大学学报(自然科学版),27(4):45-47.
- [17]汪琼,欧阳嘉煜,范逸洲(2019).MOOC同伴作业互评中反思意识与学习成效的关系研究[J].电化教育研究,40(6):58-67.
- [18]吴遐,高记,刘兵(2020).以评促学:基于三元交互决定论的同伴互评研究[J].中国远程教育,41(4):58-64,77.
- [19]许玮,祝思璇(2022).同伴互评对学习过程知识建构过程的实证研究——基于时间序列的认知网络分析[J].现代教育技术,32(1):44-53.
- [20]张红英,陈明选,马志强等(2019).基于自评与互评的网

络协作学习贡献度评价[J].现代远程教育研究,31(2):95-102.

[21]张生,王雪,齐媛(2021).评他能力:人工智能时代学生必备的高阶思维能力[J].中国电化教育,(11):24-31.

[22]钟启泉(2020).批判性思维:概念界定与教学方略[J].全球教育展望,49(1):3-16.

[23]Admiraal, W., Huisman, B., & Van de Ven, M. (2014). Self- and Peer Assessment in Massive Open Online Courses[J]. International Journal of Higher Education, 3(3):119-128.

[24]Chang, C. C., Tseng, K. H., & Chou, P. N. et al. (2011). Reliability and Validity of Web-Based Portfolio Peer Assessment: A Case Study for a Senior High School's Students Taking Computer Course[J]. Computers & Education, 57(1): 1306-1316.

[25]Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of Peer Assessment Within the Context of Spherical Video-Based Virtual Reality on EFL Students' English-Speaking Performance and Learning Perceptions[J]. Computers & Education, 146:103751.

[26]Facione, P. A., & Facione, N. C. (1992). The California Critical Thinking Disposition Inventory(CCTDI)[EB/OL].[2022-11-21]. <https://edstruments.org/instruments/california-critical-thinking-disposition-inventory-cctdi>.

[27]Filius, R. M., de Kleijn, R. A. M., & Uijl, S. G. et al. (2018). Strengthening Dialogic Peer Feedback Aiming for Deep Learning in SPOCs[J]. Computers & Education, 125:86-100.

[28]Gamage, D., Staubitz, T., & Whiting, M. (2021). Peer Assessment in MOOCs: Systematic Literature Review[J]. Distance Education, 42(2):268-289.

[29]Heyman, J. E., & Sailors, J. J. (2011). Peer Assessment of Class Participation: Applying Peer Nomination to Overcome Rating Inflation[J]. Assessment & Assessment in Higher Education, 36(5):605-618.

[30]Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving Learning Achievements, Motivations and Problem-Solving Skills Through a Peer Assessment-Based Game Development Approach[J]. Educational Technology Research and Development, 62(2):129-145.

[31]Li, H., Xiong, Y., & Hunter, C. V. et al. (2019). Does Peer Assessment Promote Student Learning? A Meta-Analysis[J]. Assessment & Evaluation in Higher Education, 45(2):193-211.

[32]Lin, H. C., Hwang, G. J., & Chang, S. C. et al. (2021). Facilitating Critical Thinking in Decision Making-Based Professional Training: An Online Interactive Peer-Review

Approach in a Flipped Learning Context[J]. Computers & Education, 173:104266.

[33]Luo, H., Robinson, A., & Park, J. Y. (2014). Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects[J]. Journal of Asynchronous Learning Networks, 18(2):119-126.

[34]Marsh, D. M., & Hanlon, T. J. (2007). Seeing What We Want to See: Confirmation Bias in Animal Behavior Research[J]. Ethology, 113(11):1089-1098.

[35]Meek, S. E. M., Blakemore, L., & Marks, L. (2017). Is Peer Review an Appropriate form of Assessment in a MOOC? Student Participation and Performance in Formative Peer Review[J]. Assessment & Evaluation in Higher Education, 42(6): 1000-1013.

[36]Pond, K., & Ul-Haq, R. (1997). Learning to Assess Students Using Peer Review[J]. Studies in Educational Evaluation, 23(4):331-348.

[37]Rico-Juan, J. R., Gallego, A. J., & Calvo-Zaragoza, J. (2019). Automatic Detection of Inconsistencies Between Numerical Scores and Textual Feedback in Peer-Assessment Processes with Machine Learning[J]. Computers & Education, 140:103609.

[38]Rotsaert, T., Panadero, E., & Schellens, T. (2018). Peer Assessment Use, Its Social Nature Challenges and Perceived Educational Value: A Teachers' Survey Study[J]. Studies in Educational Evaluation, 59:124-132.

[39]Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer Assessment as a Collaborative Learning Activity: The Role of Interpersonal Variables and Conceptions[J]. Learning and Instruction, 20(4):280-290.

[40]Yastibas, G. C., & Yastibas, A. E. (2015). The Effect of Peer Feedback on Writing Anxiety in Turkish EFL (English as a Foreign Language) Students[J]. Procedia Social and Behavioral Sciences, 199:530-538.

[41]Yu, F. Y., & Sung, S. (2016). A Mixed Methods Approach to the Assessor's Targeting Behavior During Online Peer Assessment: Effects of Anonymity and Underlying Reasons[J]. Interactive Learning Environments, 24(7):1674-1691.

[42]Zundert, M. V., Sluijsmans, D., & Merriënboer, J. V. (2010). Effective Peer Assessment Processes: Research Findings and Future Directions[J]. Learning & Instruction, 20(4):270-279.

收稿日期 2022-12-11 责任编辑 李鑫

(下转第112页)

.101.

systems. An analysis of the current state of research on intelligent technology-enabled feedback in terms of objectives, mechanisms of action, and system architecture has revealed that intelligent technology-based feedback aims to bring positive effects on learners' cognition, emotions and attitudes, and metacognition in order to motivate learners to improve their current states and learning outcomes. At the level of feedback mechanism, an automatic feedback based on correct and incorrect responses usually adopts a rule-based matching model to trigger feedback by comparing correct and incorrect answers; the key to personalized feedback is to build a learner model to adapt feedback to different students' characteristics and to realize personalized feedback by combining ontology or using multiple rounds of human-computer dialogue; a human-computer collaborative feedback can be achieved by realizing interaction and complementarity between teachers, students and computers. At the system architecture level, the feedback system consists of three layers: input layer, analysis layer and inference generation layer, generating feedback decisions through a series of algorithmic models and feedback mechanisms, whose interactions with learner models, expert models and domain models determine the degree of accuracy and adaptability of teaching feedback.

Keywords: Intelligent Technology; Instructional Feedback; Feedback Goals; Feedback Mechanism; Human-Computer Collaboration

(上接第 101 页)

Construction and Verification of Trusted Peer Evaluation Model Empowered by Artificial Intelligence

KONG Weiliang, YU Xiaoli, HAN Shuyun, DENG Minjie

Abstract: Peer assessment is an important learning strategy to cultivate learners' high-order thinking and improve learning performance. However, in practice, peer evaluation has a serious problem of low trust, that is, learners question or ignore peer evaluation and the overall acceptance is not high. Cognitive differences among learners are the key contributing factor to this problem: objectively, cognitive differences among learners lead to inconsistent peer evaluations; subjectively, learners may experience confirmation bias when faced with inconsistent evaluations. The trusted peer evaluation model empowered by artificial intelligence uses a credibility coefficient to indicate the effectiveness of the evaluation to guide learners to establish correct identification and trust, which can reduce learners' confirmation bias. The intelligent implementation of this model can be divided into four stages: the representation of evaluation and credibility coefficient, the construction of weighted graph of scoring relationship, the mining of peer cognitive level relationship based on weighted random walk algorithm, the calculation of credibility coefficient and evaluation feedback. The teaching practice based on this model shows results as follows. First, the model can calculate a reasonable credibility coefficient for evaluation based on the relative cognitive level of the evaluator and the respondent. Second, learners' acceptance of peer evaluation has improved, especially in terms of perceived usefulness and behavioral willingness, which are significantly better than traditional peer evaluation models. Third, it has a significant positive impact on the cultivation of learners' critical thinking tendency, of which the four sub dimensions including analysis, systematicness, curiosity and openness have been greatly improved.

Keywords: Artificial Intelligence; Peer Assessment; Cognitive Differences; Low Trust; Credibility Coefficient

.112.