

面向未来的下一代评估设计

——基于 OECD《创新评估以测量与支持复杂能力》报告的思考

李刚 赵佳琦 郑泽琳

(东北师范大学 教育学部, 吉林长春 130024)

【摘要】 评估是衡量和测评学生能力的重要方式,是确认学生发展水平的重要路标。经济合作与发展组织 2023 年 4 月发布的《创新评估以测量与支持复杂能力》政策文件,聚焦核心素养、深度体验、结构问题及决策发展来设计评估内容,并通过技术介入、革新概念框架、创新数据来源、促进自我调节学习、利用人工智能重构评估方式,最后通过强调跨文化的有效性与可比性、强化过程数据的科学性及其可塑性、促进学习分析与教育评估双向融合三方面革新评估结果的使用。这一政策文件对于推进我国教育评估革新具有三方面启示:一是促进教育评估内容素养化,二是促进教育评估方式智能化,三是促进评估结果使用高质量化。

【关键词】 下一代评估;高质量评估;重视证据

【中图分类号】 G423.04

【文献标识码】 A

【文章编号】 1007-2179(2023)06-0034-08

随着全球化进程的加快,各国教育改革都聚焦学生复杂能力和高阶思维的培养。2012 年,欧盟委员会发起了“重新思考教育”的改革,强调要促进学生横向技能的发展,如批判性思维和解决问题的能力等。自 2012 年起,国际学生评价项目(PISA)陆续开始测试学生协作解决问题、跨学科能力。随着越来越多的国家将创造性思维、自我调节和解决问题等复杂能力纳入课程目标,相应的评估方式也需不断变革。在此背景下,经济合作与发展组织(Organization for Economic Co-operation and Development, OECD)于 2023 年 4 月发布《创新评估以测量与支持复杂能力》政策文件。该文件包含创新评估内容、评估方式和评估结果的使用方式三方面内容,着重探讨如何设计下一代评估方案以

测量和支持学生复杂能力的发展。本研究对其进行梳理与分析,以期为我国构建高质量教育体系,实现从教育大国向教育强国的跨越提供启示。

一、创新评估的基本观点

早在 2018 年,OECD 在针对全球能力与创造性思维的 PISA 测试中发现,复杂能力的评估是一项挑战性任务,未来的教育评估需要有所改变。之后,OECD 组建了创新评估工作小组并进行多次讨论,形成了复杂能力导向创新评估的基本认识。

(一)评估作为证据推理的过程

OECD 认为,教育评估是基于证据推理的过程,旨在观察学生学习行为并产生对学生已掌握内容作出合理推断的数据和工具。OECD 将基于证据

【收稿日期】 2023-09-10

【修回日期】 2023-10-22

【DOI 编码】 10.13966/j.cnki.kfjyyj.2023.06.004

【基金项目】 2022 年教育部哲学社会科学重大攻关项目“双减背景下基础教育课堂形态及高质量发展研究”(22JZD047)。

【作者简介】 李刚,副教授,硕士生导师,东北师范大学教育学部 STEAM 研究中心负责人,研究方向:课程与教学论、科学教育;赵佳琦、郑泽琳,硕士研究生,东北师范大学教育学部,研究方向:课程与教学论。

【引用信息】 李刚,赵佳琦,郑泽琳(2023).面向未来的下一代评估设计——基于 OECD《创新评估以测量与支持复杂能力》报告思考[J].开放教育研究,29(6):34-41.

的推理过程描绘成由“认知、观察、解释”三个元素构成的“评估三角形”，三个要素分别位于顶点位置(见图 1)。其中，认知顶点指理论、数据和一组关于学生如何在知识领域表达知识和发展能力的假设；观察顶点是对评估任务的描述或一组规范，评估设计者可以使用这一功能最大化地收集数据的价值；解释顶点包含用于从易错的观察中进行推理的所有方法和工具，表达了从评估任务中获得的观察结果如何构成所要评估知识和技能的证据。在这个三角形中，三个要素本身要具有意义，且必须以有意义的方式分别与其它两个要素相联系，并同步实施以此形成有效的评估和合理的推论。

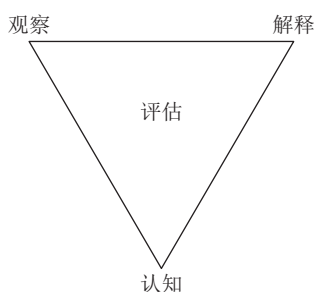


图 1 评价三角形

(二) 创新评估的三种主要趋向

欧盟委员会在“重新思考教育”改革中强调要促进学生的横向技能培养，如批判思维、解决问题能力等(Barnett, et al., 2002)。美国 21 世纪核心素养也特别指出批判性思维和解决问题、沟通、协作及创造和创新等能力是学习者应对未来挑战的基本技能。教育评估也需要进行改革创新(Burkholder, et al. 2020)。OECD 指出，创新评估需要在以下三方面展开深度探索：一是转向重要能力的评估，即评估的目标和内容不能仅停留在可评估或易评估方面，应转向学生所需具备的复杂能力；二是开展以证据为中心的评估设计，这个过程始于确定评估者希望能够对学生的知识及应该如何理解某些特定内容方面的要求；三是结果的有效解释和使用，评估结果能否得到有效解释取决于是否明确能力结构的有关概念、是否有与能力结构相关的论据及解释论据的方法。

(三) 建立更加协调一致的评估体系

我们无法通过某次评估测量出学生所掌握的

全部知识与技能，也无法使用某种工具同时评估家长、教师、政策制定者等提出的所有教育目标的达成程度，因此建立协调一致的评估体系十分必要。此外，连贯、协调一致的评估体系需要适用于从课堂到更大的组织单位(如地区、州、国家和国际)的各种教育环境，在支持教育系统的宏伟目标的同时，满足不同利益相关者的需求。科学的评估体系要求每项评估要素都能反映学习者的能力结构，每项评估结果都能反哺教学实践，因此评估体系的设计要依据以下四条标准：一是不应只强调机械记忆或操作，而要关注有助于知识迁移学习任务的设计，将基础知识和基本技能与培养批判性思维和知识迁移能力相结合；二是准确评估关键能力，如沟通、写作等复杂能力；三是使用具有教学敏感性和教育价值的内容，即评估内容应能够测量学习者对基本内容的掌握程度，并有助于将评估结果应用于教学活动；四是能够辅助开展有效、可靠、公平的评估，获得真实准确的评估结果，达到既定评估目标，从而充分利用评估结果提高教学质量。

二、创新评估内容

21 世纪以来，信息和通信技术从根本上改变了人们的生活，人类社会越来越依赖创造性思维、创新和解决问题等复杂能力来提供新的产品和服务。OECD 指出，促进学生复杂能力的发展不仅依靠教育教学变革，同时需要注意教育评估的创新设计，而评估内容作为衡量和测评复杂能力的重要组成部分，亟需更新和完善。

(一) 基于核心素养设计评估内容

进入 21 世纪以来，学校教育与学生学习的重点正由海量知识的记忆转向发展核心素养，评估是学生核心素养发展的重要路标，是表明学生应该学习什么及可以做什么的重要方式。

OECD 指出，核心素养是复杂多维的，围绕核心素养设计评估是创新评估内容的重要方面。这种创新主要表现在六个方面：一要明确核心素养构成成分及其结构和学习进度，即设计评估内容时需要对核心素养的变量进行明确的定义并确定其形成的过程，以便为任务设计、证据收集和评价提供信息；二要体现一般性和领域性特征，即

内容设置既要考虑一般意义上的核心素养内容,又要体现特定领域对特定素养的考察;三要使行为和思维过程可见,即使用开放、互动性的真实任务为学生学习行为和思维表征提供相应环境;四要产生能解释和用于可评估的证据,即评估内容的设计能够产生解释和说明学生素养发展的数据,借此将所作出的行为与使用的策略关联起来;五是评估报告要体现学生的整体发展,即评分量表要尽可能扩展测试数据的关联性,全方位地提供学生在测试中的整体表现;六是确保评估模型的一致性,即科学地评估和解释学生的表现,不受族群特征及文化背景的影响,确保评估项目的整体公平性。

(二) 基于深度体验设计评估内容

在学习科学的视域中,任何评估都涉及观察个人在少数特定情况下的表现。传统评估无法为描述和解释个体在测试环境外表现出的能力提供有效推论。学习科学认为,深入的学习体验能够为学生未来学习做好准备,评估学生如何有效地参与深度学习体验更有价值。据此,OECD从五方面提出促进学生体验的创新评估内容设计:

一是设计级联扩展的任务序列。当评估的重点转移到学生在丰富的环境中构建新知识的能力时,教师要给学生充分的时间展示他们可以想什么、做什么。这需要通过完整的任务序列,让学生能够时刻参与思考。二是评估设计和表现报告要明确说明具体的领域知识。学生素养发展是在特定环境中进行的,设计完全脱离情境的评估问题或者场景是无效的(张均兵,2013)。三是充分利用评估中的“错误”发展学生复杂能力。面对新的评估测试,学生往往会犯错,这一过程可以帮助他们识别已有解释和程序的局限性,并建立新模式与新解释。四是提供有效反馈和指导。反馈与指导必须基于任务需求为学生提供解决方案的完整模型,同时还要与其学习表现对接。五是设计“低地板+高天花板”的任务挑战。根据不同能力调整评估任务,不仅可以提高评估质量,还可以提高评估体验的真实性和吸引力。

(三) 基于结构问题设计评估内容

在21世纪,不同类型的真实问题或学习活动需要不同的知识和技能组合。智能技术的支持让

学习者可以享有多样的学习与丰富的学习资源并形成复杂的学习经验。在此背景下,创新评估设计的重点应聚焦建构一套能够充分考虑多种背景要素及全面测量学习者复杂能力发展程度的评估系统。

OECD指出,收集学生如何参与不同类型活动的的数据,可以对学生能力发展进行科学评估,并基于评估结果对其未来能力的发展作出正确决策。因此,创新评估设计应重点关注以下五个侧重点:一是以学生能力水平和知识应用领域为创新评估设计的依据。评估设计应确保大多数学生的学习取得进展、少数优秀同学能够创新问题解决方案,并涵盖多类型任务以便测试学生创造知识及解决真实复杂问题的能力。二是开展以实践为导向的学科评估设计。目前对学生学习成果的评估主要集中在学科领域,但应由侧重记忆知识转向将知识应用于在真实环境中解决新问题的能力。三是探索跨学科评估。组织跨学科领域的评估任务能促进学生更好地理解和应用学科知识。四是创设新型评估集群。OECD认为,传统的问题解决评估模型不能完全提供学习者认知、元认知、态度和社会情感技能等方面的丰富证据,故创设了三组评估集群,即搜索、评估与共享信息集群,理解、建模与优化系统集群,设计创意产品集群。五是设计集成协作任务,让学生通过协作解决复杂问题发展知识技能。

(四) 基于决策设计评估内容

解决新的复杂问题是现代社会需要的核心能力,但如何有效地评估这些复杂能力是一大挑战。OECD关于科学家与工程师解决复杂问题过程的研究发现,新手与专家在识别重要知识特征和使用组织良好的知识减轻工作记忆负担的能力方面存在重要差异。专家善于利用反思策略,应用有组织的知识库更好地定义问题,解决问题过程中采用的决策几乎决定了解决方案的每一项行动,并可以表征解决问题的过程。因此,OECD研究团队深度探索决策这一可识别、可衡量的维度,并将“科学问题解决”的结构定义为真实问题解决所需的一组技能。

在此背景下,OECD指出,使用创新评估框架

开展评估内容设计还要考虑三方面策略:一是让学生参与评估或改进同伴提供的解决方案,二是从一般原理性问题开始逐渐关注需要重点解决的特定领域问题,三是通过设置开放性问题或复杂多项选择题等方式考察学生的思考与行动等。

三、创新评估方式

在信息社会,教育评估的理念、环境与资源都发生了重大转变,教育评估方式也需相应的变革。为利用评估促进学习者复杂能力的培养,OECD 从多个视角分析了创新评估方式的六种可能路径。

(一)挖掘技术介入评估的潜力

数字技术的高速发展极大地丰富了教育评估方式。但目前数字技术主要用于复制或者简化评估实践。事实上,数字技术在如何帮助研究者理解学生评估表现、探索评估与学习之间的关系方面潜力巨大。基于此,OECD 围绕如何在多元化评估方式中嵌入技术支持的循证设计展开论述。首先,评估重点应从简单的知识记忆转向知识运用能力。技术可用于动态模拟复杂问题解决过程,不仅能测量学习者的学习结果,还能测量学习者学习过程中的思维与方法的变化。其次,在任务设计上,技术拓展了评估任务的范围,高度互动的工具和沉浸式测试环境可以为学习者提供开放式、迭代性的动态问题,而且为全面立体评估提供了更多可利用、可捕获的证据源。最后,在证据的识别与积累上,技术可对学习者的复杂反应进行解析与综合评分、数据统计与预测,使教育评估更加个性化,更有针对性与准确性。

(二)设计复杂能力评估框架

设计评估框架是开展评估工作的前提,复杂能力评估对评估框架的设计提出了更高要求(Nikolaos et al. 2019)。OECD 指出,复杂能力的评估框架需要基于证据中心设计(evidence-centered design, ECD)理论进行整体建构,将评估目的、内容、证据和表现等连贯成统一整体,并应用数理模型对学生复杂能力进行基于证据的推理与测试。据此,OECD 提出了设计复杂能力评估框架的四个阶段:第一阶段是确定评估目标领域,即整理与确认评估目标内含的知识、技能等,详细说明复杂能

力的评估要求(测量什么)、获取数据(如何测量)及科学性保证措施(测量方法的合理性);第二阶段是定义学生表现,即定义复杂能力评估的相关变量(知识、技能与态度等),并确定这些变量之间如何相互作用,同时说明每个变量的学生不同表现程度(最低水平到最高水平);第三阶段是定义任务结构,即设计能够证明其表现水平的任务,包括任务的复杂性及嵌入的资源与形式等;第四阶段是定义证据规则,即将第二阶段定义的学生表现与第三阶段定义的任务结构关联起来,建立统计模型,并将生成数据作为评估学生复杂能力的基本证据。

(三)使用技术增强评估框架设计任务

复杂能力的评估需要良好的评估任务。数字化信息技术的高速发展为设计与复杂能力相匹配的评估任务提供了可能。教育领域的数字化转型指向信息技术不断融入和渗透到教育系统各环节、各层面,并促使教育生态发生系统性和综合性变革(王贺等, 2023)。OECD 提出了技术增强评估(technology-enhanced assessments, TEA)框架设计任务的整体思路,并详细阐释了技术对复杂能力评估的任务类型、任务特征和证据来源的重要价值。在任务类型方面,技术增强评估可扩展评估的任务类型,从传统的静态任务过渡到动态交互式和沉浸式任务,即通过模拟实验室或虚拟现实环境等技术方式可真实捕捉学生对知识与技能的应用程度,提升复杂能力评估的有效性(OECD, 2023)。在任务特征方面,技术增强评估在任务的适应性、反馈性方面有较大改进:一是单级或多级的自适应任务能提升其与学生选择之间的匹配性,二是人工智能技术可对学生学习表现提供个性化反馈,提升学生的元认知水平。在证据来源方面,技术增强评估所获得的证据包括结果数据与过程数据。结果数据包括选择应达数据(如选择题等)、书面响应数据(如文本写作等)和动作响应数据(如按照指令做出的行为等);过程数据包括时间数据(如完成任务的时间、非活动时间等)、解决方案的中间状态(如提交最终方案前的状态等)、操作日志(如使用提示、鼠标单击等)和生理数据(如脉搏、眼动追踪等)。

(四)提升复杂性数据分析的可靠性

为保证收集的复杂性数据能精准地描绘复杂

能力特征, OECD 强调提升复杂数据分析和处理的可靠性, 并提出了五个方面的思路: 一是数据分析模型的交叉融合, 即通过两种分析模式推论和解释复杂性数据。例如, Mirt-Bayes 模型使用 Bayes 网络模型对数据评分, 使用 IRT 模型对评分进行推理论证; 二是建立不同粒度的推理结果, 即从粗粒度(数据模块少)到细粒度(数据模块多)进行多维度的结果分析; 三是不断修正评估目标领域模型, 即利用反复分析和推论获得的数据修改复杂任务的结构; 四是不断丰富探索性数据内容, 即通过使用大数据丰富和完善探索性数据, 从而将探索性解释转变为确认性解释; 五是采用多种技术增强评估任务测量同一复杂能力, 即开发任务族, 使用基于相同标准的相似任务开展学习评估。

(五) 支持与评估学生的自我调节学习

复杂能力评估的目标指向应从静态知识的应用扩展到动态学习过程, 即学生自我调节学习(Self-Regulated Learning, SRL)的过程。自我调节学习指学生使用元认知策略发起学习行为和调节情感状态, 并构建和应用知识的动态过程, 整体呈现了学生学习过程的行动选择和情绪管理状态(Partnership for 21st Century Skills, 2023)。OECD 指出, 自我调节学习有利于学生进行意义创造并获取新知。数字化技术能够将学生的自我调节学习过程记录下来并进行分析, 从而有效支持与评估学生自我调节学习。在支持方面, OECD 认为数字化技术可以在三方面为学生学习提供帮助: 模拟实验、情境反馈与信息搜索。其中, 模拟实验指学生通过交互模拟显示自己的探索过程, 情境反馈指通过标记错误行为并加以解释来提升学生的元认知策略, 信息搜索允许学生获得解决问题的各类提示与示例。在评估方面, OECD 指出, 设计者评估自我调节学习要注意四方面内容: 一是明确评估目标, 即尽可能将当前状态与目标状态之间的差距加以量化, 以便技术介入和支持对学生学习任务的评估; 二是设计证据规则, 即明确学生自我调节学习的行为序列与分析模型, 从而在大量学习行为证据中筛选有效证据; 三是避免泛化, 即要关注评估任务设计的有效性, 通过多个证据规则聚焦自我调节学习的根本特征; 四是避免先验知识的影响, 即考虑学生已有知识的

基本情况, 明确学生学习过程和结果的变化。

(六) 开发基于人工智能的自适应评估框架

OECD 指出, 学生复杂能力的评估不同于传统意义上对知识获得结果的考察, 它需要提供个性化的任务反馈, 即将复杂能力的评估过程视为多元交互与动态发展的过程。据此, OECD 从智能辅导系统(intelligent tutoring systems, ITS)获得启示(Vasile et al., 2013), 指出复杂能力评估需要建立基于人工智能的自适应评估框架, 即通过数字智能导师或同伴为学生提供智能反馈, 根据学生行为调整学习内容, 在评估学生当前知识水平的基础上, 确定学生达到目标状态还应作出哪些改进, 使评估更加个性化。当前教育评估变革是技术驱动导向的, 指向以学生者为中心的评价生态重塑(徐瑾劫等, 2023)。OECD 将技术驱动的评估框架分为三部分: 一是数字智能导师或同伴, 指用于指导或帮助学生个性化学习, 同时不断评估学习过程和结果的人工智能代理; 二是数字智能导师或同伴与学生之间的交互, 即由智能导师或同伴发出刺激, 随后获得学生的响应并进行验证, 同时发出新的刺激动作; 三是形成数据矩阵, 主要包括学生响应类别与智能导师反馈。响应类别分答对或者答错两类, 智能导师分别对学生自我调节学习提供积极、及时的反馈, 这样数据结构就形成了 2*2 矩阵, 这些数据矩阵成为分析学生复杂能力水平的重要依据。

四、创新结果使用

传统能力评估对于评估结果的使用还存在诸多问题。例如, 评估者仅关注评估结果的横向比较, 过分强调技术理性而忽视过程数据的使用价值等。OECD 指出, 面向未来的复杂能力评估应重视评估结果的价值和意义, 并给出三方面的建议。

(一) 将社会文化视角纳入评估设计

不可否认, 社会文化背景会影响学生复杂能力的形成、参与评估的积极性和对评估项目的理解及其在评估中的表现。OECD 指出, 开展学生复杂能力评估应将社会文化视角纳入评估设计, 提升不同社会文化背景下复杂能力评估的有效性与可比性的关键是评估设计的等效性, 并提出了保证评估等效性的三个步骤。第一, 确保评估结构的等效性,

即通过调查或专家判断识别不同社会文化中学生学习和评估的行为与理解特征等,确定学生能够以相同的方式理解评估结构,获得有意义的评估结果。第二,优化任务模型的有效性,即评估者制定任务模型需要考虑目标的定义是否因社会文化背景的差异而影响学生的认知过程等。第三,保证解释模型的一致性,即评估者使用复杂能力评估的解释模型(包括评分模型和测量模型)要不断训练模型,以获得不同社会文化背景下学生群体学习同等质量的证据。

(二)重视过程数据的使用

复杂能力评估需要特别关注过程数据的处理和使用,这是因为除了结果数据,过程数据能更充分地反映学生参与评估过程的思维方式与行为路径。OECD指出,过程数据对复杂能力评估的价值主要体现在三个方面:一是过程数据可以提高评估设计的可推广性,并增强评估得分的解释意义;二是过程数据能详细地从不同角度阐明评估结果数据的来源,既不遗漏学生的某方面能力,也不过度臆断学生某方面的发展水平;三是过程数据能测量学生与评估系统的互动过程和程度,并通过两种数据对比评估的结果,从而在一定程度上消除文化差异带来的评估误差,提高跨语言、跨文化评估的公平性。OECD还提到使用过程数据要注意的两个方面:一是过程数据只是学生思维过程痕迹的信息来源之一,实际的最终评估还需要其他信息的辅助,比如取得有声思维、眼动跟踪等多模态数据的支持与互补(臧玲玲,2020);二是要关注学生特质、教育水平与文化背景对过程数据的影响,充分利用技术优势,消解主观性对教育评估的影响。

(三)促进学习分析与教育评估双向融合

数字技术赋能的教育评估能精准描述学生复杂能力发展状况并提出可信证据,但数据的积累不是教育本身的目的,关键在于如何利用数据分析手段真正改变学习,提升学习质量。比较而言,学习分析(learning analytics, LA)关注学生个性化学习过程并使用相关数据对其效果进行解释和预测。OECD指出,学习分析和教育评估虽然目的不同,但都能为干预学生学习提供信息,因而研究者不应该分别加以研究,而要促进二者的双向融合,

完善现有的教育评估和学习分析框架。一方面,学习分析可以从教育评估中获得三点启示:1)开展学习分析时,仅获取代表性数据是不够的,还要通过透明合理的证据汇总对所得出的结论进行有效辩护;2)学习分析要基于一定的理论准备,而不是仅根据所得证据选择最佳数据拟合提出观点或结论;3)学习分析不是为了获得新的学习结构或产生新的学习见解,也不仅仅是开发新的指标,重要的是借此建构有意义的分析指标。另一方面,教育评估可以从学习分析中获得两点启示:一是真实的任务可以提升评估的精度与准确性,大幅提高学生在评估中的参与度,丰富他们的体验;二是不应过度依赖某个评估项目或评分结果的完美解释或拟合,而应关注多因素的整体诊断。

五、总结与启示

在可以预见的未来,对学生复杂能力的评估将成为国际教育评估的重要使命,也是新一代教育评估的变革趋向。在我国教育迈向高质量发展的新时期,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》,期望扭转不科学的教育评价导向,坚决克服唯分数、唯升学、唯文凭、唯论文、唯帽子的顽瘴痼疾,提高教育治理能力和水平(田腾飞等,2014)。OECD发布的《创新评估以测量与支持复杂能力》为我国教育评估改革提供了三方面启示。

(一)教育评估内容素养化

教育评估内容承载着评估体系的主旨要义,是引导教育系统发展的指挥棒,关系到教育生态能否持续运作。《义务教育课程方案和课程标准(2022年版)》更加关注学生素养的发展。实现教育评估内容的素养化就是要关注对评估内容本身的元评估,尤其关注其时代性、民族性及实施弹性,使评估内容与教育进程相匹配,既保证评估内容的信度与效度,描摹完整的学生成长画像,又保持评估内容在一定时间内能指引教育生态的发展。

教育评估是一项专业性强、知识面广的工作。社会第三方评估机构的专业化水平取决于评估队伍的专业化程度。建立一支高素质的人才队伍是教育评估机构提升服务质量的重要保障(王璐等,

2018)。随着信息化时代的高速发展,基于证据的评估受到高度关注,循证义务教育学业评价不仅要基于证据,而且要寻求到最佳证据。一般认为,最佳证据指研究者提供的与解决问题最契合、级别最高的研究证据,能够尽可能地接近事实真相,揭示出问题的症结(杨文登,2010)。在学业评价中,最佳证据在很大程度上决定了学生成长画像的適切情况。学业证据浩如烟海,可以是测评或调查数据,也可以是表现或访谈资料,包括电子证据、实物证据等。与此同时,教育现象本身的复杂性也会导致证据的多样性。由此,教师如何选择与识别学生学习成果方面的最佳证据尤为重要。

(二)教育评估方式智能化

评估方式影响评估效率,智能技术的广泛使用为教育评估改革提供了更多可能(袁莉等,2021)。首先,利用技术可使评估的重点从让学生回忆事实等高结构化的内容转向更复杂多样的开放性内容;其次,利用技术可突破传统纸笔测试的弊端,利用高度互动的工具进行更丰富的任务设计;最后,技术进步能提升数据收集和储存的能力,利用测量技术解析、统计和推理复杂的评估证据,并对其进行解释。

目前,数字信息技术可通过平台对社会各领域产生影响,教育评估也被赋予了新的内涵。互联网+教育、ChatGPT赋能教育等提法屡见不鲜。技术支持的教育评估也应向着智能化、即时化的方向不断迈进,推动一次性、单向性的评估转变为持续性、综合性、动态实时、基于证据的循证评估。虽然信息技术的发展为教育评估提供了更多可能,但是评估者不能过于依赖技术而偏离了评估的真实目的,必须时刻检验其与评估目的的契合性、一致性,以促进技术有效地服务于评估的预期目的,更好地促进教育评估的发展。

(三)评估结果使用高质化

面对新时代教育改革的新挑战,教育者首先应摒弃教育活动和教育评估隔离的观念,自觉将两者统一起来,认识到评估结果使用质量的高低将影响后继教学质量,高质量应用教育评估结果有利于形成教育教学活动和教育评估之间的闭环。首先,教育评估结果的使用要向纵深挖掘,使学有所评、评

有所用,用齐、用足各项评估结果。评估结果不是摆在眼前看的,而是握在手里用的,不应让评估结果成为无意义的数字,而应着力探究评估结果的高质量使用途径,使有意义的评估结果作用最大化。其次,高质量使用评估结果离不开围绕教学目标对结果的合理总结提炼。无论是教师、学生亦或是教育行政人员,都要衡量当前评估结果与预期教学目标的差距,对评估结果进行合理的因果分析、内化,同时考虑预期外的评估结果,必要时可修正评估内容。最后,高质量使用评估结果须依托动态的教学转化与学习转化,促进评估结果在教师与学生的双向互动中融入教学实践。教师应成为使用评估结果的带头人,积极引导、培养学生自我学习能力,使学生逐渐具备独立使用评估结果优化个人学习活动的的能力,甚至参与评估体系的调整,使评估体系不断完善,始终处于螺旋式前进状态。

[参考文献]

- [1] Barnett, C.(2002). When and where do we apply what we learn?a taxonomy for far transfer[J]. Psychological Bulletin, (128): 612-637.
- [2] Burkholder, E. W, Miles, J. K, Layden, T. J, Wang, K. D, Fritz, A. V, & Wieman, C. E.(2020). Template for teaching and assessment of problem solving in introductory physics[J]. Physical Review Physics Education Research, (16): 1-20.
- [3] Nikolaos, P., Panagiotis, F., Ioannis, K., & David, W.(2019). Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning[J]. Virtual Reality, (4): 23.
- [4] OECD(2023). Innovating assessments to measure and support complex skills[EB/OL]. [2023-05-14]. https://www.oecd-ilibrary.org/sites/e5f3e341-en/1/3/2/2/index.html?itemId=/content/publication/e5f3e341-en&_csp_=f8cd08a73f7411a73eba225857184a20&itemIGO=oeed&itemContentType=book.
- [5] partnership for 21st century skills(2023). 21st century readiness for every student: A Policymaker's Guide[EB/OL]. [2023-05-14]. https://www.commoncorediva.com/wp-content/uploads/2014/10/policy-makersguide_final.pdf.
- [6] Vasile, R., Sidney, D. M., Hu, X. & Arthur, C. G.(2013). Recent advances in Conversational Intelligent Tutoring Systems[J]. AI Magazine, (3): 34.
- [7] 田腾飞,刘任露(2014).元评估——教育评估专业化发展之必需[J].外国教育研究,41(6):111-119.
- [8] 王璐,王小栋(2018).英国第三方教育评估的发展与规范:基于英国独立学校督导团的研究[J].外国教育研究,45(2):16-28.

[9] 袁莉, 曹梦莹, 约翰·加德纳, 迈克尔·奥利里(2021). 人工智能教育评估应用的潜力和局限 [J]. 开放教育研究, 27 (5): 4-14.

[10] 王贺, 肖甦(2023). 俄罗斯学校数字化成熟度评估的价值解构、路径解构与经验反思——基于对“教育机构数字化转型监测”项目的分析 [J]. 外国教育研究, 50 (9): 97-112.

[11] 徐瑾劼, 申昕(2023). 重塑以学习者为中心的教育评价生态——基于教育评价智能化发展的全球观察 [J]. 开放教育研究, 29 (3): 40-46.

[12] 杨文登(2010). 循证实践: 一种新的实践形态 [J]. 自然辩证法研究, (4): 106-110.

[13] 张均兵(2013). 易克萨维耶·罗日叶情境化命题思想的启示 [J]. 中国考试, (6): 26-32.

[14] 臧玲玲(2020). 构建新的学习生态系统——OECD 学习框架 2030 述评与反思 [J]. 比较教育研究, 42 (1): 11-18+32.

(编辑: 李学书)

Future-generation Assessment Design: OECD Innovation Assessment to Measure and Support Complex Capacity

LI Gang, ZHAO Jiaqi & ZHENG Zelin

(Department of education, Northeast Normal University, Changchun130024, China)

Abstract: *Evaluation is an important way to measure and evaluate students' ability, and is an important signpost for students' development. OECD released in April 2023, the innovation assessment to measure and support complex ability " policy documents, focusing on the core literacy, depth experience, structure problems and decision development to design evaluation content, and through technical intervention, innovation concept framework, innovation data source, promote self-regulating learning, using artificial intelligence to reconstruct evaluation way, finally by pay attention to the effectiveness of cross-cultural and comparability, strengthen the science and effectiveness of process data, promote learning analysis and education evaluation two-way fusion to innovation the use of the evaluation results. In general, its enlightenment for promoting the innovation of education evaluation in the high-quality development of education in China in the new era has the following three aspects: first, to promote the literacy of education evaluation content; second, the second is to promote the intelligence of education evaluation methods; third, to promote the high-quality use of evaluation results.*

Key words: *next-generation assessment; high-quality assessment; emphasis on evidence*