



人工智能赋能教育评价：“学评融合” 新理念及核心要素*

□ 张生 王雪 齐媛

【摘要】

人工智能时代是在以物理世界为主的世界上新增了一个数字世界，人们的生活、生产和学习方式发生了重大改变，生产力的巨大变革必然带来生产关系和上层建筑的变革。为应对这一变革，教育必然从知识技能的培养转向学生核心素养的养成，教育评价也必然转向以促进学生核心素养的发展为核心。本文首先从诊断为核心的评价现实困境出发，提出了学评融合的评价新理念，其次从理论层面介绍了人工智能时代学评融合的重要价值和意义，详细阐述了学评融合的概念、内涵与核心特征，最后从实践层面建构了人工智能时代学评融合的评价框架，提出生成与展示、参与评价活动、反思与改进三个学评融合理念下评价作为高阶思维学习活动的核心阶段，并尝试构建各阶段的核心要素，为数字世界学生评价能力的发展与测量、评价系统的研究与研发以及教学实践活动提供具体、可操作的依据。

【关键词】 学评融合；评价理念；核心素养；人工智能；评价改革；评价能力；教育评价

【中图分类号】 G40-058.1

【文献标识码】 A

【文章编号】 1009-458 x (2021)2-0001-09

DOI:10.13541/j.cnki.chinade.2021.02.001

一、问题提出

中共中央、国务院2020年印发了《深化新时代教育评价改革总体方案》(以下简称“方案”)。《方案》指出，教育评价事关教育发展的方向，要系统推进教育评价改革，努力培养担当民族复兴大任的时代新人。十九届五中全会指出要加快推进教育现代化的新征程。随着以人工智能、大数据、互联网+为代表的现代信息技术的迅猛发展，崭新的数字世界成为人们工作、生活和学习的重要场所，生产力与生产关系的变革使得大量机械性、重复性的职业被数字世界提供的算法服务所取代，伴随着劳动力被大量解放人工智能时代对人才有了全新要求，教育也必然聚焦到以创新能力、批判意识等高阶思维培养为核心的育人过程上来。在物理世界与数字世界双重发展的基础上，

教育评价急需实践创新(齐媛,等,2020),而实践创新的根本源于评价理念创新,深入研究学评融合的新理念、新方法和新技术以及运用新模式是促进评价改革落地的关键。

在理论发展的不同阶段,评价的关注点处在由诊断到学习、由外部到内部、由知识到能力的逐步发展变化中。基于评价实施的不同目的和方法,Earl(2003, p. 26)最早区分了关于学习的评价、为了学习的评价和作为学习的评价。关于学习的评价也被叫作“总结性评价”,是指教师在教学单元结束时使用评价对学生的检查结果(O'Farrell, 2009, p. 7),其目的是向学生或父母报告学生在学校的学习进度,通常通过班级内该生相对于其他学生的排名来实现。为了学习的评价是形成性评估的代表术语,是一个以学习为导向的过程,可以促进学生接受并基于反馈采取行动,最终改善和增进对学习的理解(Bennett,

* 本文系国家自然科学基金联合基金重点项目“基于‘天河二号’超级计算机的教育系统化监控评估、智能决策仿真和应用研究”(项目编号:U1911201)的研究成果。

2011; Wiliam, 2011)。为了学习的评价将重点从总结性评价转为形成性评价,从判断到描述(Bennett, 2011),对学生进入下一阶段前尽快转变其错误理解至关重要。作为学习的评价聚焦学生的元认知能力和素养发展(Earl & Giles, 2011),鼓励学生积极参与评价过程,同时参考学生的既定学习目标,主要通过自我评估、自我监控和自我调节来缩小他们自身的学习差距并确定下一步的学习计划和个人目标。Earl (2003, p. 28)指出,在作为学习的评价中学生间几乎不存在比较,它要求教师从知识的传授者转变为学生学习的引导者,让学生体会理解自我的“认知过程”以帮助学生在学习中学会监控和及时调整。虽然Earl (2003, p. 28)认为作为学习的评价是为了学习的评价的一部分,但也有部分学者认为它应该是一种独立的教学评价方法(Gottlieb, 2016, p. 23),为了学习的评价更具公共性、交互性和教学性,突出了实践社区的本质,而作为学习的评价相对私密、个人和更具反思性,抓住了自我调节学习的特征(Allal, 2019; Lam, 2018)。

尽管为了学习的评价和作为学习的评价对学习者的积极作用得到认可,但在教学实践中迫于考试和升学的压力关于学习的评价往往更受关注(Muchlis, Ibnu, Subandi, & Marfuah, 2020)。不仅作为学习的评价由于对提高学生学业成绩的局限性而很少被实践(Knight, 2002),为了学习的评价(形成性评价)也被认为有减少的趋势,未能发挥其潜能(Gibbs & Simpson, 2004)。例如在课程中各种评价任务分配不均匀,测验往往鼓励记忆或表面的学习方法(Ramsden, 2003),无法支撑学生高阶思维能力的发展(Rust, Price, & O'Donovan, 2003)。作为学习的评价和为了学习的评价的共同特征就是反馈的使用,无论采用内部反馈还是外部反馈,二者都指向发展学生的元认知能力和达成学习标准的意识(Andrade & Brookhart, 2016, pp. 293-309)。但在实践中反馈通常是无效的,主要原因在于反馈不及时(Carless, 2006)以及反馈的数量和迭代周期有待提高(Taras,

2006)。另一个值得关注的问题是学生对两种评价的感知,特别是它是否会增加学生的额外负担。一些学生可能会将某些类型的形成性评价解释为没有分数的额外工作,或者认为它们是没有必要的(Yorke, 2003)。

教育评价改革与发展的困境催生理念的创新。评价应当是多件事情的集成体(Boud, 2000)。在人工智能时代,评价应同时涵盖诊断和学习的目的,既要评估学生成绩,也要通过评估使他们提高;它的评价标准既要有筛选价值,也要涵盖个体的增值性评价;评价的结果既要传递显性信息,也要传递隐性信息。本文基于评价改革的需求和人工智能时代的特点,从理论和实践层面提出学评融合的评价新理念,以促进学生核心素养的发展为目标,推动评价活动科学、大规模、常态化地开展。

二、学评融合的概念、内涵与核心特征

(一) 学评融合的概念与内涵

学评融合是基于数字世界的教育,统筹评价的学习性和诊断性为一体,强调以多种方式促进学生主动发展的一种评价新理念。强调通过数字世界将评价融入学生的学习活动过程中,鼓励人人展示分享成果、人人参与系列评价活动,基于评价过程与评价数据不断反思改进,促进学生自身核心素养的发展;通过数字世界记录的学习过程数据和评价过程数据,建立各种模型来计算学生认知、社会性、心理等多方面素养的发展,再以可理解的方式呈现出来。具体如图1所示。

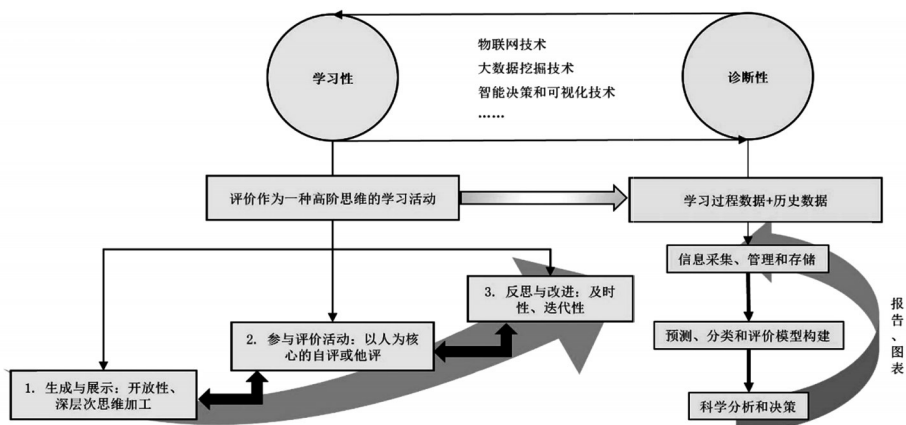


图1 学评融合的评价理念



（二）学评融合的核心特征

学评融合的评价理念不同于传统的评价理念，它有五个典型特征：一是强调评价的学习性和诊断性的融合；二是强调评价是一种高阶思维的学习活动；三是强调基于学习过程数据、利用人工智能大数据分析技术来开展过程性评价；四是强调基于数字世界开展学评融合；五是强调进步和成功的评价，保持学生与生俱来的好奇心。

1. 强调评价的学习性和诊断性的融合

先发挥评价的学习性功能，再基于学习过程数据进行数据建模评价，是学评融合的首要特征，也是区别于传统先诊断再改进的评价理念，其评价的流程更简捷，更加指向学生核心素养的发展。

以往评价实践更多地关注终结性评价，即作为诊断性的评价，其评价流程如图2所示。学生一般会经历从测验、评价、反馈到改进的过程，虽然其对一般认知能力的考察已颇有成效，但由于在物理世界中教师的课堂时间和展示空间均有限，不仅不能展示每一位学生创作的过程和结果，更难为每一位学生提供个性化的反馈和指导，学生的改进效果不尽如人意。因此，本文提出学评融合的评价新理念，其评价流程如图3所示，即依托数字世界，将评价各核心要素融合于学习活动中，以学生为主体，鼓励人人参与展示、人人参与对他人的评价，让每一位学生都能在评价活动中提高核心素养，在评价反馈中进一步改进学习。



图2 关注诊断特征的评价流程

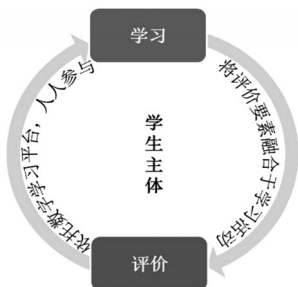


图3 学评融合的评价流程

2. 强调评价过程是一种高阶思维的学习活动

学评融合强调评价过程是一种高阶思维的学习活动，是对学生元认知、心理理论的常态化训练与实践。

学评融合强调学生在原有认知展示基础上，在阅读他人作品、评价并提出改进意见的过程中，思维得到再一次升华，思维的深度与广度得以发展。评价过程也是关注他人的创作背景、思维模式、总结他人优势与不足的过程，是对同伴深度认识的过程，是“知彼”的思维过程，这不仅是一种高阶思维的发展，也是学生社会性发展的一种方式，是直接指向学生核心素养发展的一种新的评价。学生在大量评价他人作品的过程中逐步认识到事物的多面性，再在改进自己作品的过程中认识到自己思维过程的优势和不足，通过这一“知己”的思维过程逐步树立创新意识，保持自己的特点与不同，在追求共性的基础上追求个性的发展，思维得以螺旋式上升，核心素养得以发展。最终通过评价活动达到“知己知彼”的目的，思维境界得以提升。

3. 强调基于学习过程的诊断评价

学评融合强调基于学习过程的诊断评价，即利用大数据、人工智能技术、学习分析技术对学生的各种行为进行计算、建模和可视化，具有实时性、动态性、发展性和可理解性的特点。

评价作为高阶思维学习活动会产生大量的学习过程数据，包括学生的个人基本信息、学生的作品数据、浏览数据、评价数据、修改数据以及过程中的时间点和时长等时间维度的数据。结合学生的历史数据，如作业考试成绩、学习风格、人格特质、兴趣特长等学习品质数据，通过物联网技术、大数据挖掘技术和智能决策与可视化技术，建立学生的认知发展、学业发展、社会性发展等不同方面的计算模型，构建一系列的决策预警分析模型，并基于不同地域和群体特征构建不同的学习者特征模型，可以动态地诊断、决策反馈。不仅能够描述学生的当前发展状况，展示一定时期内学生各方面的发展，还能对未来的发展进行模拟推演，从而及早采取有效措施调整教育教学。

4. 强调对数字世界的运用

学评融合必须基于数字世界开展，才能有效发挥数字世界展示、分享、交流、评价、自动计算、智能

决策和可视化的优点，才能让学评融合更加常态、更加精准和更加科学，才能有效促进深度学习和基于数据建立现代治理制度与体系。

在人工智能时代，数字世界的时空泛在、物理世界的时空拓展，人人展示、同伴评价、个人反思与改进可以常态化开展，评价内隐于学生的学习过程成为可能，学评融合的高阶思维特点可以得到常态化的训练和发展。同时，数字世界对学习过程能够自动记录、存储，结合不同的目的建立不同的计算模型，开展智能化的诊断，自动及时反馈。另外，数字世界能够大规模、常态化、个性化地进行分析，不仅能给个人提供有针对性的学习过程报告，提出有针对性的建议，助力个性化学习的开展，还能支持科学管理与决策，及早发现区域性、群体性问题，及早进行模拟推演，快速科学决策，进而大幅度、大规模提升教育教学质量。

5. 强调保持学生的好奇心

保持学生的好奇心是学评融合最关键的特征。评价是为了更多地让学生体验自身的进步和成功，保持学生与生俱来的好奇心，保持高效的学习动力，减轻学生的负担。在学评融合理念下，人人展示自己的优点和不同，人人发现他人的优点和不同，发现他人的美和成长，进而营造一种积极向上的育人氛围。

传统的评价理念受制于测试与学习分离，整个评价过程主要是教师在开展，学生则是被动地参与评价，教师更加关注成绩优秀的学生，而对其他学生的表现视而不见。同时，由于时空限制大多数学生的努力程度会被忽视，评价开展得越多，排名靠后的学生受到的负向激励越多，学生参与学习的兴趣、动机越难以保持上升。由于学生不能参与评价过程，学生思维无法得到提升，使得学生的

认知停留在原有水平，不能有效改进自己的作品，只能照搬教师眼中的优秀作品，学生不仅无法获得成就感，更会产生无力感和失败感，潜移默化学生会觉得自己不行，进而丧失学习的兴趣和动机。另外，先评价诊断再改进的评价理念还会加重师生的负担。诊断的评价是为了测而测，这就形成了多种测评系统、多方测评主体的一同参与，加重了教师和学生的负担。如果基于这个理念强化过程性评价，只会简单增加终结性评价的频率和采集点，要求教师和学生提交大量的照片、文件等过程性资料，不仅不能有效改进教学，反而会加重学生、家长、教师 and 学校的负担，难以持续，也难以发挥以评促学的作用。

三、学评融合中学的实施阶段与核心要素

学评融合在实践层面和操作层面依赖数字世界的运用，重点强调评价的学习功能和诊断功能，核心前提是评价的学习性的运用，因此本文聚焦评价的学习性的阶段和核心要素。具体而言，评价作为一种高级思维的学习活动，有生成与展示、参与评价活动、反思与改进三个核心阶段和多个核心要素，如图4所示。

(一) 生成与展示及其核心要素

学评融合理念落地的基础是强调学生思维的生成与作品的展示，强调教师要利用数字世界重构育人环境，优化学与教的方式，强调学生从消费为主到生成为主，形成为生成而消费的新型学与教方式。在学

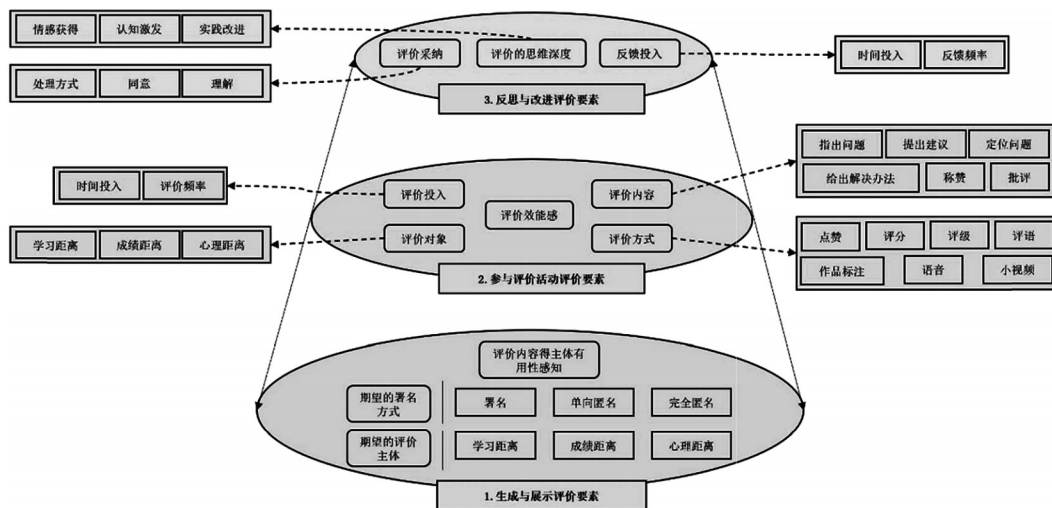


图4 学评融合中评价作为高阶思维学习活动的要素框架



生生成的基础上强调人人展示,这可以是在物理世界展示,当然,更强调在数字世界的常态化展示与分享。即便作业是选择题一类客观性、标准化的任务,也需要教师开展基于答题板的任务设计,学生基于选择分布来进行深度思维。例如,我能否说服他人同意我的选择是对的?我能否判断他人思维的局限导致的错误选择?我要修改我的答案吗?学生基于答题数据开展深度的讨论和学习,选项分布如图5所示。



图5 答题版的选项分布可视化

生成与展示是学生参与评价活动的前提条件,学生在评价过程中关注的核心要素是学生签署方式、评价主体和评价内容有用性的心理期望与感知。

1. 期望的评价署名方式

同传统的面授课堂相比,学评融合理念下的评价活动内置于数字化学习空间,更关注对他人作品的理解与评价。在交互环境中,评价的署名方式对学生生成与展示的行为和心理都存在影响(张倩,2007)。随着同伴互评逐渐成为一种有效的评价方式(李菲茗,2018),圈子效应和权威效应也逐渐凸显,即给自己的好友较高分数或给平时表现好的同伴较高分数。这种效应与评价的署名方式密切相关(舒存叶,等,2015),学生期望的署名方式从某种角度而言反映了依赖该效应的倾向,进而影响生成与展示部分的投入程度。对同伴互评的形式进行分类,从评价者、被评价者角度可以分为署名、单向匿名和完全匿名。以往研究发现,完全匿名和单向匿名条件下的评价更具有批判性,且不存在显著差异,但是评价质量不如署名条件下的评价(Yong Zhao, 1998; 王佑镁,2006)。因此,学评融合理念下的评价要关注学生期望的评价署名方式,体现个性化。

2. 期望的评价主体

学评融合强调学教并重的教学理念,评价活动也更加注重主体的多元化。评价主体在同伴互评中占有

重要地位,而学生自身期望的评价主体也将潜在影响学生的生成投入和评价内容的有用性感知等。该部分评价主体的分类将同下文中对评价对象的分类,不同学生期望的评价主体具有差异性,学评融合理念下的评价应该在捕捉这种个性化差异的基础上设计精准的评价主体推荐功能。

3. 评价内容的有用性感知

不同的评价内容具有不同的作用,如激励性评价有助于激发学生的积极性和创造性等,但研究也显示,不同环境下同一评价内容的作用也不是一成不变的。例如,随着年级的增长男生会更倾向于选择“客观”的评价,而女生则相反,有从“客观”转向“称赞”的趋势。因此,富有不同个性特征的学生对不同评价内容的感知是学评融合理念下的评价需要关注的,在此基础上考虑教师、同伴、家长等不同多元主体的参与,将评价内容和评价主体衔接,确定不同学生对评价内容的主体有用性感知,推动评价的落地实施。

(二) 参与评价活动及其核心要素

学评融合理念落地的关键是以人为核心开展自评或他评的学习活动,鼓励跨学科的教研方式,核心就是评价是以人为出发点。自评或他评都是学生主动参与评价的象征,也是从他人角度制定战略,将开放性的生成转变为客观的价值判断。其中,面对同一开放性的生成任务,衡量每个人的要求和标准也应不同,故以人为核心的增值评价便自然得到关注。无论是自评还是他评都要求学生主动参与,评价的方式也多种多样,涵盖了点赞、打分、评级、评语等,在学生主动参与评价的过程中多种思考问题的角度和方式都会对其形成冲击,这就促使学生不再停留在原有简单认知上,而是会萌发出更深层次的理解、困惑、迷茫。同时,学生的主动参与也会让其更加了解评价标准,有利于师生关系和同伴关系的发展。

参与评价活动作为学评融合评价的核心,各要素的合理、共生发展是重中之重。从功能上可以将参与评价活动的核心要素划分为基本要素和关键要素两部分。评价对象、评价方式和评价内容是表征评价活动参与情况的基本要素,评价投入和评价效能感是表征参与评价活动质量的关键要素。

1. 评价对象

评价对象分为评价者和被评价者,从关系角度看包括学习距离、心理距离和成绩距离三种类型。学习

距离是指物理世界中的距离,可以分为同桌、同一学习小组的人、学习小组之外的人和其他同学等;心理距离是指亲密关系所带来的距离,可以分为好朋友、普通同学和我不喜欢的同学等;成绩距离是由平时的学习表现所塑造的距离,分为平时成绩好的人、平时成绩一般的人和平时成绩不太好的人。近期关于中小学生的评价活动现状调查显示,在学习距离中评价其他同学的比例最高,高达37%;在心理距离中评价好朋友的比例高达80%;在成绩距离中评价平时成绩好的比例高达58%。这说明现阶段的评价还未合理、公平地考虑三个距离的评价对象安排。同时,也加重了同伴评价中的圈子效应和权威效应的发生风险。

2. 评价方式

数字世界让学评融合理念下多样化的评价方式成为可能。在以往的学生互评中,以评分反馈和评语反馈为主,而学评融合理念下的评价方式包括点赞、评分、评级、评语、价值判断、答题板的运用等,评价的媒体包括文本、语音、小视频和作品标注等。其中,不同的评价方式表征了学生参与评价时的不同思维层级。了解学生的评价方式喜好并结合不同评价方式在不同教学阶段的适用性进行合理设计将有助于提升评价活动效果。

3. 评价内容

评语内容可以从评价者和被评价者的角度分为评语焦点和评语形式两个层面,评语焦点分为关注语言类和关注内容类,评语形式分为认知特征类和情感特征类(Patchan & Schunn, 2015; Gao, Schunn, & Yu, 2019)。本文讨论广义下的评价,只关注评语形式。认知特征类是一些可采纳评语,如“指出问题”“提出建议”“定位问题”“给出解决办法”。情感特征类主要指“称赞”“批评”两类评语。“指出问题”是明确、直接指出作品中的问题或错误的评语。Matsumura等(Matsumura, Patthey-Chavez, Valdes, & Garnier, 2002)发现指出写作文本整体性问题更有利于采纳和修改。“提出建议”是针对项目问题或错误给出的修改意见,其特点是笼统、不具体(Nelson & Schunn, 2009),如“请你再好好想一想”。“定位问题”是明确指出问题的具体位置,如“我认为这句话跟前面有点重复”,有利于被评价者理解问题进而采纳评语(高瑛,等,2019)。“给出解决办法”是提供如何修改或提高作品质量的评语,收到该评语的

被评价者极有可能提高采纳率并改进作品质量。总的来说,相较于情感特征类,认知特征类对学生后续评语采纳和修改影响最大(Patchan, et al., 2015)。同时,不同类型的评语使学生产生的情绪体验也有显著差异(柏宏权,等,2019)。这就要求在设计学评融合理念的评价活动时合理利用不同类型的评语内容。

4. 评价投入

评价投入是评价他人作品时的时间投入和评价频率。两者都表征了学生自评或评他活动的参与度,是学评融合理念下评价活动量的表现。由于学评融合理念更注重学生的评价生成,因此在评价投入中将主要关注评价频率。Chen和Tsai分析了同伴评价的频率对作品质量的促进作用。结果显示,在前两轮的学习中评价频率会影响学业成绩,但在第二轮到第三轮中的影响不明显。李艳燕等(2020)的研究发现,小组成员发帖越多,成员之间的社交网络越联通,越对小组成绩有重要影响。因此,较好的设计评价的频次是学评融合理念下的评价需要关注的要素。

5. 评价效能感

评价效能感是指学生基于对自身评价能力的主观评价而形成的一种自信感。是在评价活动中,学生相信自己给出的评语经过了仔细思考、对他人很有用、有充分的依据、符合一定标准等。目前,关于评价效能感的研究较少,但与其相近的自我效能感和评价行为间的研究表明,自我效能感可能是影响学生评价行为的一个重要因素(Hsia, Huang, & Hwang, 2016)。Wang和Shu-Ling(Wang & Wu, 2008)等基于社会认知理论的视角,在探索同伴评价以及自我效能感对网络学习的影响时发现,自我效能感作为影响自我调节学习的关键变量之一,与学生的评价行为、学习策略的使用均存在显著的相关性。

(三) 反思与改进及其核心要素

学评融合质量的提升要求及时性和迭代性的反思与改进。不管是从评价还是从支持自我调节学习方面看,大多数研究的重点还很少聚焦反思阶段(奥尔加·维伯格,等,2020),而学生利用评价行为表现信息在反思阶段做出改进也至关重要。在参与评价活动后,学生将发现很多相似或完全不同的作品,进而明白有很多思考问题的角度和方式,由此拓展了思维的广度,在此基础上的反思和改进将是又一次认知升



华,它需要学生有元认知的加工。在该阶段学生虽然知道会有其他的想法和观点,但并不会完全参照、变成它们,而是会潜移默化地吸收某些优点,进而将自己的作品发挥出特色。也就是说,评价行为让改进更容易。

反思与改进的核心要素可以从表征学生态度的评价采纳要素、表征学生行为的反馈投入要素和表征学习过程的评价思维深度要素三个方面进行归纳。

1. 评价采纳

评价采纳是判断学生间互动有效性的一个重要环节(高瑛,等,2019)。学评融合理念的评价强调以人为中心和评价的增值性,因此,被评价者的评价采纳行为是说明该评价高质量进行的重要标志。同时,在评价和采纳之间很有可能存在某种潜在中介对反思与改进行为产生影响,从认知、情感和行为三个角度出发,与评语采纳关系密切的分别是“理解”“同意”“处理方式”,但研究结果不一(Patchan, Schunn, & Correnti, 2016)。“理解”是指一种认知上了解或解释意义及因由的能力。“同意”是指情感上对评价信息或评价能提高改进质量的认同。“处理方式”是行为上对待评价信息的及时性。因此,学生“理解”“同意”“处理方式”的程度和及时性应在设计学评融合理念的评价活动时被考虑。

2. 反馈投入

反馈投入是学生在得到他人评价后,进行反馈的时间投入和频率。两者都表征了学生在参与评价活动后进行反思与改进的广度,是学评融合理念下评价活动质量的保障,同时对促进评价活动三个部分的螺旋上升具有重要作用。由于学评融合理念更注重学生的反馈生成,因此在反馈投入中将主要关注学生反馈的频率。

3. 评价的思维深度

学评融合理念的评价关注学生高阶思维能力的培养,而培养效果的表征与评价的思维深度密切相关,即学生在评价活动中的情感获得、认知激发和实践改进程度。情感获得即学生在多大程度上觉得评价他人时很开心、很有成就感等。认知激发即学生在多大程度上认为评价拓展了自身视野、激发了灵感等。实践改进是学生对评价能够让自己更好地完成作品、看到不同解决办法、更加了解评分标准的认可程度。高阶的评价思维深度是学评融合理念的发展目标,因

此在设计评价活动时要充分考虑学生评价思维深度的发展状况。

四、小结

本文面向人工智能时代下的教育评价变革,提出了新的学评融合理念,并详细阐述了学评融合理念的概念、内涵与五大特征,提出了学评融合中评价的学习活动落地的实施框架和核心要素。当然,学评融合的落地还需要更多的探索研究与实践。

一是学评融合理念强调自下而上的评价改革,是针对一线实践层面提升质量而设计的一种新型评价体系。当前我国的教育评价体系还处在不断完善的阶段,在国家、省级层面开展大规模的以诊断为核心的终结性评价有其必要性,但不能停留在知识技能掌握的导向上,要强调和关注对学生核心素养的评价,快速获取学生发展的状况、特点,并挖掘背后的成长规律。

二是学评融合理念的落地有利于实践一线提升教育水平,提高教师的育人能力。建议更多研究者能够聚焦评价的学习功能的研究,避免过度强调评价的诊断性,导致学生、家长、教师、学校的评价负担不断加重,但评价改进却原地踏步。研究人员可以更多地利用PISA数据、我国对外发布的质量监测数据以及已有的研究数据,加强基于评价数据改进的能力,提出可切实落地的实践方法,助力我国基础教育质量的提升,助力学校育人方式的转变。

三是期待更多的评价第三方聚焦利用数字世界科学诊断、科学提升学生的核心素养,强化评价的学习性,建立有利于学生、教师发展的评价环境,尤其在加强过程性评价、探索增值性评价的背景下,科学推进、科学实践,不做应试教育的助推器,而是探索开展有利于学生发展的评价活动,基于学习活动数据,利用人工智能、大数据等优势研发科学的计算模型和个性化学习模型,开展大规模的质量提升与科学诊断。

虽然学评融合理念能够提升学生的高阶思维能力,解决以评促学低质、低效的问题,但在当前教学实践中仍可能存在诸多挑战:一是对学评融合理念的理解不够,不能构建基于学评融合理念的育人生态;二是数字世界的建设不够,很多地方由于多种原因还

不具备常态化、大规模开展的条件,不少地区目前还处在物理世界育人的阶段,欠缺对数字世界育人的规律和方法的认识和运用,不具备基于数字世界常态化育人的条件和能力,学评融合的落地还有待于更多、大规模的研究和实践。

[参考文献]

- 奥尔加·维伯格,默罕默德·哈利勒,马丁尼·巴斯,肖俊洪. 2020. 在线学习环境下的自我调节学习和学习分析实证研究述评[J]. 中国远程教育(12):28-41, 58, 93.
- 柏宏权,李婷. 2019. 同伴互评中评语类型对情绪体验的影响研究[J]. 电化教育研究,40(04):92-98, 111.
- 高瑛,汪溢,Christian D. Schunn. 2019. 英语写作同伴反馈评语采纳及其影响因素研究[J]. 外语电化教学(02):17-24.
- 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(05):63-67, 78.
- 李艳燕,彭禹,康佳,等. 2020. 在线协作学习中小组学习投入的分析模型构建及应用[J]. 中国远程教育(02):40-48, 77.
- 齐媛,张生. 2020. 学评融合:落实评价改革的重要路径[N]. 中国教育报,10-31(03).
- 舒存叶,张海萍. 2015. 网络环境下同伴评价影响因素的实证研究[J]. 电化教育研究,36(02):89-93.
- 王佑镁. 2006. 电子学档中同伴评价的实验研究[J]. 电化教育研究(02):62-65.
- 张倩. 2007. 网络环境下同伴匿名评价实证研究[J]. 电化教育研究(07):55-57, 71.
- 中共中央、国务院. 2020-10-13. 中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL]. [2020-11-10]. http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202010/t20201013_494381.html
- Andrade, H., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation*. Berlin, Germany: Springer, Cham
- Allal, L. (2019). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 1-18.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Education*, 22(2), 151-167.
- Carless, D. (2006). Differing perceptions in the feedback process, *Studies in Higher Education*, 31(2), 219-233.
- Earl, K., & Giles, D. (2011). An-other look at assessment: Assessment in learning. *New Zealand Journal of Teachers' Work*, 8(1), 11-20.
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin.
- Gao, Y., Schunn, C. D. D., & Yu, Q. (2019). The alignment of written peer feedback with draft problems and its impact on revision in peer assessment. *Assessment and Evaluation in Higher Education*, 44(2), 294-308.
- Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning, *Learning and Teaching in Higher Education*, 1, 3-31.
- Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity: Connecting academic language proficiency to student achievement* (2nd ed.). Thousand Oaks, CA: Corwin.
- Hsia, L., Huang, I., & Hwang, G. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology*, 47(4), 618-632.
- Knight, P. (2002). Summative assessment in higher education: practices in disarray, *Studies in Higher Education*, 27(3), 275-286.
- Lam, R. (2018). Understanding assessment as learning in writing classrooms: The case of portfolio assessment. *Iranian Journal of Language Teaching Research*, 6(3), 19-36.
- Matsumura, L. C., Patthey-Chavez, G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *The Elementary School Journal*, 103(1), 3-25.
- Muchlis, Ibnu, S., Subandi, & Marfuah, S. (2020). Students' Result of Learning at Chemistry Department through Assessment of, for, and as Learning Implementation. *International Journal of Instruction*, 13(2), 165-178.
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375-401.
- O' Farrell, C. (2009). *Enhancing student learning through assessment*. Dublin, UK: Institute of Technology.
- Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*, 43(5), 591-614.
- Patchan, Melissa M, Schunn, Christian D, & Correnti, Richard J. (2016). The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions. *Journal of Educational Psychology*, 108(8), 1098-1120.
- Ramsden, P. (2003). *Learning to teaching in higher education* (2nd ed.). London, UK: Routledge.
- Rust, C., Price, M. & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes, *Assessment and Evaluation in Higher Education*, 28(2), 147-164.
- Taras, M. (2006). Do unto others or not: equity in feedback for under-

(下转第16页)

台,体现了我国政府立足新的时代背景,对全面发展的人所应具备的创新精神与实践能力和综合素质的重大关切。在推动信息技术与教育深度融合的历史新时期,深入探索智能评价机制、理论体系、建模方法与关键技术,将学习视为复杂的现象,探究不同认知因素、社会情感因素、背景变量等因素影响学习的规律,助力真实教育需求的满足,既是难得的历史机遇,也是教育技术人理应具备的使命担当。

[参考文献]

- 陈丽,逯行,郑勤华. 2019. “互联网+教育”的知识观:知识回归与知识进化[J]. 中国远程教育(07):10-18,92.
- 冯成火. 2020. 高考科目改革的轨迹与推进策略——兼论新一轮高考改革的深化与完善[J]. 中国高教研究(05):78-82,102.
- 国务院. 2020-10-13. 中共中央国务院印发《深化新时代教育评价改革总体方案》[EB/OL]. [2020-12-10]. http://www.gov.cn/zhengce/2020-10/13/content_5551032.htm
- 国务院. 2017-07-20. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. [2020-12-10]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
- 黄荣怀,刘德建,刘晓琳,徐晶晶. 2017. 互联网促进教育变革的基本格局[J]. 中国电化教育(01):7-16.
- 中国教育和科研计算机网. 2020-12-11. 雷朝滋:智能技术支撑教学改革与教育创新[EB/OL]. [2020-12-20]. http://www.edu.cn/xxh/focus/li_lun_yj/202012/t20201211_2055540.shtml
- 田爱丽. 2020. 综合素质评价:智能化时代学习评价的变革与实施[J]. 中国电化教育(01):109-113,121.
- 新华网. 2020-09-22. 习近平:在教育文化卫生体育领域专家代表座谈会上的讲话[EB/OL]. [2020-11-22]. http://www.xinhuanet.com/politics/leaders/2020-09/22/c_1126527570.htm

(上接第8页)

- graduates, *Assessment and Evaluation in Higher Education*, 31(3), 365-377.
- Wang, Shu-Ling, & Wu, Pei-Yi. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers and Education*, 51(4), 1589-1598.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37,3-14.
- Yong Zhao. (1998). The Effects of Anonymity on Computer-Mediated Peer Review[J]. *International Journal of Educational Telecommunications*, 4(4), 311-345.
- Yorke, M. (2003). Formative assessment in higher education: moves towards theory and enhancement of pedagogic practice, *Higher Education*, 45, 477-501.

收稿日期:2020-11-22

- 张琪,李福华,孙基男. 2020. 多模态学习分析:走向计算教育时代的学习分析学[J]. 中国电化教育(09):7-14,39.
- 张琪,王红梅. 2019. 学习投入的多模态数据表征:支撑理论、研究框架与关键技术[J]. 电化教育研究,40(12):21-28.
- 张琪,王红梅,庄鲁,赖松. 2019. 学习分析视角下的个性化预测研究[J]. 中国远程教育(04):38-45,92-93.
- 张琪,武法提. 2018. 学习仪表盘个性化设计研究[J]. 电化教育研究(02):39-44,52.
- Albert, R. S., & Runco, M. A. (1999). The history of creativity research. In R. Sternberg (Ed.), *Handbook of creativity* (pp.16-31). New York: Cambridge University Press.
- Wu, F., & Lai, S. (2019). Linking prediction with personality traits: a learning analytics approach. *Distance Education*, 40(1),1-20.
- Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits*. New York: Cambridge University Press.
- Lai, S., Sun, B., Wu, F., & Xiao, R. (2020). Automatic Personality Identification Using Students' Online Learning Behavior. *IEEE Transactions on Learning Technologies*, 13(1), 26-37.

收稿日期:2020-12-22

定稿日期:2021-01-04

作者简介:张琪,博士,教授,硕士生导师;王丹,硕士,讲师。淮北师范大学教育学院(235000)。

责任编辑 刘 莉

定稿日期:2020-12-23

作者简介:张生,博士,副教授,博士生导师;王雪,硕士研究生。北京师范大学中国基础教育质量监测协同创新中心(100875)。

齐媛,博士,助理研究员,中国教育科学研究院(100088)。

责任编辑 刘 莉

Abstracts

AI-enabled educational assessment: an integrated approach to assessment of, for and as learning

Sheng Zhang, Xue Wang and Yuan Qi

In the era of Artificial Intelligence (AI), a digital world is added to the physical world, which has brought about radical changes in the way people live, work and learn, thus resulting in transformation of relations of production and superstructure. In light of this transformation, the focus of education will have to shift from teaching knowledge and skills to fostering core literacies, which will become the primary purpose of assessment. This article sets out to analyze the dilemma of diagnosis-centered assessment and proposes an integrated approach to learning and assessment. It makes a theoretical case for the approach in the AI era by illustrating its concept, connotation and defining features. It also develops an assessment framework which is comprised of three key stages of higher order thinking activities, i.e., generation and demonstration, engagement in the assessment activities, reflection and improvement. Key elements in each stage are identified to facilitate assessment of, for and as learning.

Keywords: assessment of, for and as learning; assessment concept; core literacy; Artificial Intelligence (AI); assessment reform; assessment skill; educational assessment

Educational assessment in the intelligent era: implication, role and pathway

Qi Zhang and Dan Wang

Educational assessment impacts the overall quality enhancement and supply-side reform of education. It is argued that educational assessment should evolve accordingly by taking advantage of big data and intelligent technology to address problems of existing assessment practice. This article focuses on the “whole person” assessment with learning as the main target, and expounds on how education assessment should adapt to the changing learning modes and the multi-dimension learning outcomes. It explores the pathway of “knowing students”, “implementing teaching” and “developing overall quality”, drawing upon Confucius philosophy of education. It then goes on to discuss the role of intelligent technology-enabled assessment in terms of learning engagement assessment, identification of students’ personality and characteristics, innovation assessment, and expression ability assessment. The article concludes by arguing that the reform of educational assessment should be in line with the nature of education and make appropriate use of technology to deliver the goals of assessment.

Keywords: intelligent technology; educational assessment; individualized teaching; knowing students; scene calculating; model-building

Online teaching and innovation of Chinese elementary and middle schools during COVID-19: an interactive teaching perspective

Yijun Wu and Li Chen

Face to face teaching is not possible when schools are closed during COVID-19. In response to the initiative of “Disrupted Classes, Undisrupted Learning”, online teaching becomes the most important option to guarantee learning continuity in elementary and middle schools, resulting in an unprecedented mass online teaching practice. This study reports on a qualitative analysis from an interactive teaching perspective of 100 cases selected from the BNU collection of excellent online teaching classes in prima-