

【质量与评价】

人工智能推动教育测评范式 变革的机遇与挑战

杨丽萍 辛涛

【摘要】人工智能与信息技术的发展催生个性化学习的发生,疫情带来的影响加速推动世界范围内教育数字化产业的转型与升级,线上与线下学习相结合的大规模混合式教育形态倒逼教育测评范式从“以评促学”向“学评融合”演进。在这场变革中,以往教育测评的参与者、测评形式、数据来源、测量模型与技术方法都在发生深刻的变化,教育测评研究者与实践者在核心素养测评、基于过程数据的测评和个性化学习等重要方面的积极探索随之迎来新的契机。新一代教育测评的核心问题是,如何将人工智能与教育测评纳入统一框架以提升整个教育系统的效率。与此同时,教育测评的实质意义、效度、伦理和公平仍需坚守,以应对新时代教育发展和社会需求的挑战。

【关键词】教育测评;人工智能;新型测评;教育测评范式;效度;可解释性

【作者简介】杨丽萍(1981-),女,北京师范大学中国基础教育质量监测协同创新中心博士后;辛涛(通信作者)(1968-),男,北京师范大学中国基础教育质量监测协同创新中心教授(北京 100875)。

【原文出处】《中国考试》(京),2022.10.13~21

【基金项目】2020年度国家自然科学基金联合基金重点支持项目“基于‘天河二号’超级计算机的教育系统化监控评估、智能决策仿真与应用研究”(U1911201)。

引言

教育的本质是服务,教育形态随着社会发展而演变,社会生产力发展水平决定教育应该承担的社会功能。随着以信息技术为代表的第三次工业革命的深入,社会岗位的设置和定义都在迅速发生变化。21世纪以来,很多常见的岗位正在迅速消失或衰减,对岗位适配人员的要求也不再整齐划一。由于社会人才需求急剧变化,以班级制授课为特征的大规模系统化培养技能型人才的教育模式逐渐无法适应社会发展节奏。审视全新的社会政治、经济与文化环境,以学习者为中心理念的个性化教育在促进人的全面发展方面的作用不断凸显。未来教育的形态和特点是当前国内外教育舆论的热点。不论是联合国教科文组织发布的《教育2030行动框架》,还是我国的教育发展规划和政策文件,如《中国教育现代化2035》等,无一例外地强调信息化对教育的再造和推动。在政策倡导下,各地积极加强教育信息化建设。近年来,尽管有人工智能等先

进技术辅助教育实践的成功范例,但由于没有紧迫性需求,因此推动各级各类教育信息化产业进一步升级的动力不足。

2019年底至今,新冠肺炎疫情给全世界各国教育带来巨大的冲击。由于传统线下课堂无法开展教学,世界各国纷纷推出教育信息化建设升级措施以增强教育系统的弹性^[1]。在信息技术与疫情的双重推动下,虚实结合的教育生态环境正在快速形成,课程形态、教学组织、学习方式、教育测评和反馈的形式都发生了一系列变化。后疫情时代,线上与线下并存的数字化教育生态产生大规模、系统性的数据和信息资源,如何从繁杂的数据中挖掘和寻找教育规律,从而服务于学生成长和社会发展是教育新生态下亟待解决的诉求。

教育评价在教育改革与发展中处于核心战略角色。2020年10月,《深化新时代教育评价改革总体方案》发布,强调教育评价与信息技术相结合,通过科技赋能深化教育评价改革,以更好发挥教育评

价的导向、诊断与激励作用。虽然国家有系统化的教育改革政策设计,现实有紧迫的学习与评价需求,但在信息技术环境下,利用人工智能技术辅助教育测评仍缺少可操作的途径。在数字化教育生态转型中,如何调整教育测评范式以适应教育新生态,同时对学生学习和发展给予有效帮助,从而培养出国家与社会发展需要的人才,是教育测评理论研究者与实践者要共同回答的话题。

一、教育测评范式演进的机遇

教育测评范式的变化不但受到社会、政治、经济与文化,教育教学规律,以及国家政策导向的影响,同时也越来越得益于技术力量的创新与推动。虽然人类科技的发展对教育的贡献较其他领域相对有限,但仍能发现教育测评范式缓慢却愈加清晰的演进路线。20世纪五六十年代大规模推进“测评学习”(assessment of learning, AoL),测评是对学习和教学外在的、刚性的评价。由于这种范式忽视了学生自身发展,因而受到批判和质疑。21世纪以来,无论课堂评价、区域评价还是国家层面的评价,各国都大力倡导“以评促学”(assessment for learning, AfL),即测评要真正有助于教学改进,进而促进教育质量的提升^[2]。疫情的出现加速了教育测评范式的转向,随着线上和线下教学与学习的混合式教育环境快速形成,“学评融合”(assessment as learning, AaL)^[3]比预期更早到来。“学评融合”是未来个性化学习的有机组成部分,在面对以学习者为中心的教育生态转型时,尤其需要实时的学习反馈和调整,以保证学习效果。在“学习”与“评价”融合的过程中,测评的目的、测评的参与者、测评的形式、测评的数据、测评的模型和应用场景都因信息技术与疫情的双重影响而快速变化。

(一) 测评目标从显性走向隐性

面对未来社会岗位的多样性与不确定性给教育带来的巨大挑战,首先要回答的是教育应该培养什么样的人?党的教育方针总目标是培养德智体美劳全面发展的社会主义建设者和接班人。2014年《教育部关于全面深化课程改革,落实立德树人根本任务的意见》正式提出核心素养概念并明确其内涵,即学生在基础知识之上要形成能够适应社会发展和个体发展需求的基本素质、价值观、关键能力和必备品格^[4]。核心素养的提出具有重要的政策导向,从2017年发布的普通高中课程方案和课

程标准,到2022年发布的义务教育课程方案和课程标准,进一步强调如何通过设置不同学科的内容来实现学生核心素养的养成。从课程政策的制定者、教育教学管理人员到教学一线教师,都在探索如何培养学生的核心素养。核心素养既是国家政策导向的要求,也是未来社会学生学习和测评的目标。

当教育测评目标从显性走向隐性,由“知识立意”走向“能力立意”和“素养立意”时,测评的理论价值层面都发生了非常大的变化。早期的测评目标大多是基于布鲁姆框架考查学生对知识的记忆、对规则的理解、方法的应用等,主要指向学生学习的学科知识。在当前快速转型的数字化生活、学习与工作环境中,对人才的需求不再强调简单的知识理解、记忆和特定技能,测评目标指向的也不再是学科知识,而是学生在获得基础知识水平之上形成的跨学科的关键能力和基本素养。因此,站在教育管理者 and 一线教育工作者的角度来看,更需要了解学生必备能力和核心素养的发展状况,进而改进教育教学。随之而来的困难是,对核心素养中深层与隐性指标的测量无疑比知识成就测验难度更高,如何有效测评这些指标,是教育测评专业人员面临的挑战。

(二) 测评主导者从教师走向教师和学生

在“学评融合”范式下,测评的参与者从教育管理者或教师主导走向以学生为中心,由“外部使用”走向“改进学习”和“自我监测”,从而帮助学生成为终身学习者。早期的教育测评一般是为满足外部较为刚性的要求做测评设计与测评方案,例如,无论是学生学习还是教师教学都被简单地按成绩排队,在测评的过程中,教育管理者和教师是测评的主导者。近20年来,在特别强调测评对学生学习与教师教学改进的“以评促学”范式下,学生、教师、同伴、家长等多种角色都有参与测评的机会,学生在测评中被赋予更多主动性,通过测评的反馈来改进学习。在“学评融合”的典型在线学习场景下,学习者的学习与测评依托信息化平台形成了一个相互推进的系统性过程,对学生线上学习状态的及时测评成为推动学生学习的有效手段。

(三) 测评形式从纸笔走向数字智能化

在测评的形式上,由标准化的纸笔测评走向数字技术支持的交互式、虚拟式等基于人工智能技术的测评。无论是“以评促学”还是“学评融合”范式,

都对传统纸笔测评提出了更高要求。虽然纸笔测试具有非常成熟的理论框架和操作范式,但是要实现《深化新时代教育评价改革总体方案》提出的要求,使综合评价落地,就需要超越传统的纸笔测评,在新技术和新方法的支撑下使过程性评价真正发生。目前,以自动化、交互式、AR或VR等虚拟化为特征的增强现实测评不断出现,为测评学生素养和能力提供了超越传统测验的手段和途径。例如,基于交互式测评的复杂任务解决测评、基于游戏的创造力测评、基于脑认知的测评等,其本质都是构建从认知或行为到能力水平的映射,是有待深入研究和应用的一系列新型测评形式。在教育数字化转型时代,教育测评专业人员担负着重要的角色,教育测评正处在机遇与困难并存的风口,既有政策倡导和跨学术交流的良好环境支持,也面临当下理论与实践研究不足的窘迫,在“学评融合”范式下,教育测评工作者需要积极探索以各种人工智能技术为工具开发新型教育测评的可能性。

(四) 测评数据来源从单一走向多样

在数字化教学、生活与工作场景中,海量的个体在虚拟或网络环境中的痕迹与行为数据留存在数字化空间中,测评能够触及、获取与分析的数据形态越来越异构化和高维化。以往的教育测评大都使用纸笔测验、自评或他评的心理量表采集数据,而在信息丰富的教育环境下,测评数据的来源除纸笔测验数据外,还有基于人工智能测评的多模态测评数据、课堂录像与教学数据、社交媒体数据、伴随式设备收集的数据等。如何在多源杂糅的数据信息里发现规律,利用多模态数据克服传统测验或自陈式报告的固有局限,准确捕捉与测量心理特质是核心问题。这些新场景数据的采集与处理既需要从大数据计算或人工智能的角度来突破技术难点,更需要从教育测评的理论角度对新技术提供支撑、解释与牵引。

(五) 测评模型从经典走向跨领域

在心理和教育测量领域内,经典测量理论仍然在实践中占有主导地位,其基本概念(如难度、区分度、信度与效度等)的应用仍然非常广泛。从经典测量理论到项目反应理论再到认知诊断,下一代教育测评的理论基础是什么还需要领域内的研究者与实践者共同探索。更进一步地,当超越心理和教育测量学术群体时,随着信息技术的更新迭代,当基于人工智能与大数据方法的分类模式越来越精

确时,完全可以脱离心理与教育测量理论和测量模型,用人工智能或大数据实现对个体特征的精准分类。在信息技术环境下,教育测量模型,例如认知诊断模型需要与机器学习、脑科学等技术方法相融合,从而拓展测评模型在新教育场景下的适用性。因此,对于心理与教育测量专业人员来说,已有的测评理论与测评模型如何与人工智能或大数据方法对话与接轨是必须面对且亟待解决的难题。

二、人工智能与教育测评的融合

当审视教育测评和人工智能两种截然不同的研究范式时,能够发现人工智能是从多模态数据中抽取重要或典型特征作出最终判断的过程,是数据驱动“自下而上”的典型计算机科学研究,在实现精准分类的同时存在难以解释的因果关系和理论机制的“黑箱”问题。与之相对,教育测评通常基于理论概念和框架收集证据,是“自上而下”的研究。教育测评对学生学习或教学效果的判断是基于理论框架的解释,对于概念与分数的赋值有明确含义。在教育信息化环境下,两种学科在核心素养的测评、基于多模态数据的测评与自适应学习等教育测评的重要应用场景中碰撞与合作,成为实现“学评融合”的有效路径。

(一) 基于人工智能的核心素养测评

虽然我国政策明确提出了以核心素养为导向的教育改革,但是在一线教育教学实践中,关键问题是教育测评专业人员如何帮助政策制定者和实践者提供评价学生核心素养的可操作性方案。长久以来,与知识应用与固化技能的教育测评相比,对更加隐性且具跨学科特性的核心素养测评的研究与实践都相对滞后。目前关于人格、能力、成就、态度等方面的测量已有很多著名的测验,但对于核心素养中类似创造力、复杂问题解决等高阶能力或个体心理品质的评价在传统教育测评范式下一直难以解决。随着心理学领域对学习规律和影响因素研究的不断深入,在“学评融合”范式下,利用人工智能与大数据技术为学习者的元认知策略、问题解决能力、协作能力等素养的发展提供更多反馈和指导,是跨学科研究需要形成合力来重点突破的难题。

以创造力测评为例,在数字化的游戏任务环境中,通过收集与分析学生在问题解决过程中的表现和结果等证据来测量学生的创造力。测评人员开发基于游戏的创造力测评分为三个步骤:首先,需

要界定创造力的概念与内涵,构造透明且稳定的“尺子”作为测评标准;其次,模拟真实任务构建游戏情境,将创造力测评目标在数字化的环境下进行表现性的要素解构,将游戏任务解构为一系列指标,从而使抽象的测评概念和标准与具体可操作的任务表现相对应;最后,通过引导学生在游戏中思考问题,分析推断,实施创新的操作方案,获取数字化环境中学生的认知与行为信息与证据。在测评的开发与实施过程中:心理学家参与创造力概念与标准的定义;教育测量学家提出以证据为中心的设计,为游戏化测评提供理论框架,通过定义学生模型、任务模型,使游戏情境与心理测量理论融合;计算机技术人员辅助开发游戏化测评,在创造性问题解决的测量理论框架下开发任务模型,获取、存储与分析学生在任务解决过程中留存的认知与行为数据信息,结合心理测量模型估计学生的创造力水平。概要之,对于测评学生未来发展必备的高阶能力,任何有效的新型测评技术都需要心理学家、教育测量学家、计算机领域工作者的共同参与和努力推动。

在心理品质测评方面,以疫情期间学生心理健康问题测评为例,基于网络环境下学生的学习与活动的行为表现来预警学生的心理状态。具体来说:首先,通过采集文本、图片、音频、视频,以及学生在社交媒体上的网络行为痕迹数据,对学生的认知与行为进行客观刻画与分析;其次,基于心理学家、咨询专家的经验,对学生在网络环境中的产出进行分析与标注,从理论与经验的角度归纳心理品质与学生行为表现之间的潜在关系;最后,基于多元数据,以人工标注为预测目标建构心理品质的分类模型,并对模型进行检验与迭代。在此类测评的构建过程中,将人工智能和大数据分类方法与心理学家、心理咨询工作者的研究或经验相结合,既能规避传统的心理咨询或健康问卷自陈式报告的局限,也不会给学生带来额外的负担和压力。

(二)多模态数据与教育测评的融合

随着测评形式日益丰富,教育测评由强干预方式逐渐向弱干预方式发展,积累了越来越多在过去传统测评形式下无法获得的多模态测评数据。在典型的在线学习场景下,网络平台上除了学生的学习的结果数据之外,还有大量在线学习的过程性信息。结果数据和信息化背景下的过程性数据能够被共同使用于测评中,过程性数据带来的信息增益

有助于更加准确地判断学生的学习状况。尤其值得注意的是,从测量的角度来说,这意味着构念的定义从过去基于“作品”的定义,转化为基于“过程”的定义。学生在学习过程中与社会、自我和领域知识之间的交互构成了整个构念的范围。在这个交互过程中,学生的非认知产出、高阶思维与元认知策略也从多源数据中被捕获。从政策的角度来说,这些过程性交互中所有认知元素的合集本质上是将核心素养中比较抽象的政策性表述概念(如自我发展、社会性)对接为心理学研究中惯常使用的学术概念,使一些抽象的构念更容易分解为测评人员能够有的放矢的测量元素。

以写作自动测评为例,目前计算机评分大多聚焦于写作产出的作文的措辞、语法、流畅性等浅层语言质量,而人工评分一般关注作文的立意、主题、思想性等深层特质,将二者的评分信息有效结合,并在以下几个方面深入探索:一是计算机评分如何用来消解或平衡人工评分中存在的趋中性或评分偏误等问题;二是学生写作的过程数据有利于扩大写作自动测评覆盖的构念范围,例如在写作自动评估中,可以通过分析学生在写作、模拟和其他表演任务时产生的击键日志,在对作文语言质量测评的基础上加入对学生写作策略的测评,并给学生提供及时反馈;三是建立心理学范式的写作评价标准与计算语言学的分析处理方式的关联^[5],即将作文的思想性等特质与自然语言处理的底层语言表征数据进行结合,从而实现心理学范式下解释性和计算机测评下准确性的统一,同时,写作自动测评模型亦反哺写作测评的理论与标准。随着数据的日益丰富与变化,在原有技术下难以测量的写作测评维度被新技术所捕获,传统测量形式的载体被新的测评形式与测评数据信息所替代;同时,写作能力的概念与内涵在数字化环境中被更新与拓展,心理学的相关概念与理论得到相应发展,写作测评的维度与其发展规律在新的数据环境中被重新定义与解释。

以增强现实测评学生复杂问题解决能力为例,增强现实是多模态数据应用于测评的重要方式,研究者们越来越热衷于探索使用增强现实、虚拟世界和沉浸式任务来测评未来社会人才必备的技能。例如,通过虚拟现实来补充物理现实,让学生使用手持设备探索宇宙空间等。基于学生与虚拟现实交互的日志文件分析学生的动作模式,能够为每位

学生构建能力模型,并对学生在环境中的行为与互动实时更新,进而创建个性化的学习或测评体验。需要注意的是,这些方法目前仍处于研究或试验阶段,还需要足够的时间来评估其有效性。

(三)从自适应测评迈向个性化学习

从传统课堂与标准化考试到测评改进课堂教学,再到通过学习过程中的实时测评与反馈为不同学习水平的学习者提供不同的学习路径,使得发展自主学习与自我评价能力在“学评融合”的范式下成为可能。在个性化学习情境下,单次考试被移植到学习过程中,实现学习与评价的融合。事实上,基于认知诊断的自适应测评与学习进阶本质上就是面向学习者个体层面的学习路径推荐。将认知诊断应用于个性化学习的探索既是人工智能与教育结合的主要发力点,也是实现“学评融合”的前瞻性研究。

以自适应测评系统的发展为例,第一代测评系统根据考生的能力调整测试的难度与呈现的题目,并且反馈关于考生能力、变化过程和中介策略的信息。在这个过程中,项目反应模型与认知诊断模型被应用于估计考生的能力。随着计算机与测评模型进一步发展与融合,第二代测评系统允许多次连续测量学生能力,并以这种方式将测评集成到教学活动中,形成对学生能力发展的评价,第三代测评系统实现自动反馈与学习内容的推荐。随着多模态数据在测评中更多的应用与尝试,如对考生言语、面部表情或其他行为的分析,将学习者的元认知与情感状态纳入测评,在基于多模态数据对学习者的复杂反应进行评分的基础上,根据学生在系统中留存的长期学习档案进行解释,并在连续测量期间提供指导与建议。从测评模型来看,目前的个性化学习研究的典型框架是构建学习者模型,通过呈现适当信息与学习材料,基于测量模型的分析结果提供反馈和建议来创设符合学习者需要的智能化学习与进阶路径。其中,学习者模型包括学生对学科领域知识的掌握、认知能力、思维和社会性等方面的刻画,测量模型将学习者的特征在多维度上以时间序列进行判断,机器学习通过推荐算法寻找适切的方式代替教师给出学习策略。从对学习者的建模的“学习者模型”到“测量模型”再到“推荐模型”,不同于来源于外部测验产生的简单机械刺激反应,学生的学习行为在这个过程中一直真实发生并能得到及时反馈。这种嵌入学习中的测评允

许通过数字环境对学生进行持续的监控和指导,有助于在学习过程中结合过程性测评和终结性测评的证据以优化学生学习和课堂教学的进步,标志着测评从自适应测评向个性化学习转变。

三、教育测评在未来面临的挑战

当教育测评用于更好地促进教学与学习时,或者教育测评作为教学与学习的一个有机组成部分时,测评的目标和导向都发生了巨大变化,尤其是人工智能与教育测评融合的智能化测评,其研究定位、研究逻辑都是值得领域内外研究者与实践者深入思考的问题。教育测评不但要回答混合式教育教学新场景下学生学习状况并帮助其改进学习,还要关注学生的学习结果是否与国家政策导向以及社会发展的需求相适应。更重要的是,当教育测评的目标、参与者、形式、数据与模型都与过去传统考试有明显区别时,如何保证对学习效果和教学效果测评的有效性与公平性,无疑也是新时代教育测评研究者与实践者面临的艰巨挑战。

(一)智能化测评的效度与可解释性

首先,在系统地探讨心理与教育测量和人工智能与大数据之间的接口时,其关键仍取决于测量的有效性。只有将大数据特征与要测量的构念建立起联结,才能真正推进人工智能在教育评价中的应用。换句话说,就是将人工智能或大数据的方法引入教育测评时,如何证明这种方法测评的就是测评开发者想要测量的学生特质。在知识导向测评为主的应用场景下,测评开发者可以使用大量题目来不断探查学生水平以触及所要测量的构念;但是,在新型交互式、游戏式、虚拟式智能化测评中,活动设计的任务量往往非常有限,需要结合有限作答结果数据与大量过程数据对学生的某些能力或特质作出精准判断,在这个范式下,如何保证测评结果的效度是智能化测评必须回答的问题。

其次,心理与教育测量理论框架下的效度研究是不断发展变化的。从基于外部标准的效度、基于内容的效度、基于结构的效度,到21世纪以来以图尔敏模型为典型代表的效度论证,需要测评开发者与测评使用者在每个环节收集证据为效度提供解释^[6]。在效度的这条发展线上,目前智能化教育测评的效度研究才刚刚起步,非常类似于历史上效度发展的早期阶段,即20世纪初期基于外部标准的测评。在可见的教育测评公开研究与产品应用中,

智能化测评大都是以传统心理与教育测量测评结果或人工评判结果为外部标准来建构测评模型,与完整的效度论证研究相比还有很远的距离。从论证视角看智能化测评的效度时,在测评的参与者、测评数据的采集、模型的训练与测验、特征分析、测评的稳定性等一系列环节上,都需要与教育测评效度研究的理论与框架接轨。

最后,当前应用于教育测评中的人工智能与大数据算法的推演,包括实现个体分类的内在逻辑往往高度抽象且难以解释。无论是事前解释还是事后解释,大都与构念的心理学概念或测评标准相分离。从理论上来说,心理学能够提供对个体特质和赋分较为清晰且明确的理论与解释框架,将教育测评范式的概念和框架嵌入大数据和人工智能模型的“黑箱”中,既能提高分类效率,也能使人工智能和大数据方法得到的分类结果具有更好的可解释性。从研究实践来说,这既符合对2种学科理想融合的预期,也是近年来人工智能领域内的研究热点^[7]。因此,效度与可解释性是智能化教育测评的有效性检验需要兼顾的2个方面,二者既有重叠也有区隔,应避免只关注数理计算逻辑关系、无效度证据与解释而使教育测评陷入迷航。

(二)新时代教育测评的伦理与公平

新时代的教育测评面临前所未有的伦理道德与公平性挑战。一方面,在后疫情时代,很多学校的日常教学与考试机构的一些高利害的选拔性考试都采用了远程在线方式,信息泄露风险与考试安全隐患增加对信息安全与远程监考提出了更高要求。如何规范地使用数据,在更好地发挥大数据功能的同时,又能保证考试安全、保护个人隐私是新时代教育测评必须着手解决的问题。另一方面,当线上与线下的混合式教学与学习成为常态,教育与学习的公平性需要审慎思考与研判。教育测评的核心价值之一就是要保证每个孩子都有学习的权利^[8],即测评要有助于教育公平。由于在线学习必须依赖于网络与数字设备,社会经济发展与教育信息化建设的失衡很有可能进一步放大学习资源配置的差异。为了保障处于不同社会经济地位和教育环境中学生的学习权利,无论是国家政策制定者、教育管理者还是教育工作者,都要基于科学的教育评价结果,在教育资源分配、社会和家庭教育工作上付出更多努力。

科技与疫情的双重影响加速推动了世界各国教育系统为适应正在形成的教育新生态而作出应对措施,如何在“学评融合”范式下评价学生在未来社会必备的关键能力是整个教育领域的关注焦点。在教育测评范式转向的关口,将教育测评理论或框架与人工智能或大数据方法进行有效融合,既能利用计算机数据驱动优势提高分类效果,也能基于教育测评理论对所测评的构念和赋值分数给予理论化的定义和解释。需要强调的是,智能化教育测评并没有统一的测评形式或技术方法,人工智能与教育测评在实践中的融合与应用需要考虑到各地情况与教育系统的实际特点。无论是教育政策制定者、跨领域的研究者、教育一线工作者还是教育评价专业人士,都需要审慎理解与思考这场变革带给教育测评的机遇和挑战,同时坚守教育测评的标准、原则与实质性意义。通过科学有效地开展融合人工智能的教育测评,实现促进学生全面发展、培养担当中华民族伟大复兴时代新人的目标是使命,更是责任。

参考文献:

- [1] UNESCO. UNESCO warns 117 million students around the world are still out of school[EB/OL]. (2021-09-16)[2022-07-21]. <https://en.unesco.org/news/unesco-warns-117-million-students-around-world-are-still-out-school>.
- [2] OECD. Synergies for better learning: An international perspective on evaluation and assessment[M]. Paris: OECD Publishing, 2013.
- [3] 张生,王雪,齐媛. 人工智能赋能教育评价:“学评融合”新理念及核心要素[J]. 中国远程教育, 2021, (2): 1-8.
- [4] 教育部关于全面深化课程改革,落实立德树人根本任务的意见:教基二[2014]4号[EB/OL]. (2014-03-30)[2022-07-21]. http://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/201404/t20140408_167226.html.
- [5] 杨丽萍,辛涛. 人工智能辅助能力测量:写作自动化评分研究的核心问题[J]. 现代远程教育研究, 2021, (4): 51-62.
- [6] 谢小庆. 测验效度概念的新发展[J]. 考试研究, 2013, (3): 56-64.
- [7] MOLNAR C. Interpretable machine learning [M]. München: eint Publishing, 2020.
- [8] SIRECI S G. NCME presidential address 2020: Valuing educational measurement [J]. Educational Measurement: Issues and Practice, 2021, 40(1): 7-16.