

江苏教育考试科研月报

2025年第7、8期（总第126、127期）

江苏省教育考试院编印

本期内容简介

- 【热点透视】 以“教育评价的时代变革”为主题，从宏观战略到微观方法，系统聚焦教育评价领域的前沿理论与创新实践。
- 【研路同行】 教育考试命题的“四性”及其实现路径；国家教育数字化战略背景下自学考试卷库管理系统建设研究。
- 【域外资讯】 经合组织发布《提升 PISA 无障碍性：PISA 题目修订与设计指南》；UNESCO 发布《全球教育监测报告：引领教育技术发展东亚区域版》；拉美地区高等教育公平的挑战与改进策略；法国制定职业教育质量提升与行动治理计划。

【热点透视】

教育评价的时代变革

【编者按】教育评价作为事关教育发展方向的指挥棒，既是衡量教育质量与人才素养的准绳，亦是引导教育改革、优化资源配置、促进教育公平的关键机制。构建科学、多元、契合时代发展需求的教育评价体系，已成为当前我国教育事业发展的重点任务之一。本期月报重点关注教育评价领域的前沿理论与创新实践，内容涵盖从宏观战略到微观方法，从国际比较到本土探索的多个维度。通过系统剖析新时代教育评价体系的定位，在方法论创新层面深入探讨人工智能、大数据等技术对人才培养与评价模式的革新路径与优化策略，为教育评价实践探索提供学术参考。

论教育评价的战略属性

谢维和（清华大学）

教育评价的战略属性是教育战略属性的组成部分，也是实现教育战略属性的重要途径与载体。教育评价战略属性的主要特点是跳出教育评价教育，而且是从国家经济社会发展与科技进步的宏观视野评价教育发展。

教育评价战略属性的引导功能

所谓教育评价战略属性的引导功能，主要

是通过教育评价宣传与推广教育质量的新标准与新内涵，引导教育的高质量发展。根据高质量发展要求和《强国纲要》有关内容，新的教育质量标准可以概括为“六个力”，即思政引领力、人才竞争力、科技支撑力、民生保障力、社会协同力及国际影响力，分别从德育地位、德育实施有效性、人才素质、科学教育定位、

办好人民满意教育、与经济社会协同发展以及国际合作与交流等方面,全面地提出了新时代中国教育发展的质量标准。以思政引领力为例,当前,教育的根本任务是立德树人,强调德育与其他四育的内在统一性,不再单纯地通过分离的方式强调德育的重要性,而是通过强调德育对其他四育的引领和促进作用,最终实现树人目标。这种表述更加注重教育的整体性、系统性,强调德育的有效性和针对性。德育工作的成效不应仅仅以工作的完成情况来衡量,而应以是否能带动学校整体发展、是否能促进学生全面发展、是否能真正激励学生成长为评价标准。这种评价体系体现了教育理念的转变,从单纯强调德育的重要性转向关注德育如何在五育并举中发挥引领作用,最终实现培养全面发展人才的目标。

再如,将人才的竞争力作为人才培养质量的新标准,十分明确地反映了创新素养在新时代人才培养中的重要性。**对基础教育而言**,创新素养的培养主要体现为对中小学和高中阶段学生的差异化教育方面。在中小学阶段实行的“沃土计划”强调基础性,注重为学生成长提供适宜的土壤、环境和条件,使其能够发展和形成创新素养。该阶段的教育重点是构建良好的教育生态,为更多的孩子创造公平发展的教育机会。在高中阶段实行的“脱颖计划”则聚焦于让不同类型、特点和优势的人才脱颖而出。该阶段的教育目标是建立科学合理的识别和选拔机制,以及为学生提供个性化发展路径,帮助他们在各自的优势领域实现突破。**在高等教育领域**,这种创新素质与创新人才的培养更多地表现在学科、专业与课程建设等方面,特别是在新知识的发展与新领域的拓展中,以及通过集成创新与原始创新在经济社会发展与科技进步方面形成突破,等等。

显然,教育评价的战略属性要求我们重新审视新时代教育质量的内涵,进一步深化改革教育评价的指标体系与评价方法等。这种转变既是对教育质量标准的重新定义,更是对教育发展方向的精准把握。

教育评价战略属性的协调功能

教育评价活动复杂且涉及诸多因素与变量,把握与协调各方面的因素与利益相关者的要求,是实现教育评价战略属性的重要任务。

这种协调性既包含经济社会发展对教育的要求与教育活动自身逻辑之间的矛盾与协调,也包含教育活动本身诸多影响因素与变量之间的差异性与协调。

一是转换,指的是经济社会发展等对教育的外部规定性与教育活动自身规律之间的协调。教育质量具有外部规定性,其评价依据不能仅由教育系统内部决定,而应由社会、用人单位乃至未来需求判定。但需要明确的是,教育评价绝不是简单地按照这些外部规定性评价教育活动,不能简单地将社会要求直接作为评价依据。它必须通过转换,即把这些外部的规定性转化为教育本身的内在规定性与质量标准,进而创新教育评价标准。

首先,这种转化需要将教育外部的各种因素(社会目的、任务和要求)转化为适合教育教学和学生学习的形式,包括设计适当的课程体系、学科建制,以及体现循序渐进和成长规律的学习境界。**其次**,这种转化还必须将教育的外部目的和要求转化为合理的学习内容、学习目标和评价标准。

二是中和,是借用我国优秀传统文化的重要理念说明教育评价的战略属性,也是我国教育中优秀传统文化的特点。教育评价中的“中和”并非简单的折中,而是一种具有高度关联性的、恰当的“点”,由此实现教育评价中协调的战略功能。**从宏观视角看**,它包括支撑国家战略和满足民生需求、知识学习和全面发展、培养人才和满足社会需要、规范有序和激发活力、扎根中国大地和借鉴国际经验等重大关系。**从微观视角看**,它涉及教育活动的方方面面。第一,高考评价中服务选才和引导教学之间需要协调。高考承担着服务高等学校选才和引导中学教学的双重任务,如何在整体安排与命题设计中实现两者之间的平衡,是一个需要协调的重要问题。第二,专业评价中对口与适应之间需要协调。在人才培养中,既要满足企业、行业 and 用人单位对人才的现实需求,又要注重培养学生适应经济社会发展变化的能力,使其能够在未来的数字技术和人工智能发展的背景下不断提升自我。第三,学校管理评价中规范有序与激发活力之间需要协调。一方面,需要强调国家和政府的集中统一领导,确保教育管理的规范性和有序性;另一方面,要通过横向

交流和资源共享激发学校的办学活力。第四，人才培养中自主性与开放性之间需要协调。既要不断提高人才自主培养的质量，还必须坚持开放，在不断学习国际先进经验、提升自主性的同时体现判断力和选择能力。第五，在基础教育评价中把握全面发展与个性发展之间的协调。全面发展并不是抽象的，而是一种有个性

的全面发展；个性发展并非标新立异或者别出心裁，而是共性基础上的个性表现方式。综上，教育评价中的“中和”并不是简单的调和与机械的相加，而是一种有机的结合，是相得益彰的互补，是提升教育活动有效性的重要形态。

（摘编自《中国考试》2025年第6期）

新时代教育评价体系的价值定位：国际趋势与中国方案

辛涛、洪倩、李刚（北京师范大学）

国际教育评价改革的价值透视

1. 服务发展：教育评价功能的价值定位。

总体可以概括为两种取向：一种取向认为评价应该服务管理，通过对被评对象水平、等级等的鉴定服务于后续招生、奖惩、问责、晋升等管理相关环节；另一种取向则认为评价应该服务发展，通过为评价对象提供证据和改进策略，促进学生、教师或学校不断改进。从当前趋势来看，虽然大家并不否认评价具有管理功能，但是更为强调和坚持服务发展的价值取向。这一价值定位又可凝练为两个方面。

一是为了发展。诸多国家和国际组织均认为，应以学生的发展来定义发展，但是对此的认识却大不相同。有国际组织和国家认为，对学生发展的评价应该聚焦于学生的核心素养。但是在对区域和学校的评价与对教育从业人员的评价中，一些国家和地区仍然主要用学生的学业成绩加以衡量。这也反映出，虽然人们明确学生发展的重要性，但是在现实衡量中，简单主义的操作仍然占据巨大的市场。**二是加强改进。**主要存在两种路径。一种是强调基于证据的合作改进路径，形成性评价便是其中的典型样例。另一路径则强调通过基于评价结果的奖惩问责来督促被评价方改革。关于这一路径，有人认为体现了责任制的落实，也有人认为带有过于浓重的绩效主义色彩。

2. 精准刻画：教育评价内容方法的价值定位。**第一，关注过程。**例如，在2015年英国出台的《共同督导框架》中，学校评估指标体系包括领导力与管理的效能（18个指标），教学、学习与评价的质量（9个指标），个人发展、行为与福祉（13个指标），学生学业成就（8个指标）四个维度，前两者都是学校办学过程

的维度。一些方法的兴起迭代也反映了人们对过程的日益关注。例如，档案袋评价可以实现对学生学习过程和教师专业发展过程更为详细地记录，在20世纪80年代成为欧美国家过程性评价的重要手段。随着建构主义评价范式的深入，关注过程已经不再只是指对时序意义上的过程进行评价，更是强调要重视意义的建构、对多元资料特别是质性资料的收集等。**第二，重视增值。**增值评价是指追踪学生不同时间点的学业成绩，利用一定的统计分析方法对学业成绩变化情况进行分析，并注意剥离掉学生家庭背景等非教育因素影响，从而实现对学校或教师效能更为精准的考察。

3. 多元参与：教育评价组织实施的价值定位。**一是强调第三方的广泛介入。**第三方广泛地参与评价标准的制定、数据采集、分析反馈等诸多环节。**二是重视自我评价的力量。****三是重视评价的沟通协商。**

新时代中国特色教育评价体系的价值建构

在价值体系的顶层设计上，澄清学生全面发展的时代内涵，厘定根本价值遵循。学生全面发展是我国教育目的最为核心的表达，也是新时代中国特色教育事业改革发展的根本价值定位。**首先，结合新时代背景下人与社会发展的需要，进一步澄清学生全面发展的时代内涵，并以标准体系加强对学生全面发展的制度化定义。**当前，诸多政策文本对学生全面发展均有论述，但是相互之间的侧重与维度却并不兼容，需要加强顶层的统合。需要厘清学生全面发展的各维度要求及其相互关系，特别是要避免将学生体美劳等方面的发展过分技能化，要澄清学生各类能力共同、协调发展的水平，避免将全面发展视为平均发展、同等发展。**其次，结**

合不同评价领域，分析不同领域评价中学生全面发展的指向。例如，在课堂教学评价领域，学生的全面发展应更为侧重对学生的核心素养进行评价，让学生全面发展更加彰显时代特色，更具操作性；对学校的评估则更应关注学校落实“五育”并举、促进学生全面发展的举措与成效，以评价督促学校更好地促进学生的全面发展。最后，结合评价的要素与环节，分析学生全面发展对于评价的具体要求。例如，对于评价内容，要从促进学生全面发展的角度加以衡量筛选；对于评价方法，则要考虑其是否可以精准衡量学生全面发展的水平，并据此调整优化。

在功能定位的价值定位上，立足服务立德树人，辨清并统合服务发展的价值定位。要将立德树人作为检验教育工作成效的价值标准，一方面要让教育评价回归教育本身，破除教育评价价值定位的管理主义异化。要明确评价的根本目的是提升立德树人的质量、促进学生的全面发展，而不是只是满足外部的排序、分等、控制等。这就要求统筹立德树人的质量追求与效率追求。诸多管理主义的评价举措虽然更有效率，但是忽视了潜在的、隐性的大量投入，可能并不利于教育事业与学生的可持续发展。所以评价活动必须立足立德树人的质量来考察效率，需要从学生的长远发展来考察效率而不仅仅立足于短期目标的达成。这还要求不能简单以分数、升学率等来检验教育工作的成效。当然，破除“五唯”的顽瘴痼疾不能仅凭价值倡议，要将这样的价值定位制度化教育从业人员评价的指标体系，将评价活动中的“五唯”行为作为红线。另一方面，可以立足立德树人，对服务发展与服务管理的价值取向进行统合。评价要服务学生发展，但是并不排斥其对管理的直接功效。从服务管理的角度而言，改革的重点是聚焦立德树人，实现更为科学合理的评估，如对于教师的评价要关注其在课堂教学方面的改革创新与成效，而非仅仅对从业时间、工作量、活动参与次数等简单指标加以评价，从而激励教师、学校、区域真正致力于落实立德树人根本任务。就服务发展而言，改革的重点是通过评价明确立德树人的要求，提高评价

结果的科学性和认可度，让从业人员可以基于评价改进立德树人工作。

在评价的内容与方法的价值定位上，以开放的态度对待“精准刻画”的技术倡议，加强研究与试点。关于评价内容与方法的诸多价值倡议，应当秉持开放的态度积极学习反思，既不能因为原有评价的路径依赖而盲目排斥，也不能因为迷信其“科学”而盲目推崇。一方面，要加强对不同价值取向的深入研究。通过研究明晰每种价值取向的核心观点、优势与潜在风险，整理反映这一价值取向的样例，从而帮助利益相关方真正辨清并加以筛选。比如，当前实践领域对增值评价话语盲目推崇，很大原因就是未能明晰增值评价的观点、优势与不足。另一方面，要审慎地推动价值导向大规模落地。技术层面的价值导向从理想走向现实，需要大量的证据与实践支持。诸如过程性评价、增值性评价等倡议，可以先进行试点试验，验证其信效度与可行性，总结有效模式，形成操作指南，然后再思考利用政策工具大规模推广。最后，对于诸多价值倡议，要打好“组合拳”，切勿设置非此即彼的硬性要求，如强调过程评价便反对终结评价，强调增值评价便视根据最终结果的评价为不科学。

在组织实施的价值定位上，加强多元参与导向的制度与文化建设，探索评价育人的可能路径。要通过制度明确评价诸方的权责分配，明确参与的具体途径和流程，确保多元参与乃现实期望而非理念空谈。聚焦促进协商，要建立相应的议事机制，鼓励各方围绕评价工作理性交流、达成共识。要加强评价中协商参与相关理念与实践的介绍、培训与推广，形成座谈访谈、问卷调查、议事会、接待日等多种载体，让相关利益主体认同多元参与、实践协商参与。倡导教育评价中的多元参与，不仅仅是希望通过民主的方式获得更为合理、更具共识的评价结果，更重要的是将其作为育人的重要途径，评价各方不仅可以审视对教与学的理解，还可以切实提高自身的评价素养及其他能力，真正推进评价育人。（摘编自《国家教育行政学院学报》2024年第2期）

破界创生：数智时代创新人才培养与评价的新范式

郑泉水（清华大学）

一、传统教育评价的局限性与个性化成长的困境

传统教育评价体系往往侧重于对知识掌握程度的考查，以考试成绩为主要衡量标准。这种评价方式在一定程度上能够筛选出基础知识扎实的学生，但其局限性也日益凸显。首先，它无法全面衡量学生的综合素质与个性特长。每个学生都是独一无二的个体，拥有不同的兴趣、特长和潜力，传统评价试图用一把尺子去衡量所有人，这无疑会压抑学生的个性发展。其次，这种评价方式容易导致学生过度追求分数，忽视对知识的深入理解和应用能力的提升。许多学生为了在考试中取得高分，往往采取死记硬背、刷题等机械式学习方法，缺乏对知识的真正理解和创新思维的培养。

在数智时代，这种局限性更加明显。随着人工智能等技术的发展，许多重复性、规律性的知识和技能将被机器取代。未来社会需要具备创新思维、社会适应技能和解决复杂问题能力的人才。然而，传统教育及其评价体系难以培养出这样的人才。

二、创生教育理念的提出与实践

创生教育强调以创生为内生动力，鼓励学生自主探索、获取资源，促进学生创新能力的提升和个性化成长。未来的教育评价不应仅关注学生的知识掌握程度，而应更加注重学生的综合素质、创新能力和社会适应能力的培养。

清华大学钱学森班长达16年的人才培养探索便是一个成功案例。钱学森班的学生从清华大学优秀学生中经二次选拔产生，本身已经具备较高的学术水平和综合素质。钱学森班的培养目标是让每个学生都能在个性化成长的道路上取得成功，培养他们的自信和创新能力。以个性化的研究课题、丰富的学习资源和自由的学术氛围为依托，钱学森班的学生在创新能力和个性化成长方面取得了显著成果。许多学生在本科阶段就开始进行自己的研究项目，取得了优异成绩，并在国内外学术会议上发表研究成果。

创生教育理念已推广到南昌大学等其他高校，这些高校也创办了类似钱学森班的实验班，取得了良好效果。这些实验班的学生在个性化成长 and 创新能力方面表现出色，证明了创生教育理念的可行性和有效性。

三、X型学生与A型学生：教育评价的两种路径

在创生教育理念中，学生可以分为X型和A型两类。X型学生指具有探索精神、敢于冒险、注重个性化成长的学生。他们善于发现自己的兴趣和特长，勇于尝试新的事物，不拘泥于传统的学习方式和评价标准。而A型学生则指那些追求完美、注重成绩、按照传统教育模式学习的学生。他们在学习过程中往往更加注重分数和排名，力求在每门课程中都取得优异的成绩。

传统教育评价更适合A型学生，而创生教育理念则更注重培养X型学生。X型学生在个性化成长 and 创新能力方面具有更大潜力，能够在未来社会中更好地适应变化、解决复杂问题。然而，目前的教育评价体系却往往压抑X型学生的发展，使他们逐渐变成了A型学生。因此，我们需要改变教育评价方式，为X型学生提供更大的发展空间。

高考作为我国最重要的教育评价方式之一，对学生的成长和发展具有深远影响。高考改革是实现创生教育理念的关键。具体来说，高考应创新试题设计，更加注重考查学生的思维能力和创新能力。通过这种方式，可以为X型学生提供更多机会，使其在高考中脱颖而出。此外，还应重视高考评价的多元化。高考评价不应仅局限于考试成绩，还应包括对学生的综合素质、创新能力、社会实践能力等多方面的评价。通过多元化的评价方式，更加全面地衡量学生的能力和潜力，为学生的个性化成长提供更多机会。

（摘编自《中国考试》2025年第6期）

以学生为中心的综合评价

王殿军（中国教育发展战略学会副会长兼教育评价专委会理事长，教授）

1.教育评价中的突出问题。第一，教育的标准化、统一化特征明显，有利于均衡，但不利于学生的个性发展，难以适应学生的个性特点；第二，过分强调知识的学习，能力培养比较薄弱，如创新能力、动手实践能力和发现问题、研究问题、解决问题的能力等，这些能力很难通过纸笔考试来检测和评价，因此往往不受重视；第三，过于重视单学科学习，钻得特别深、特别难，而学生跨学科的学习特别弱；第四，对学生身心健康、全面发展关注不够。当最终选拔只看考试分数的时候，教师、家长和学生心目中就只有分数，孩子们就会牺牲身心健康和全面发展来先保分数，导致全面发展流于形式。

2.国际经验借鉴。我们系统考察了美国名校。它们在选拔录取学生的时候所考虑的因素比较多，包括申请表、个人陈述、教师推荐信、能力证明、标准化考试（SAT、ACT、TOEFL）成绩、高中的成绩单、大学预备课程成绩等。大学招生录取时关注的因素越多，学生在学习过程中就会更加注重全面发展和个性发展。在这样的选拔评价体系的引导下，学生能够享受到多元、丰富的校园生活，身心得到全面发展。

2017年9月，常青藤联盟联合80个高校建设了一个新的申请平台（CAAS），记录高中生4年的日常学习记录。CAAS系统具有三大功能：一是**储存柜功能**，存储学生在校期间的重要学习成果和个人成就，包括作品、证书、奖项等；二是**合作交流空间**，学生可以在平台上将积累的素材与校内外升学指导老师讨论，与大学招生办联系等；三是**申请门户**，学生可以将存储的所有材料进行整理编辑，按照不同

大学的要求提交申请。这个系统类似国内的高中学生综合素质评价系统。

另外一个重要的国际经验是重视以学生为中心的大规模学情调查，如美国中学生学习参与度调查（HSSSE），用以了解中学生对于学习、学校环境、与学校互动情况等方面的看法，目的是记录、描述并加强中学教育活动中的学生参与，并为有关教育教学变革的讨论提供信息。

3.国内实践探索。我们可以借鉴海外经验和国内高校成功的运行模式，建立以中学生为中心的综合评价系统（CHSSA）。CHSSA的核心功能是更系统全面地调查和分析，了解学生哪里做得好，哪里出了问题，以更好地服务师生。CHSSA还可以帮助学校和主管部门进行教育监测，对于学生和家长也有帮助。学生的学习参与情况可以从认知参与、行为参与、情感参与三个维度进行考查。

应该把综合素质评价和学情调查两部分结合起来。综合评价包括学生综合素质评价，也包括大规模学情调查，两方面结果相互印证才能更全面地评价学生。其功能和应用场景有四个主要方面：一是**服务学生的专业选择**，通过大数据画像帮助学生了解自己、填报高考志愿；二是**生成成长记录**，把学生的成长过程、成长状况和个性特点收集起来；三是**具有诊断功能**，帮助学生了解自己的长处和短处，为学生制订个性化的指导方案、成长方案；四是**根据数据分配学习资源**，为学生提供多种学习资源、反馈渠道、测评工具、分析报告、规划方案等。

（摘编自《中国考试》2024年第1期）

增值评价的优化策略

李俊飞、谭顶良（南京师范大学）

在教育公平与高质量发展理念的引导下，增值评价为教育评价提供了全新的视角和方法。增值评价理念扎根于教育公平与高质量发展观、以学评教的学生主体观和服务于改进的评价功能观，借助增值评价技术模型，从本质上探寻教育投入与产出的因果关系。但在实践中尚存一定的问题需要解决，主要包括技术模

型尚不成熟、评价内容片面单一，以及评价结果反馈效果不佳等。立足于增值评价的基本理念以及目前面对的实践困境，提出应用增值评价的优化策略。

1.系统测量和收集数据。由于增值评价需要借助相应的统计模型来计算增值结果，就必须测量和收集一定规模的数据作为支撑。完备

的数据支持是有效开展增值评价的基础。在收集数据之前,需要先全面系统地测量准确的数据。对于学业水平测试等标准化考试而言,这类数据是容易测得的。但如果涉及道德水平、审美素养等高维度变量时,测量会有一定的困难,传统的纸笔测验和自陈式量表已经无法完全代表学生的真实水平。因此,对于道德水平、审美素养、身心健康、劳动能力等方面的测量,需要更加关注学生多模态数据下的表现,并辅以表现性评价、质性评价等方面的结果作为参考。

此外,如果涉及教师或学校评价,也要考虑到学生、教师、学校等多个层面的数据。在收集数据的过程中,需要以学生、教师、学校三个层次分别规划变量集,如学生层的家庭经济水平、教师层的教龄、学校层的乡镇或城市区位等。这些都有可能影响最终增值结果的变量,需要分层处理,并从学生层面的结果层层推进至学校层面。通过数据分层可以搭建“学生发展—教师效能—学校办学”的评价证据链,以学生发展为中心,兼顾教师效能,在控制学校办学规模、投入资金等非教育教学变量的差异后,估计学校各方面办学条件对学生和教师发展所产生的影响。通过收集搭建系统的增值评价数据集,可以有效证明名校不一定适合所有学生,普通学校也不一定不适合所有学生。以评价教师教学效能为例,教学效能存在效果、效益和效率三个层面。数据集需要反映教师教学的投入和产出的关系。利用增值评价评估教师效能的实操重点是通过测量学生的发展,并结合系统的教师教学评价指标,估计教师对学生学习的影响。在应用过程中尽可能全面地收集教师的教学行为和个体特征变量,并利用学生的学业增值情况反映教师教学效能,以对教师教学开展增值评价。

2.妥善选用和改进增值评价模型。一方面,结合评价需要妥善选择增值评价技术模型。如果学生体量较大,可以选择使用成长百分比模型,在区域范围内开展增值比较;如果学生体量较小并且需要开展相应的学习诊断,则使用可以纳入协变量的回归模型。在选择模型后需对不同评价内容作假设分析,部分高维变量可能并不符合所选增值模型的统计假设。比如道德、身心健康的发展不一定是线性增长的,其可能是有规律波动向上的非线性样态,这就需要结合学生评价的内容全面考虑选取适合的模

型。对教师教学效能和学校办学的评价建议使用分层模型。不过分层模型的统计假设较为复杂,实践中能够在多大程度上满足前提假设,精准判别教师及其他影响因素对学生学习的影响,这是长期以来增值评价方法的争议所在。因此,在开展增值评价时需要结合评价项目的特征,综合评估教育教学活动有关的变量对学生学业发展的效果、效益和效率所产生的影响。这个过程中对一些高维度变量的测算需要使用符合数据特征的模型,各类模型的统计假设应得到充分检验。

另一方面,基于增值评价技术模型自限性问题,从减少统计假设的限制和提升模型的可解释性两个方面提出改进方向。通过减少统计假设的限制可以使模型更广泛地应用到不同的增值评价场景中,进而避免模型选择带来的不利影响。为此可以考虑进一步研究非线性模型和智能化模型在增值评价中的应用。相较于成绩作为因变量的传统线性模型来讲,核心素养等方面的增值建模或涉及非线性、潜变量测量、数据类别复杂等问题。解决这些问题的关键在于扩大模型的包容性和适应性,一种具有大规模并行处理、分布式信息储存、自组织、自学习等特点的智能化增值模型可能是未来增值评价研究的一个有希望的突破口。此外,提高模型的可解释性是指增进评价的利益相关者对模型内部运算机制与参数的理解,以提高测算结果的可信性。技术模型需要确保纳入运算的变量是有意义的。为了支撑不同的变量类型可以纳入模型中,需要对模型的入口结构做相应的调整,放宽对数据类型的限制,甚至减少标准化程序,以适应未来变量多样化的特征。

3.综合运用增值评价结果。针对增值评价结果难以理解和不够精准的问题,在此提出综合运用增值评价结果的优化策略。《方案》中“四个评价”并不是相互独立的,结果评价、过程评价、增值评价、综合评价四者既是一个整体又相互联系。首先,增值评价的主要内容之一来自结果评价,也就是每一次学生在独立测试中获得的评价结果。从根本上讲,增值评价就是以纵向追踪的形式使用结果评价的信息。无论增值评价的内容如何,它都需要基于结果评价来判断增值,也就是增值比较的基准是由结果评价中得来。其次,过程评价是运用增值评价结果改进教育教学的载体,因为过程

评价用于描述学生和教师的教与学过程,增值评价也是通过捕捉一段时间内的教育过程来开展诊断并为改进提供证据。最后,增值评价结果需要搭配其他评价方式的结果,也就是增值评价、结果评价以及过程评价等评价方式均要回归到综合评价的体系中。增值评价最好与其他教学评价结果结合使用,如来自课堂观察和专家评价的信息。这是使增值评价可以真正参与到“四个评价”的关键。因此,应将增值评价融入“四个评价”的体系中,不应以单一的增值评价结果作为主导。综合运用增值评价结果促使“四个评价”形成整体合力来推动我国教育评价改革,这或将是未来完善增值评价的

重点之一。

最后需要强调的是定期开展对教师和教育管理者的专业技术培训是破解增值评价实践困境的关键。无论增值评价的数据如何完备、技术模型多么先进、评价内容是何,只有当教育工作者全面深入地了解增值评价的理念、技术、应用等方面的知识后,增值评价才能更为科学有效地落实,且真正成为一种可为的评价方式,以助力我国新时代教育评价的改革和发展。

(摘编自《中国教育学刊》2024年第9期,原文题为《增值评价的基本理念、实践困境与优化策略》)

中国教育评价改革的发展趋势

苏启敏、陶燕琴(广州大学)

1.评价目的从选拔人才转向人的全面发展

从评价目的来看,改革逐渐从“选拔部分优秀学生”转向“促进全体学生全面发展”。20世纪90年代中后期,改革开始与素质教育接轨,试图改变应试教育一切为分数、过于重视智育以及片面追求升学率的做法。注重所有人全面发展是素质教育的核心。相应地,改革重心亦逐渐从以选拔智力较高或某方面天赋较好的人才为目的逐渐转向以发展具备良好综合素质或多方面核心素养的人为旨归,突出学生德、智、体、美、劳全面发展。

这种趋势表现为弱化以选拔人才为目的的升学考试比重。选拔性考试政策有所减少,关注学生发展质量的政策有所增加。随着人的全面发展目的日渐受到重视,改革严格区分选拔性考试和合格性考试,并缩小了选拔性考试的范围。

以人的全面发展为目的,评价是否有助于该目的达成,成为衡量评价改革成败的关键标准。基于该标准推动的改革,侧重通过对学生发展状况的教育质量监测与评价,了解学生五育发展达标情况,确保人的全面发展目的得以实现。

2.评价内容从知识技能为本转向核心素养为本

第一,不同学科课程确立了各自的学科核心素养发展目标。以学科核心素养为导向选取学习/考试内容、情境和任务,以及衡量三者关系的适切性,从而更具体地考查学科核心素养发展目标达成度。同时,高考亦重点考核学科

核心素养发展目标达成度。

第二,涵盖五育的通用核心素养发展目标日益成为评价重点。2021年《教育部关于印发〈普通高中学校办学质量评价指南〉的通知》将涵盖五育的学生核心素养概括为:“包括品德发展、学业发展、身心健康、艺术素养和劳动实践等5项关键指标,旨在考查学生德智体美劳全面培养全面发展情况……培养学生适应终身发展和社会发展需要的正确价值观、必备品格和关键能力。”

从评价内容改革的发展趋势来看,改革初期只注重学科课程内的评价内容(学业成就)。改革中期在继续重视学科课程内的评价内容基础上,初步提出把综合素质作为评价内容的设想。改革后期拓展了评价内容的内涵与外延,将改革中期的学业成就与综合素质二元评价内容逐步转变为基于核心素养且覆盖五育的评价内容。整个过程恰好契合“健全综合评价”的改革思路。

3.评价形式从偏重结果增值转向过程与结果增值并重

增值评价作为一种动态评价理念,使结果评价从一次性评价逐步转变为连续性评价,符合改革精神。

运用增值评价改进结果评价的改革探索主要聚焦于考试结果或其他量化评价结果的增值。早期考试结果增值主要体现在分数或者位次的纵向比较,而在新近的改革当中,基于中考、高考等考试结果综合应用多种增值评价模

型的做法逐渐成为一种趋势。除考试结果外，教育质量监测领域也试图对较难测量的学生能力或品质，实施量化增值评价。

然而，单纯探索评价结果的增值评价，虽然改进了结果评价，但依然以评价结果论成败，难免重回“唯分数论”“唯结果论”窠臼。因此，不仅需要探索基于结果的增值评价，还需要探索基于过程的增值评价。最终通过增值评价把“改进结果评价”和“强化过程评价”有机结合起来。探索基于过程的增值评价实质上是把评价重心从结果质量增值拓展到过程质量增值。过程质量增值评价能有效评价影响学生学习过程诸影响因素的真实状况，了解学生在学习过程中心理健康、学习动力、学习习惯、学业负担等方面存在的问题及发展趋势，并加以诊断和改进。

4. 评价手段从人工操作转向技术赋能

数字技术的时空优势推动教育评价由单一横截面评价转变为全时空立体评价。

评价手段数字化使教育评价得以兼顾过程与结果评价。改革初期为减少误差提高评分科学性，标准化考试通过机器阅卷改进主观题评价方式，是改革主旋律之一。新的改革开始提倡纸笔考试使用网上评卷系统，智能测评技术在大规模教育评价中得到广泛应用。尽管技术支持的结果评价其精准性得到极大提升，但仍存在局限，因为它缺乏对过程评价的支持。未来改革将进一步强调运用多种数字技术手段监测

学生学习过程与结果，并实现结果反馈可视化。

数字技术带来的数据优势促使教育评价由主观评价、局部评价转变为精准评价和整体评价。改革初期和中期受限于技术，在评价学生学习情况或内隐能力、品质时难免局限于主观评价或局部评价。随着数字技术在教育评价域的推广，在人工智能等技术加持下，借助对表现特征（如语言、神情、动作）加工所得数据信息，教师将能更准确、全面地评价学生。数字技术亦推动教育评价从智育为重转向五育并举。长期以来受重智取向影响，改革在强调数字化评价的同时首先注重智育的数字化评价，如利用数字网络优化考试形式。改革亦日益关注数字化综合评价。2022年《教育部办公厅关于开展信息技术支撑学生综合素质评价试点工作的通知》提出，“用5年左右的时间，形成百万级规模中小学生学习综合素质发展基础数据库，创新评价工具，利用人工智能、大数据等现代信息技术，……主要包括思想品德、学业水平、身心健康、艺术素养、劳动与社会实践等五个方面，形成数据驱动的学生综合素质评价解决方案”。可见，数据驱动将使学生综合素质评价（核心素养评价）实现从有限维度延时评价到全方位即时评价的飞跃。

（摘编自《福建师范大学学报（哲学社会科学版）》，2024年第3期，原文题为《中国教育评价改革的演变、取向与趋势——基于国家政策文本的分析》）

教育评价领域的新方法与新趋向

熊新惠（美国教育考试服务中心）

一、教育评价领域的新方法

1. 基于项目的档案袋评价。在该方法中，学生通过整理自己的作品集合，可以展示他们在各种项目和作业中的学习历程和成果。档案袋可以提供学生学习过程的整体视图，使教育工作者不仅可以评价学生知识掌握情况，还可以评价他们的批判性思维、创造力和沟通技巧等内容。通过参与实际项目，学生既能加深对概念及其实际应用的理解，也能增强真实的学习体验。此外，该评价方法鼓励反思和自我评价，有助于学生掌握自身学习情况并设定个人学习目标。

2. 同伴互评与反馈。指学生之间互相评价作业完成情况，并提供建设性反馈意见。这种方法不仅可以减轻教师负担，而且可以促进学

生更深入地学习和提高批判性思维能力。要想顺利完成对同伴作业的评价任务，学生不仅要更深入地理解课程内容，还必须能运用所学概念进行客观评价。此外，由于同伴反馈比教师反馈更容易被接受，因此也提高了学生的参与度和改进动力。同伴互评还有助于培养学生的沟通能力、合作能力及同理心等重要技能，这些技能对他们日后的学业和职业顺利发展非常重要。

3. 游戏化评价。是将游戏设计元素整合进评价过程，从而增强学生的参与度、学习动机和成果产出。通过竞争、奖励和进展等原则，游戏化评价将以往枯燥乏味的评价任务转变为沉浸式、互动式的体验，以一种有趣、交互式的方式激励和评价学生。学生通常会面临一系

列挑战、任务或情境,这些设计旨在以有趣和引人入胜的方式评价他们的知识和技能。

在教育环境中,游戏化评价具有促进深层学习和提高知识内容保持的潜力。首先,可以提升学生的参与度和学习热情;其次,游戏化评价提供的形成性反馈和适应性学习体验的机会,可以帮助教育工作者根据学生的个性化需求量身定制教学策略;最后,游戏化评价的互动性还可以促进学生合作和社交互动,进而推动同伴学习和知识分享。

4.虚拟现实评价。通过模拟具有专业背景的真实生活场景或环境,提供具有实践性、体验性的沉浸式学习机会,从而有助于学生将理论知识应用于实际情境。模拟有多种形式,如虚拟模拟、角色扮演练习或互动案例研究等。通过让学生参与真实挑战,有助于培养他们的问题解决能力、批判性思维及与学习领域或未来职业相关的决策能力。通过提供即时反馈信息,学生可以反思自身行为并做出及时调整,从而提升学习效果。此外,由于参与者需要相互协作解决复杂问题或应对动态情况,因此该方法还有助于锻炼团队合作能力。

在高等教育领域,虚拟现实评价越来越多地被教育、科研机构用来弥合学术学习与专业实践之间的鸿沟。

5.数字徽章和微证书评价。是识别和验证学生在特定技能或领域中取得成就和具备能力的一种新方法。与传统的文凭或证书不同,数字徽章是学生通过各种学习经历(如课程、项目或评价)获得成就的一种数字化呈现。数字徽章包含元数据,能提供有关达到标准的详细信息及提交证据,具有数据透明和可验证性特点。微证书是一种正式认证,通常与行业标准或专业发展目标一致,可以证明学生具备某些特定的技能或能力,一般通过完成一系列相关任务或评价即可获得。

数字徽章和微证书相较于传统证书而言具有多种优势。首先,它们能精细化地认可某项技能或成就。其次,更加灵活和便捷,可以通过在线课程、研讨会或基于项目的评价等多种学习经历获得。最后,可以促进学生持续学习和技能发展。通过提供清晰的途径,该方法帮助学生获取新技能并得到认可,同时还激励个人参与持续的教育和培训。

6.自我评价与反思。是深化学习经验、促进元认知发展的有效工具。在具体评价过程中,学生通过完成结构化的自我评价活动,如完成

评分表、检查表或日记提示等,参与对自身进展、优势和成长领域的评价,进而有助于深入了解自己的学习过程和结果。通过自我评价和反思,学生还可以全面深刻地认识自己的学习风格、偏好及面临的挑战,进而制订有效的自我调节与改进策略。此外,自我反思鼓励学生对自身学习经验、个人与现实世界联系、个人目标等内容进行批判性思考,促进学生对学习内容的深刻理解。

7.多模态评价。采用多种形式的媒介并将各种表达形式纳入评价过程,承认学生具有多样化的学习风格和个人专长,允许他们通过文本、图像、音频、视频、交互元素等多种形式展现个人能力。多模态评价提供了一种更具包容性和公平性的评价方式,可以满足不同学习偏好、不同学习能力人群的差异化要求。多模态评价鼓励学生发展批判性思维和媒体素养,因为他们必须批判性地评估不同交流方式的有效性及其对意义建构的影响。此外,多模态评价对创造力和创新发展也具有推动作用。

8.自适应评价。代表传统测试方法的范式转变,主要表现为可以根据学生的不同表现动态调整测评难度。自适应评价可以通过算法根据每位学生的不同能力定制差异化的测试体验,以确保学生接受的挑战水平适当且有吸引力,从而更准确地评价他们的知识和技能水平。具体而言,随着学生展示能力越来越强,评价系统提出的问题也越来越具有挑战性,但可以在学生感到困难时主动降低难度,从而促进学生的持续学习和成长,允许他们以自己的步调前进。此外,评价结果还可以向教育工作者提供关于学生优势和改进建议的信息,有助于调整教学策略并提供更有针对性的教学。

目前,自适应评价在中小学教育、高等教育和专业认证项目等领域日益流行。在中小学教育中,自适应评价可以识别学生个体的学习需求并提供有针对性的干预措施,教育工作者可以根据评价结果提供差异化教学和个性化学习体验。在高等教育和专业认证项目领域,自适应评价可以对知识和能力进行更为高效和精准的测评,有助于学习者根据个人特点制订学习计划。

9.开放式探究项目评价。是一种以学生为中心的评价,有助于培养学生的批判性思维、问题解决能力和独立探究能力。在该项目中,学生要对开放式问题进行研究、收集证据并得出结论,尤其是可以自主选择探究范围和研究

方法,并以体现自身学习风格的方式展示研究结果。与传统的有预设答案的评价不同,开放式探究项目允许学生持有多样化观点并鼓励创造性探索,因此有助于学生深入研究感兴趣的主体项目。

开放式探究项目有助于促进学生的高阶思维技能发展和元认知意识培养。通过处理复杂问题和面对不确定性情景,学生学会批判性地分析信息、评价来源,并将探究发现整合为系统论证或解决方案。此外,开放式探究项目需要学生联合解决复杂问题并与同伴分享探究发现,有助于发展学生的团队合作能力和沟通技能。

10.基于社区的评价。将现实世界背景和社区实践参与纳入学生的评价过程,增强了评价的相关性和实际应用性。在该评价中,学生要与本地的组织、企业或社区成员合作,解决真实问题。通过与课堂外的利益相关者直接合作,学生可以获得实践经验,并为有益于社区的重要项目作出贡献。评价强调在实际生活情境中应用知识和技能,以促进更深层的理解并激发学习动机。

在高等教育领域,主要用于评价学生的学习成果和促进公民参与。此外,该评价方式为大学提供了与周边社区建立有意义合作伙伴关系的机会,增强了机构在校园外的影响力和关联性。

上述十种新方法虽然各有特点,但并非独立运行,而是相互补充,从而实现更全面、更有效的教育评价。在选择评价方法时,结合具体的目标和目的选择最优组合,为学生提供有意义的成长和成功机会。

二、教育测评技术的新趋向

1.个性化学习和评价。应用先进技术和数据分析,根据学生需求量身定制教育方案,从而提高学习的参与度和效果。由于该评价方式可以根据学生作答情况动态调整问题难度,因此有助于教育工作者给学生提供即时反馈和有针对性的支持。

个性化评价的关键要素之一是开展形成性评价,以持续监测学生学习进展并提供教学调整的数据支持。形成性评价通常被嵌入学习过程,以便教育工作者实时收集学生对知识和技能的理解情况。例如,数字平台可以跟踪学生的学习情况,提供可能需要额外练习或支持的数据信息。这种持续的评价使教育工作者可以给每位学生定制个性化的学习路径,以确保学生获得与自己特征相匹配的任务和支持。

2.基于能力的教育(CBE)评价。以学习者为中心,强调学生以个人节奏学习特定的能力或技能,而不是局限在传统的课堂时间内。与传统的评分系统不同,CBE评价通常依赖固定间隔的总结性评价,持续进行并允许学生根据熟练程度自主确定进度。

首先,CBE评价的关键要素之一是强调形成性评价实践。与大规模、高风险考试不同,形成性评价为学生提供定期反馈,帮助他们发现问题并引导整个学习过程。这种持续性反馈不仅有助于学生保持学习积极性,也有助于他们了解自身学习进展并随时调整学习策略。其次,CBE评价强调过程公开透明且与能力保持一致。测评开始时,学生收到明确的学习目标和评分标准,向他们描述了应该掌握的具体能力及需要达到的熟练程度。这种做法有助于培养学生的责任感和自主性。最后,CBE评价要求学生在有意义的情境中应用知识和技能完成任务,尤其是完成现实世界中的真实任务。评价可以采取项目探究、模拟、案例研究或绩效任务等形式,有助于学生真实、恰当地展示个人能力。

3.技术整合评价。主要包括数字工具和自适应测试,既是测量和支持学生学习的新方法,也代表了评价实践领域的重大进步,还为教育工作者提供了评价学生进步和知识点掌握情况的更为高效的方式。技术整合评价中的一个关键要素是数字化评价,包括在线测验、交互式模拟和基于计算机的测试。数字化评价不仅可以提供即时反馈、进行自动评分,还能根据学生需求和学习目标选择个性化的评价题目。

技术的高速发展和全面整合,使自适应评价系统得以广泛应用。技术整合还允许评价中使用视频、音频、交互式图形等多媒体元素,即拓展为多模态评价。这些评价不仅测评内容知识,还测评包括媒体素养和数字沟通技能在内的多种能力,从而有助于更全面地了解学生的学习情况。

教育评价中的技术整合有助于收集和分析大量有关学生表现的数据。教育数据分析工具可以帮助教育工作者识别学生的学习模式和趋势,从而做出更具针对性的教学决策和干预。然而,在将技术整合到评价实践中时,必须考虑公平和可获得性问题,因为技术获取和数字素养技能的差异可能加剧现有的成绩差距。

4.形成性评价。与终结性评价不同,形成性评价贯穿整个学习过程,常用技术包括小测

验、互动投票、快速写作等,有助于教育工作者识别学生的优势和薄弱领域,并及时开展干预和支持。通过持续监测学生进展,形成性评价可以创造一个响应式的学习环境,并能根据学生的不同需求进行调整。

在个性化学习环境中,利用形成性评价结果生成个性化的学习指导策略非常重要。通过一系列评价,教育工作者可以根据学生的学习风格 and 不同需求调整教学策略。

形成性评价的基本特征之一是强调持续改进。通过关注学习过程而不是最终结果,形成性评价鼓励学生反思学习过程并设定改进目标,使学生看到努力的价值。这不仅有助于激发他们的内在学习动机,还有利于形成良好的课堂文化。在这种课堂文化中,错误被视为学习的自然组成部分,因此极大地减轻了学生焦虑,也培养了他们应对困难和挑战时的韧性。该方法还能为教师提供教学改进建议。

5.21 世纪技能评价。强调批判性思维、创造力、协作和沟通,为学生应对未来挑战做好准备。传统评价通常优先考虑基本记忆技能,但这些方法无法捕捉批判性思维、创造力、合作和沟通等关键能力。为解决该问题,教育工作者和政策制定者正在开发新的评价工具,用于评价重要的软技能和复杂问题解决能力。

创建基于表现的评价是一种有效方法,要求学生参与现实世界的任务和项目。这些评价允许教育工作者观察和测量学生在实际场景中应用知识的能力,反映他们在职业生涯中可能面临的挑战。例如,学生可能被要求组成团队、制订商业计划、解决社区问题或进行科学实验,这些任务不仅评价学生的学科知识,还评价他们的批判性思维、有效沟通和合作能力。通过将元素整合到评价中,教育工作者可以更全面地了解学生能力。此外,数字工具和模拟技术的发展为 21 世纪技能评价提供了新的途径。这些技术可以用于创建沉浸式环境,需要学生实时解决复杂问题并做出决策。

6.真实性评价。将学习与现实任务相结合,使教育的各个要素更加有关联和有意义,也代表从传统测试到反映真实世界挑战和应用的的评价范式的转变。与通常侧重机械记忆和孤立技能的传统评价不同,真实性评价旨在反映现实任务的复杂性和综合性。这种方法强调知识和技能的实际应用,提供学生在真实场景中表现能力的更准确的测量。例如,学生不是接受关

于历史事实的多项选择测试,而是被要求分析相关资料并推断历史事件影响,从而展示他们的批判性思维、研究能力、沟通能力等。

真实性评价的重要组成部分之一是对项目、演示和基于绩效任务的应用。这些类型的评价要求学生深度领悟所学内容,以有意义的方式应用他们对知识的理解,并产生可以公开展示和批判的作品。真实性评价的另一个标志性特征是强调实际应用。通过让学生参与实际情境中的任务,教育工作者可以更好地评价他们在应对职场要求方面的准备情况。

7.教育评价中的社会文化敏感性和公平性。设计具有文化包容性和公平性的评价工具,需要研究者深入理解学生群体的多样化背景、语言和学习风格等特征。这包括创建的评价工具能反映广泛而多样的文化背景,确保没有任何群体被不公平对待。

在评价工具和评价过程中减少偏见是实现公平性的关键。为了减少偏见,测试设计者必须严格审查任务并在多样化学生群体中进行试测,以识别和消除偏见内容。

8.教育评价中的数据隐私与安全。对学生而言,实施强大的数据保护对于保护敏感信息免受未经授权的访问和泄露日益迫切。这需要采用先进的加密技术、安全的身份验证过程和定期的安全检查,以保护数据的完整性和机密性。学校和教育机构必须制定全面的数据治理政策,明确数据收集、存储和访问等程序,以确保只有获得授权的人员才可以接触和处理学生信息。数字化评价中一个主要关注点是保护学生信息,包括测试成绩、作答模式和可能的个人识别信息等,这些都必须防止被滥用。同时,教育数字化平台必须遵循隐私保护设计原则,并将其贯穿于评价开发、实施的各个阶段,包括在必要时对数据进行匿名化处理、使用安全的数据传输方法、实施严格的数据访问控制以阻止未经授权的访问等。

遵守数据安全法律和道德标准是确保教育评价中数据隐私的另一项关键内容。这不仅有助于保护学生信息,也有助于学生、家长和教育工作者等主体建立对敏感信息处理过程的信任。此外,教育机构还必须确保数据在收集、使用和实践等流程中尊重学生隐私。

9.全球和跨文化能力评价。旨在评价学生的跨文化能力,包括对文化差异的意识和敏感性,在多样化环境中工作的能力,以及有效进

行跨文化交流与合作的能力等。该评价涉及的场景或任务通常需要学生展示在不同文化背景中对文化规范、价值观和沟通风格的理解,以帮助学生适应全球互联的社会发展要求。评价内容包括尊重文化多样性、文化适应能力,以及有效的跨文化沟通技能等。

在评价任务中融入多元文化视角,不仅可以丰富学生的学习体验,还可以为学生提供探索不同文化观点的机会。评价任务可能要求学生比较和对比文化实践、信仰和世界观,展示他们能够欣赏文化多样性并认识跨文化互动的复杂性。

全球意识和跨文化沟通技能的评价工具包括自我评价调查、同伴评价,以及衡量学生知识、态度和跨文化能力的表现性评价。

10.终身学习的成果认证。未来,传统的学术学位将不再是唯一评价标准,教育评价将越来越集中于通过创新方法认可和验证终身学习

成果,承认通过在线课程、研讨会、职业发展项目、实际经验等多种途径获得的知识和技能。这种转变以尊重个体多样化的学习经历为前提,反映整个社会对于终身学习价值的广泛认同和对更包容的评价实践的需要。

随着教育形势的不断发展,用于技能认证的微证书和技能徽章的发展急剧增加,为通过非正式和非传统学习方法获得的特定技能和能力提供灵活、精细化的认可和认定途径。

此外,认证终身学习的评价方法超越传统的大规模考试和标准化测验,主要评价在实际环境中知识和技能的应用,包括作品集评价、体验性学习项目、基于能力的评价以及反思性文章或面试。通过整合多样化的评价方法,教育工作者能够更准确地测量学习者的能力,包括批判性思维、问题解决能力、创造力和协作能力等。(摘编自《中国考试》2024年第10期)

美国最新科学教育评估中试题情境的特征及功能旨趣

荆鹏(东北师范大学,美国俄亥俄州立大学)

情境是科学教育评价环节实现“价值引领、素养导向、能力为重、知识为基”综合考查的关键载体。美国国家教育进步评价(NAEP)作为美国唯一具有全国代表性的教育持续评估项目,于2024年1月对NAEP科学评估框架予以结构性修订,并将其命名为“2028 NAEP科学评估框架”(简称NAEP 2028),给出了未来一段时间美国科学教育评价的方案流程和行动指南。该框架将试题情境作为NAEP 2028中“NAEP科学学科概念”“NAEP科学与工程实践”“NAEP科学跨学科概念”三大测评维度考查的核心载体,并给出了高质量情境的设计理路和典型范例。

一、NAEP 2028中试题情境的基本特征

1.现实关联:基于具体实例的真实性。NAEP 2028强调科学评估应通过具体、真实的情境评估学生的科学成就,而非仅考查抽象的理论知识;注重考查学生对科学知识的可迁移性,即需要融入现实世界的现象或实验这类生动具体的科学实例,使试题对学生具有吸引力和激励性,以考查学生所学知识与实际情境之间是否构建了有意义的联结。但值得注意的是,NAEP 2028所强调的真实性的具体实例情境的本质特征是“真实性”而不是“真实”,它是

具有人为改造的指向学生科学学习测评目标的现实实例。具体可从三个维度理解。一是评估问题情境需要基于学生日常生活中可能遇到的真实科学现象或问题。二是评估情境尤为关注当前热点科技话题或重大科学事件等社会性科学议题。三是模拟科学家、工程师等专业人士在工作中可能面临的实际问题或任务情境。

2.进阶贯通:蕴含不确定的非良构性。评估试题中情境不能漫无目的地铺开,它在包含着激发认知冲突的同时蕴含着进阶连贯的构造理念。所以,NAEP 2028要求试题情境设置必须由内在的脉络贯通,让学生在经历完整的问题解决过程中体悟解决问题的价值蕴含。试题情境应呈现“真实的不确定性”,即情境需蕴含一定挑战性的真实性任务,这些任务的不确定性源于科学发现历程中无法预知的不确定因素,进而可以衡量学生的批判性思维、迁移能力等高阶品质。非良构问题情境是那些具有真实不确定性、进阶性和复杂性的评估情境。在NAEP 2028中,引入了非良构试题情境设计理念旨在模拟真实世界中的复杂问题,以全面测量学生在充满不确定和动态的条件下科学素养的综合表现。这些非良构试题情境通常呈现真实的不确定性、进阶性和复杂性,它们衍生的

问题任务往往没有唯一的正确答案或明确的解决路径,具体表现为三个特征。一是**真实不确定性的引入**。鼓励学生运用批判性思维,考虑多种可能性,而非简单寻求唯一“划一性”标准答案。二是**多维度科学实践的整合**。使用“真实的问题或其他提示”,映现科学实践与研究的综合性和跨学科性,引导学生使用“三维科学学习”来解释现象或找出问题的解决方案。三是**进阶性和复杂性的蕴含**。评估将完全采用数字化工具进行,以多模态方式呈现试题情境,同时,通过多元化的试题项目呈现进阶性和复杂性的问题情境。可以说,NAEP 2028 试题情境的进阶连贯的非良构性特征蕴含了对于学生对科学现象的解释能力、实验设计与数据分析能力、问题解决能力与批判性思维等高阶素养能力的考查,并持续减少对事实性知识和纯粹数学计算的考查比重的评价理念。

3.重视差异:关注学生群体的异质性。一是**关注文化响应性和情境相关性**。框架强调试题情境应“与特定群体相关”,认为学习是嵌入特定社会文化背景中,通过选择与不同群体相关的情境,可有效提高评估的文化响应性和公平性。二是**进行语言考虑和认知负担平衡**。提出要“减轻那些可能在学校同时学习英语和科学的学生的认知负担”,但也强调不应“过度限制语言参与,以至于无法有意义地评估概念”,这种平衡反映了NAEP 2028 试题情境设计中对可及性和有效性的审慎考虑。三是**采用多模态呈现和通用设计**。照顾不同学习风格和能力的学生,减少评估中的非必要障碍,进而更准确地测量学生的科学综合表现,提高评估的包容性以及针对不同背景学生的公平性。

二、NAEP 2028 中试题情境的功能旨趣

1.秉持整合性逻辑,赋予学生体验解决问题的现实意蕴。NAEP 2028 秉持整合性逻辑将真实性任务、跨学科统整和社会性科学议题融入科学评估试题情境中,让学生在解决复杂情境中的问题时,通过情境这一媒介调用元认知系统与操作性知识,体验知识的现实应用价值,促进其科学素养的全面发展。例如,引入气候变化、能源危机或公共卫生社会性科学议题等试题情境,要求学生分析温室效应的物理机制,探讨不同能源方案的环境影响,或评估声波传

播在城市噪声污染中的作用。通过这些试题情境,学生还了解到科学发展与社会进步的相互作用,可培养其作为社会公民的责任感和参与意识,为应对全球性挑战作好准备。

2.遵循系统性逻辑,提供学生经历问题解决的完整历程。梳理NAEP 2028 试题情境创设的要点和试题样例不难发现,试题情境的系统性逻辑框架明晰,以前置情境为后一情境铺陈,后置情境则是对前置情境的具体化、深入化。情境链条之间环环相扣,设计一系列难度递进的任务情境序列,形成一个有机统合的整体,保障了其进阶连贯性,为学生认知参与情境构建而非对情境的简单“应激反应”提供了一套可行性路径指南,旨在让学生经历解决科学问题或现象时的“情境化”“去情境”“再情境”的完整历程,产生一种完全置身于试题情境中的身心体验,进而成功激活学生的思维动机,多维地实现对学生真实科学素养的全面衡量。在科学测评中,“经历情境”尤其是经历完整的情境,有助于学生认知参与和情感参与,进一步保障了测评学生科学学业成就的信效度。

3.关照协同性逻辑,导引学生树立科学学习的身份认同。身份建构超越了对科学和科学课堂活动的短期态度和情感反应的考量,从自我过程的角度构建对科学的参与和更长远的愿望,塑造了一种自我意识。NAEP 2028 关照协同性逻辑,强调科学学习和评估应是一个社会性、互动的过程,而非孤立的个体活动,为有效地促进学生的科学身份认同,增强其在科学测评和未来科学学习中的参与度,将多元文化视角、协作问题解决、科学话语实践和科学态度培养等重要元素融通到评估试题的情境设置设计中,尤其是将关涉科学资本、包容的科学经验和实践以及科研伦理和价值观等关键要素融入试题情境。除NAEP外,PISA等国际大规模测评也十分关注对学生科学身份认同的测评,PISA 2025 科学素养测评框架中便将科学身份作为与科学能力、科学知识并列的第三大科学素养维度。

(摘编自《上海教育科研》2025年第2期,原文题为《美国最新科学教育评估中试题情境的特征、旨趣及启示》)

【研路同行】

教育考试命题的“四性”及其实现路径

孙建峰（省教育考试院命题中心）

近年来，教育考试命题质量出现“四性”这一提法，即政治性、科学性、公平性、规范性。出现“四性”的文本，既有官方层面的规范性文件，也有社会层面的新闻报道，涵盖多类考试项目。命题质量包括宏观层面的命题管理质量、微观层面的测量评价质量、中观层面的试卷试题质量等多个维度，“四性”属于命题质量中观层面的核心内容，是与试题试卷密切相关的基本质量要求。

“四性”的形成植根于我国教育事业土壤，其形成与发展反映了从测量评价驱动到全面育人驱动的考试理念跃升。1992年《高等教育自学考试命题工作规定》提到试题“不出现政治性、科学性错误”。“两点论”是指政治上不走歪、科学上不出错，为“四性”奠基。教育部2006年《普通高等学校招生全国统一考试分省命题工作暂行管理办法》提到“分省命题工

作，要有利于考试的科学公正、安全高效与准确规范”。这一要求针对当时展开的分省命题实际，重视科学性、规范性，强调精准实现考试目标。2015年新课改高考在全国范围内实行，教育部对高考命题提出“一点四面”改革要求，“以立德树人为核心，加强社会主义核心价值观、依法治国、中华优秀传统文化和创新能力等内容的考查”，深化对立德树人的认识，凸显命题育人导向，发挥考试育人价值。此后，“四性”概念逐步清晰，提法相对固定，目前已成为业内共同遵循的命题质量标准。

“四性”是命题底线要求，体现系统合力，具有较强适用性。实现“四性”，要聚焦命题规律，坚持正确的政治方向，依据规范性文件，细化技术标准；要聚焦命题管理，建立审查校对、集体研讨、全面查重等管理制度；要聚焦命题人员，提升命题能力，提高综合素养。

国家教育数字化战略背景下自学考试卷库管理系统建设研究

陈智丽（省教育考试院命题中心）

自学考试卷库管理系统是在国家教育数字化战略背景下，为解决自学考试命题规模大、管理环节多、过程性评价匮乏等难题，由江苏省教育考试院自主研制开发的命题全流程数字化管理系统。该系统的顺利上线，为完成新一轮自学考试专业及课程改命题任务提供了坚实支撑。

系统建设目标是实现命题全流程数字化管理，支持各业务环节的功能需求和多种命题模式，解决传统命题模式下的痛点和堵点问题，支持命题教师、审题教师、管理人员间的数据交互，实现试卷修改、查询的全流程留痕和可回溯管理，以及依托系统开展命题质量评价。

卷库管理系统层次模型设计共分为5层，包括用户界面层、业务执行层、功能设计层、数据存储层和基础设施层。网络结构上以B/S结构为基础，在分散命题环节搭配C/S应用，通过建立统一的数据标准，实现对命题业务的

有效支持。体系结构按照项目划分为系统管理、全国统考试卷、社会自考试卷、专接本试卷、助学专业试卷等5个子系统，以确保项目数据安全隔离。此外，作为卷库管理系统的延伸，离线使用的命题辅助工具也包括加密文件导入、试题编辑、试卷组配等功能。

与传统命题模式相比，系统的安全性较好，适配性较为灵活，具备命题全流程留痕功能，满足高并发访问需求，智能化和自动化程度高，支持各类报表的统计和导出等特色功能。系统投用以来，经过省级考试机构、主考学校和命题审题教师的共同努力，系统功能得到不断完善和提升，在推动管理模式转型、提升命题质量、强化安全保密等方面取得良好效果。使用系统后，2024年存在问题试题的试卷初稿比例较2023年下降65%，同时降本增效，年节约排版费近100万元，纸质耗材使用量也大大降低。

【域外资讯】

经合组织发布《提升 PISA 无障碍性：PISA 题目修订与设计指南》

在全球教育评估领域，如何确保所有学生，无论其是否具有特殊教育需求，都能公平地参与并展示其真实能力，已成为一个亟待解决的关键问题。近日，经合组织发布了《提升 PISA 无障碍性：PISA 题目修订与设计指南》工作论文，为 PISA 测评在题目设计、工具开发及实施流程等方面的无障碍改进提供了系统性框架。

一、背景：PISA 无障碍性的历史与挑战

PISA 作为全球最具影响力的国际学生评估项目，每四年对全球数十个国家和地区的 15 岁学生实施阅读、数学和科学能力的评估。然而，由于历史原因和操作复杂性，PISA 在无障碍性方面长期存在不足。传统上，PISA 对参与学生的特殊教育需求考虑有限，仅允许在极少数情况下提供有限的支持措施，如小规模施测、听觉放大器使用等，而明确禁止如放大设备、题目朗读等常见无障碍工具，导致大量有特殊需求的学生被排除在评估之外。

随着全球对教育公平性的关注日益加深，如何确保 PISA 测评能够包容所有学生，尤其是那些具有特殊教育需求的学生，成为了一个迫切需要解决的问题。为此经合组织于 2023 年启动了“PISA 研究、开发与革新（RDI）项目”，旨在通过开发新的无障碍测评形式，提升 PISA 的包容性和公平性。

二、指南内容：四大理论框架与具体设计准则

（一）四大理论框架。指南基于四大核心理论框架，为 PISA 题目的无障碍设计提供了坚实的理论基础。

1.通用设计。强调产品设计应尽可能满足最广泛用户群体的需求，确保所有人都能平等使用。在 PISA 测评中，这意味着题目和测试平台的设计应考虑到不同学生的能力差异，提供多样化的呈现和交互方式。

2.认知负荷理论。关注信息处理过程中工作记忆的负担。指南建议通过简化题目表述、减少非必要信息、优化视觉呈现等方式，降低学生的认知负荷，确保他们能够将认知资源集中在评估目标上。

3.无障碍理论。将测试者的技能分为物理、感知、接受、情感和认知五个维度，并匹配相应的测试特征进行调整。

4.网页内容无障碍指南。为网络内容无障碍提供了详细的技术标准。通过借鉴了网页内容无障碍指南的原则，确保 PISA 测评平台在感知性、操作性、可理解性方面达到国际无障碍标准。

（二）具体设计准则。指南详细列出了从题目开发流程到具体内容设计的多项准则，包括：

1.开发团队构成。建议包含无障碍和残疾专家，确保从设计初期就充分考虑无障碍需求。

2.题目审查清单。开发了详细的题目无障碍审查清单，涵盖文本、视觉元素和布局等多个方面，确保每个题目都经过系统性评估。

3.文本优化。建议使用简洁明了的语言，避免复杂句式和冗余信息；关键信息加粗显示，减少非必要文本；多选题选项不超过三个，以降低认知负荷。

4.视觉改造。确保视觉元素清晰、无干扰，使用高对比度颜色组合；重要信息通过箭头、高亮或标签等方式突出显示；为所有视觉元素提供替代文本描述。

5.布局重构。尽量将题目所有必要信息呈现在一个屏幕上，减少滚动需求；视觉元素与文本紧密集成，避免分离；确保鼠标指针和导航元素清晰可见，提高操作便利性。

三、实施案例：PISA 无障碍科学测评试点研究

1.试点研究概况。作为指南应用的具体案例，经合组织在七个国家（巴西、加拿大、哥伦比亚、丹麦、法国、葡萄牙和瑞典）开展了 PISA 无障碍科学测评的试点研究。

研究聚焦于具有中度视力障碍、阅读障碍以及注意力不集中和行为控制困难的学生群体，通过提供读屏软件、视觉调整工具（如对比度、字体大小调整）及灵活的测试安排（包括休息时间），评估这些适应性安排的有效性。

2.试点结果与分析。试点结果显示，适应性测评形式显著提升了目标学生群体的参与度。超过 80% 的学生表现出高度或中度的参与度，其中读屏软件 and 对比度调整工具最受欢迎。学生普遍反映这些工具“非常有用”，有效帮助他们克服了阅读和理解题目的障碍。同时，灵活的测试安排也获得了学生的认可，有助于他们更好地管理测试过程中的疲劳和注

意力分散问题。

3.反馈与改进建议。尽管试点研究取得了积极成果,但学生和观察员也提出了一些改进建议。例如,提升读屏软件的语音质量、增加更多对比度选项、引入高亮工具和语音转文字工具等。此外,学生还建议为视觉元素中的文本提供读屏软件支持,并增加缩放特定视觉元素(如图表)的功能。

(“上师大国际与比较教育研究院”公众号)

UNESCO 发布《全球教育监测报告:引领教育技术发展东亚区域版》

2025年8月18日,联合国教科文组织(UNESCO)发布《2024/5全球教育监测报告:引领教育技术发展东亚区域版》。该区域版报告由全球教育监测报告组(GEM Report)与北京师范大学、日本上智大学及韩国教育发展研究所联合编制,聚焦教育领导力在推动东亚地区教育数字化转型的作用。东亚在推动教育技术改革方面走在全球前列,报告记录了中、日、韩三国在培养地方教育行政人员、学校校长及教师方面的主要举措,以确保这些关键角色能够有效且公平地推动改革落地。

区域报告指出,教育的数字化转型需要能够在不确定环境中适应并管理系统性变革的领导者。他们肩负着实现雄心勃勃的数字改革的责任,同时还需应对日益增长的相关工作任务,从制定指导方针、支持专业发展,到战略性资源分配、基础设施维护,以及实施过程的监督等。

校长是教育数字化转型的“守门人”。他们在学校社区中构建安全、包容的数字文化方面发挥着关键作用,这一角色正日益受到各国数字战略和行动计划的重视。

报告借鉴东亚地区的经验,呼吁各国在推动数字教育改革的过程中,应进一步重视并支持教育领导者。这一支持应通过有效的选拔程序、持续的专业发展机会及配套支持措施来实现,并在相关的教育发展规划、政策文件和实施指南中明确体现,从而助力教育系统平稳迈入数字时代。

报告通过对中国、日本和韩国教育领导者在数字化转型中的实践经验的案例研究,发现以下主要结论。

1.教育数字化转型在东亚地区加速推进。中国超过一半的学校已配备独立数字化管理平台。日本因新冠疫情影响,将“一人一设备”目标的实现时间从2023年提前至2020年。

人工智能(AI)在教育中的应用日益推广。韩国计划于2028年前全面推行AI教科书计划;中国则计划在2023年前在所有中小学普及人工智能教育。

2.学校领导层在教育数字化转型中的作用日益凸显。校长可促进校内数字文化发展、组织教职工能力建设,并根据学校需求选择资源。其职责正日益复杂,涵盖采购、网络安全、教学方法及学生福祉等多个领域。

这一趋势已被纳入新的数字化战略。在中国,校长被赋予“首席信息官”职责;在日本,2023年的教育数字化转型战略明确将校长定位为数字创新的核心推动者。

3.学校教职工正被赋予新职责以支持课堂数字化转型。在中国,《2025年智慧教育白皮书》提出赋予教师“新角色与使命”。在日本,核心人物如“GIGA领导者”正在推动学校的数字化创新。在韩国,每所学校任命2至3名骨干教师,制定符合本校情况的实施策略。

4.地方、学校及教职工需要政策协同、指导和支持。全国性的学习管理和评估系统至关重要。中国已将32个省级学习管理系统整合至统一的国家智慧教育平台,按教育阶段与类型分类管理。

制定技术应用指南与标准,有助于确保数字化安全有效落地。日本最近发布了关于在学校使用生成式AI的指南。

需要配备专业人员协助学校进行数字化转型。在韩国,“数字导师”负责管理学校技术基础设施的维护。

5.领导者需促进不同治理层级及校内的协作。学校领导层之间的非正式工作组有助于他们分享经验。一项针对中国校长的调查显示,他们最偏好的培训形式是同行交流。

地方教育行政部门可促进跨地区协作。中国青海省与上海、江苏、浙江等数字化发展较为先进的地区建立了合作关系。

通过协商制定更契合实际的数字教育政策。日本与韩国在制定数字化战略时均征询了教师意见。

6.需要开展领导力培训以应对数字化转型。三国均要求新任校长接受培训,并更新教师能力框架以纳入人工智能应用能力。

系统层面的领导力培训同样重要。中国与韩国均为未来管理者开设了信息化课程;在日本,对815个市镇的调查显示,谈判与协调能力是落实数字化战略的关键技能。

报告最后结合三国案例研究的证据,提出相应的政策建议,旨在明确系统领导者和学校引领教育技术发展可采取的具体步骤。

报告认为,教育数字技术的应用形式多样且日益融合。从硬件设备到人工智能等一系列创新,正深刻影响教育管理者的管理方式、教师的教学方法以及学生的学习与互动模式。技术变革的步伐持续加快:物理世界与虚拟世界的界限日益模糊;人机互联程度不断提升;数据存储与处理能力呈指数级增长。保障学生的安全与身心健康成为首要任务,同时确保所有学习者在数字环境中能够负责任地参与与行为亦至关重要。

数字技术与其他技术一样,是人类为满足需求、改善生活而创造的成果。教育系统在不断扩大对数字技术的依赖时,应始终将学生的利益置于核心位置。数字技术不应取代以人为本的教育,而应成为其有力支撑。

东亚是全球教育数字化的热点地区。尽管政策导向鲜明,但中央政府、地方政府、学校及其社区等各方仍需携手合作,推动有意义且符合伦理的数字化转型。

教育系统领导者、学校领导者及教师领导者在当前大规模改革中发挥着关键作用,但其重要性可能尚未得到充分关注。这些领导者不仅是推动改革在学校和课堂层面有效落实的桥梁,也是汇集并反馈改革过程中所遇瓶颈与问题的重要纽带。

报告提出的具体建议如下:

建议1:国家政策框架需保持清晰一致。领导者应对技术改革及其目标有清晰的认识,能够运用互操作性强且稳健的监测系统,并获得充足的资源支持。

建议2:学校领导需获得清晰指导与配套

支持。明确的指导方针与标准有助于领导者理解并推动学校数字化转型的协同性。清晰的职责划分、充分的支持以及适度的自主权,有助于提升学校领导者的影响力。

建议3:系统及学校领导应参与教育技术决策。应广泛征求并充分反映贴近教学一线人员的意见,确保改革方案在教学层面具备适切性。

建议4:协作管理文化是推动可持续变革的必要条件。领导者应通过实践共同体促进校内及校际间的协作,同时学校与家长之间的合作对于化解技术使用中的合理顾虑同样至关重要。此类实践有助于促进相互支持,围绕共同目标营造学校共同体意识。

建议5:教育领导者的选拔与培养需契合数字化转型目标。教育领导者的能力建设应得到政策制定者更多关注。教育领导者的选拔与培训政策应持续完善,以适应时代需求并反映最新的实证经验。

建议6:学校及地方教育部门需配备专职人员予以支持。受过专业培训的专职人员能够减轻领导者在技术方面的负担,指导系统和学校领导者高效推进数字技术改革的实施。

(“国际与比较教育研究所”公众号)

拉美地区高等教育公平的 挑战与改进策略

近日,联合国教科文组织拉丁美洲和加勒比地区国际高等教育研究所(UNESCO IESALC)所长弗朗切斯科·佩德罗发表演讲,讨论了高等教育的责任和性质,并为如何迈向更具包容性的高等教育提出了策略。

目前,拉丁美洲的高等教育面临着这样一个十字路口:是继续复制不平等,还是成为社会变革的真正工具?对此,佩德罗认为:高等教育不仅是一种公共产品,而且是一种公共利益,这意味着整个社会都必须对此方面的资源进行保护和培养,而大学也必须将追求包容性、公平性和教育质量作为其发展的基本原则。

在演讲中,佩德罗首先分享了拉丁美洲和加勒比地区高等教育的现状,强调了在该地区

对高等教育进行结构改革、政策更新和资金保障的重要性。在演讲中,佩德罗对学术价值的传统概念提出了质疑,认为这些概念经常成为排除结构性弱势群体的借口。他强调,高等教育应被视为普遍受教育权的一部分。同时,佩德罗承认高等教育机构的自主性使得公共政策在高等教育系统中的实施具有复杂性。然而,他强调这种自主性不应成为各国履行其在指导教育发展实现公平、可持续和国家目标等方面的障碍。此外,佩德罗还指出,拉丁美洲目前有70%的学生就读于私立高校,这需要对资助和监管计划进行重新考虑,从而能够帮助系统缩小所有参与者的差距,而非扩大差距。

其次,佩德罗就如何迈向更包容的高等教育问题列出了几种应对策略。这些策略在拉丁美洲及加勒比地区已经得到了应用,并显现了良好效果:

平权行动政策:在入学配额上,对非洲人后裔、土著和农村学生给予特别支持。

院校多元化:在农村或城郊地区建立跨文化大学和高等教育中心,尽可能确保更大的地域覆盖面。

经济援助:除了奖学金,还有对放弃工作学习者的经济资助,作为对其“机会成本”的经济补偿。

同时,高等教育质量也是佩德罗关注的重点。佩德罗提到,尽管拥有大学学历者的平均收入是中学学历者的一倍,但目前只有不到50%的学生能够从大学毕业,并且高等教育机构数量正呈现出无序增长的状态。佩德罗认为,这反映了高等教育内部质量和内容缺乏相关性的问题。

最后,佩德罗警告了高等教育国际排名的影响,认为尽管国际排名起源于商业,但仍然是感知高等教育质量唯一的重要基准。同时他提出:高等教育的转型不仅是国家或大学的任务,而且是一项集体承诺,是一项共同利益。因此,包括政府、机构、教师、学生和民间社会在内的所有利益相关者都必须成为它的守护者。

(“上师大国际与比较教育研究院”公众号)

法国制定职业教育质量提升与行动治理计划

法国职业教育体系当前呈现多重结构性缺陷:一方面,成本压缩不仅导致供给波动,更直接推升了培训质量的离散度;与此同时,部分机构因教学投入不足而教育质量低下,从而侵蚀整个系统的公信力;加之标准化流程长期阙如,进一步抬升了建立统一高阶基准的难度。由此引发的信息不对称,使学生难以在充分知情的前提下理性择课,既损坏了优质机构的声誉,也削弱了职业路径规划的总体效能。

为回应上述结构性缺陷,法国近日制定了一项兼具变革性与系统性的质量提升与行动治理计划。该行动首次打破部门壁垒,构建跨部委协同框架,以立法、监管与实践三线并进:一方面清除既有监管盲区,强化政府职能;另一方面为监管机构赋权,使其手段更具靶向性、威慑力与协调性,既避免重复执法,也不误伤合规优质机构。最终,计划始终贯彻公共服务理念,以青年学员为中心,确保其获得高质量、高透明度且高保障性的培训,从而提升职业教育体系的公平度与经济-社会效益的双重收益。

该计划以四项行动展开,旨在以制度化手段回应系统缺陷:

1.提升培训质量:通过立法强化培训机构的法定义务,贯通CAP(职业资格文凭)至BTS(高级技术员文凭)的资格认证链,前置筛查潜在监管风险,并在CFA(学徒培训中心)系统内纳入工伤预防与扫盲任务。

2.保障青年与劳动力:完善学生与学徒的权益保护,优化培训与学位信息的公开渠道,并升级问题报告系统以提升信息透明度。

3.规范机构流程:在Qualiopi认证标准(教育领域的国家质量标准)中增设学徒制指标,细化审核程序,并逐步扩大其适用范围。

4.治理低信誉机构:建立市场禁入机制,提升监管效率,协调跨机构检查,并对行业商业行为实施全过程追踪。

(“上师大国际与比较教育研究院”公众号)