



借鉴与参考

第 2024 (03) 期

研究生教育发展动态

【主题】

深化新时代教育评价改革

【导语】

教育评价事关教育发展方向，事关教育强国成败

【观点】

- 01 以高等教育评价改革为抓手推进高等教育强国建设
- 02 建构中国特色高等教育评价体系
- 03 数智时代我国高等教育评估体系的转型与重构
- 04 新时代大学教师教学评价：反思与重构
- 05 破局、重构与协同：
生成式人工智能视域下的“破五唯”审思

导语

“

为深入贯彻落实习近平总书记关于教育评价改革的重要指示要求，进一步推动《深化新时代教育评价改革总体方案》落地落地，加快构建多元主体参与、符合我国实际、具有世界水平的教育评价体系，教育部印发《教育评价改革2024年重点推进事项》，以教育评价改革牵引深化教育领域综合改革，为教育强国建设激活力、增动力。

完善拔尖创新人才选拔、培养和评价机制，提升人才自主培养质量；加快建立卓越工程师培养评价新标准，着力破解工程技术人才培养与生产实践脱节的突出问题；加快破除“唯帽子”难题，推进人才称号回归学术性、荣誉性本质；强化重大人才工程牵引示范，牢固树立创新价值、能力和贡献的人才评价导向；深入推进高校科技人才评价改革试点，引导各类科技人才人尽其才、才尽其用、用有所成；深入推进高校评价改革，引导高校提升服务国家战略的能力和水平；建好用好教育督导评估信息化平台，提高评价效能。

”

来源：《教育评价改革2024年重点推进事项》节选，2024-03-15

以高等教育评价改革为抓手

推进高等教育强国建设

文 | 王智超, 东北师范大学教育学部

高等教育是教育强国建设的龙头, 推动高等教育强国建设是当前和今后一个时期高等教育的重要工作。高等教育评价作为指挥棒、风向标, 指引高等教育强国建设的发展方向, 关乎对“建设什么样的高等教育强国和如何建设高等教育强国”这一重大时代命题的正确认识、理解与把握。2020年, 中共中央、国务院印发《深化新时代教育评价改革总体方案》, 提出要充分发挥教育评价指挥棒作用, 从党中央关心、群众关切、社会关注的问题入手, 破立并举, 推进教育评价关键领域改革取得实质性突破。当前, 高等教育强国建设进入关键阶段, 我们必须以高等教育评价改革为抓手, 破除制约高等教育强国建设的障碍, 推动高等教育体系实现结构性变革、系统性重塑。

一、准确认识高等教育评价存在的两个问题

党的十八大以来, 我国高等教育事业取得了举世瞩目的伟大成就, 高等教育规模跃居世界第一位, 中国特色高等教育体系初步建立, 有力支撑了国家科技进步和人才培养, 成为富国强民的重要引擎。高等教育在快速发展的同时, 与之相匹配的高等教育评价体系仍未形成, 在助力高等教育内涵式发展上仍有较大不足。受制于不科学、不完善的高等教育评价体系, 我国高等教育独立性、自主性仍有不足, 办学特色还不够彰显, 在支撑高水平科技自立自强、探索人才自主培养、满足人民群众高等教育多层次多样化诉求等方面仍有差距, 高等教育强国建设任重道远。

一是高等教育评价调控功能异化, 导致高等教育呈现功利化发展倾向。依据高等教育评价结果对教育资源进行有序调配, 借此实现对教育结构、布局的调整以达到国家需求与个人需求的平衡是高等教育评价调控功能的重要表现。我国高等教育依据教育评价结果进行资源配置时有较强的排他性倾向, 使得高等教育评价的调控功能异化, 诱发了高等教育发展的功利化倾向。一方面, 不同高校在科研实力、人才培养能力、社会服务水平等方面存在巨大差异。高等教育评价本应基于不同高校发展存在差异的客观事实设计有针对性的资源配置策略, 以引导不同高校分类发展。但从现实看, 分类多元、扶弱促强的高等教育评价指标体系还不完善, 统一的评价标准湮没了高校之间存在的客观差异, 依据高等教育评价结果进行资源的排他性配置迫使一些高校为了获得更多资源不得不脱离学校现实, 对项目、人才及学校层次、规模进行盲目追求, 使学校畸形发展, 破坏了高等教育格局的合理建构, 造成了高等教育资源的浪费。另一方面, 我国高校自身资源创生能力极为有限, 在发展所需资源上过度依赖政府供给。高等教育评价结果作为政府资源配置的重要依据, 高校在发展过程中不得不唯评价指标, 在日常管理、内部评价体系构建等方面走向工具主义与绩效主义优先的简单化、片面化发展道路, 继而导致一些高校在办学中违背基本规律, 大量学科、专业简单重复设置, 在发展中丧失办学自主权。

二是高等教育评价导向功能失灵，导致高等教育呈现无序化发展倾向。高等教育评价作为指挥棒、风向标，对牵引高等教育有序发展具有重要作用，可以为高校发展供给必要且稳定的基本规范。从现实看，高等教育评价受外部环境影响较大，独立自主的评价理念仍未形成，不稳定、不连续的无序发展导致高等教育陷入低质发展陷阱。一方面，高等教育评价缺少一以贯之的内在价值，相对稳定的评价标准缺失，相关主体及利益相关者在学科建设、人才培养、社会服务等方面缺少稳定、持续、规范的参考标准，致使我国高等教育总体规模虽庞大，但各层级、各类型高等教育处于失序发展状态，未能形成服务国家需求、满足人民群众多元教育诉求的高等教育供给体系。另一方面，我国独立自主的高等教育评价指标体系尚不成熟，对高等教育发展和评价体系构建的价值澄明和经验凝练存在不足。一些人热衷于简单移植西方高等教育发展范式来评价我国高等教育发展水平，但西方高校与我国高校在办学体制、发展目标和育人目的上并不相同，在高等教育评价指标体系构建上的“拿来主义”既不利于我国高等教育独立发展，又容易出现方向性错误，这也是高等教育评价导向功能失灵的重要表现之一。



以上问题的存在阻滞了高等教育强国建设的有效推进，不利于高等教育现代化的实现。时代发展呼唤高质量高等教育，社会主义现代化国家建设要求高等教育必须以新的姿态回应国家诉求，在迈向中国式现代化的时代征程上进一步推进高质量高等教育体系建设。

二、正确处理高等教育评价改革中的三对关系

扭转高等教育评价存在的这些问题，有力推进高等教育强国建设，我们必须瞄准高等教育评价的堵点、难点，重点处理好当前焦点与长远布局、重点突破与全局改进、自我模式与域外经验的关系，发挥社会主义体制优势，走中国特色社会主义高等教育建设之路。

1. 处理好高等教育发展过程中当前焦点与长远布局的关系

高等教育强国建设既要关注当前，更要着眼未来。一方面，高等教育评价要引导高等教育人才培养聚焦到适应现实需求上。高等教育强国建设的核心在质量，高等教育质量在很大程度上体现在人才培养质量上，高等教育能否有效回应党、国家、社会各方的诉求，核心也正体现为能否培养出符合党、国家、社会需要的高质量人才。人才培养质量是高等教育的生命线，高等教育评价必须聚焦高等教育人才培养“能不能”“行不行”这一关键问题，将人才培养质量作为评价高校办学水平的核心标准，着力构建既符合当前高等教育发展实际，又契合未来高等教育发展趋势、立体综合多维的人才培养质量评价体系。我国要创新人才培养质量评价标准，做到人才培养过程与人才培养结果并重，重点考察高校所培养人才的社会贡献度，将评价由关注学生自身发展的单一焦点转为关注学生自身发展和学生社会贡献的二维焦点，以此为核心推进学科结构和教师队伍结构的优化调整。另一方面，高等教育评价应引导高等教育注重长远发展。高等教育强国建设是一个渐进发展过程，我国既需要坚实的基础，更需要创新引领，这就要求我们在考察高校办学水平的同时，必须为高校未来发展预留可能空间。因此，高等教育评价需要转变内在价值，着眼高等教育未来发展，在促进高校发展优势学科的同时注意对弱势学科的培育，在以市场需求为导向决定资源配置的同时注重发挥政府调节作用，保护当前呈现弱势却有长远战略意义的学科。此外，高等教育强国建设中的高等教育应该是均衡发展的。教育部门应发挥教育评价在高等教育布局结构调整中的引导作用，使高等教育资源在区域间合理分布，结合各地优势形成良性互动格局，促进高等教育效能有效发挥。

2.处理好高等教育发展过程中重点突破与全局改进的关系

高等教育强国建设既要点位突破，实现重点领域的高位引领，又要筑牢基础，惠及全体人民。一方面，高等教育强国建设要求高等教育发展具有引领性。从国家发展战略需求看，高等教育必须在某些领域、学科引领世界科学发展水平，培养在世界范围内具备核心竞争力的优秀人才。高等教育评价应聚焦高质量发展这一核心要义，构建更有利于一流大学、一流学科建设的评价指标体系，瞄准世界科技发展前沿，对标国家重大战略需求，不断提升人才培养质量。同时，许多教育问题的症结在教育之外，教育现象背后往往有着深刻的社会及制度基础，因此我们有必要从更加宽广的社会视角审视教育评价改革。高等教育评价改革要在分类评价上下功夫，推动高等教育在发展侧重上有所为、有所不为，实现不同类型高校的分层次、分阶段、有步骤发展，优先满足国家迎接不确定性新挑战、对接经济发展新态势、引领世界科技新潮流的战略需求，提升高等教育与经济社会发展的耦合度，为中国式现代化提供强有力支撑。另一方面，高等教育强国建设要求高等教育整体质量得到有效提升。高等教育强国追求人才培养结果的多样化，要求高等教育能够满足人民群众多样化、个性化的高等教育需求，实现不同类型高校的特色自主发展。相对于重点突破，高等教育质量的全面提升更具有重要性和基础性，高等教育在追求引领性的同时必须注重自身的厚重性。高等教育发展应具备更为稳重、庞大、坚实的普及化程度和高质量的发展水平，注重高等教育体系的金字塔样态和院校间协同式发展生态系统的形成，如此才能满足不同人群差异化的高等教育需求。因此，高等教育评价既要有适应引领性的高质量标准，又必须有适应一般性、特色化的标准，确保相关资源的合理有效配置，由此改变单一评价标准导致的高校同质化发展倾向，确保形成高等教育建设的良性生态。

3.处理好高等教育发展过程中自我模式与域外经验的关系

高等教育强国建设既需要形成具有自身特色的经验模式，又需要合理借鉴域外经验；既需要坚持中国高等教育办学特色，又需要吸收不同国家高等教育办学经验。这就要求高等教育评价要在主体、指标、范式、手段、结果使用等方面兼容并蓄、博采众长，形成符合高等教育强国建设实际的高等教育评价体系。一方面，高等教育评价要扎根中国教育实际，聚焦中国高等教育现实问题，为高等教育强国建设保驾护航。我们要系统总结和凝练中国共产党领导下的中国特色高等教育发展经验，注重在高等教育评价体系建构过程中从中国高等教育的独特价值理念、治理模式和改革方法论中寻找内在逻辑，逐渐形成更富中国特色、更符合中国高等教育发展实际的评价体系。同时，我们要坚持问题导向、系统观念，以立德树人成效为根本标准，发挥高校主动性，对不同类型高校有底线、有区别、有目的地进行综合评价，注重发挥评价的发展性功能。另一方面，高等教育强国建设不应故步自封，在坚持中国特色的同时应具备全球视野。我们要多渠道、多途径开展与国外高等教育的联系与合作，面向世界科技前沿、面向经济主战场、面向国家重大需求、面向人们生命健康，加快实现高水平科技自立自强。同时，我们既要广泛吸收借鉴不同国家高等教育评价体系建设建设的有益经验，又要为世界高等教育发展提供可供借鉴的中国高等教育评价经验，体现高等教育强国的辐射性和影响力。高等教育评价体系建设要以有效促进更全方位、更多层次、更宽领域、更加主动的教育国际交流与合作新格局的加快形成为目标，积极参与国际高等教育评价标准制定，提出中国高等教育评估方案，掌握高等教育评价的世界话语权，保持大学精神的开放与自由。

• 来源：《高校教育管理》，2024年第1期

建构中国特色高等教育评价体系

文 | 周光礼，中国人民大学教育学院



建构中国特色高等教育评价体系，对推进高等教育治理现代化、建设高等教育强国意义重大。中国在高等教育督导和评价方面积累了一定经验，取得了明显成效。对照党的二十大提出的目标任务，对标建设高质量高等教育体系的新要求，高等教育督导和评价在评价标准、分类指导、体系优化、数字化转型等方面仍存在问题。建构中国特色高等教育评价体系，应在“中国特色”上下功夫，基于中国独特的历史、独特的文化、独特的国情，从价值理念、治理模式、改革方法论中寻找建构中国特色高等教育评价体系的底层逻辑，坚持价值引领、分类评价、问题导向、守正创新、数字化转型等五项基本原则，优化政府评价、规范社会评价、强化高等学校自主评价。

关键词：高等教育强国；教育治理现代化；高等教育评价体系

党的二十大报告提出“加快建设高质量教育体系”，这是高等教育改革与发展的中心任务。高等教育评价体系是推动高等教育高质量发展的重要抓手。在推进中国式高等教育现代化的新征程中，加快建构中国特色高等教育评价体系，对推进高等教育治理现代化、建设高等教育强国意义重大。改革开放以来，我国“从无到有”建立起评估、认证、监测“三位一体”的高等学校（以下简称高校）评价体系，为保障高等教育有序发展发挥了积极作用。然而，现行评价体系难以适应建设高质量高等教育体系的时代要求，建构中国特色高等教育评价体系迫在眉睫。

一、高等教育评价的现状与问题

高等教育评价体系包括两个部分，一是来自高校之外的评价，即高校外部评价体系；二是来自高校之内的评价，即高校内部评价体系。由于组织惰性的存在，没有外部压力下的高校内部评价很难产生积极效果，因此，高校外部评价体系被率先建立起来。高校外部评价体系主要包括三个部分：政府主导的认可性评价、第三方机构提供的认证性评价和商业机构推出的市场性评价。其中，商业机构推出的市场性评价即大学排行榜。这种评估方式旨在适应高等教育市场的多元主体和多元需要，用老百姓听得懂的语言提供高等教育的质量信息，但只是高校声誉的一种检验，不是真正意义上的教育评价。

我国高等教育评估体系以政府主导的认可性评价为主，主要包括以教学评估为核心的本科院校评估和以办学水平为核心的学位和研究生教育评估。

我国在高等教育督导和评价方面积累了一定经验，取得了明显成效，初步建立适应中国国情的高等教育评价体系。进入中国式现代化的新时代新征程，对照党的二十大提出的目标任务，对照建设高质量高等教育体系的新要求，对照应对高等教育普及化、市场化、国际化、数字化的新挑战，对照高校对高等教育评估改革的反馈意见，对照社会公众对高校评价体系建构的新期待，我国高等教育督导和评价在评价标准、分类指导、体系优化、数字化转型等方面尚

有问题待解决。

1. 中国特色高等教育评价标准亟待确立

2018年，在北京大学师生座谈会上，习近平总书记提出了两条中国特色高等教育评价标准：一是以立德树人的成效作为检验学校一切工作的根本标准，二是师德师风是评价教师队伍素质的第一标准。然而，迄今为止，这两条标准尚停留于指导思想层面，没有落实落细于具体的高校评价标准。建立能够反映高校立德树人实际的评价指标体系仍是一个尚未解决的问题，教师评价的破“五唯”、立新标任重道远。

2. 高校分类的逻辑亟待厘清

分类评价是《深化新时代教育评价改革总体方案（2020）》（以下简称《评价总体方案》）对高校评价改革提出的基本要求。如其所言，“推进高校分类评价，引导不同类型高校科学定位，办出特色和水平”。据此，教育部提出，要立足普及化阶段高等教育的多样性、多元化发展的特点，对不同类型高校开展不同导向的评估评价。然而，时至今日，国家层面尚未出台作为评价依据的高校分类标准，分类评价体系建设使命艰巨。

3. 高等教育评价体系亟待优化

高等教育评价体系从评价主体的角度可以分为政府评价、社会评价和高校自我评价。现有高等教育评价体系以政府问责评价为主，其系统性、整体性、协同性不够，主要表现为三个“不强”。一是政府评价的统筹性不强。以国家层面为例，高等教育评价项目除以教学评估为核心的本科院校评估和以办学水平为核心的研究生教育评估，还有教育部不同司局根据业务需要，开展重点实验室评估、“双一流”建设动态监测和成效评价、中外合作办学



评价等行政性专项评估。由于缺乏顶层设计，各部门各自为政，在评价时间、评价对象、评价内容、指标体系等方面缺乏协调机制，导致多头评估、重复评估、同时评估，高校不堪重负。二是社会评价的规范性不强。鼓励社会机构参与高等教育评价是落实“管办评分离”的重要举措。然而，到目前为止，我国并没有真正的第三方教育评价机构。三是高校自我评价的主动性不强。高校是教育质量的责任主体，只有激发高校的主动性，才能推动教育教学质量持续改进，最终形成质量文化。然而，当前高校质量主体意识不强，建立内部质量保障体系的积极性、主动性、创造性还没有充分发挥出来，对高校内部质量保障体系的构成要素、运行机制及成效评价监测的认识还不到位，难以形成持续改进的质量文化。

4. 高等教育评价的数字化转型亟待强化

数字化、智能化蓬勃发展推动高等教育评价转型升级。然而，全国统一的高等教育评价大数据库尚未建立，“信息孤岛”现象普遍存在，高等教育评价数据治理亟待突破。一方面，教育管理部门、各高校都建立了各自的数据平台，形成大而全、小而全的数据库，社会各门户网站也积累了海量的教育数据。另一方面，在教育管理部门之间以及教育管理部门与高校内部，尚未充分实现数据的共建、共享、共用，重复建设、壁垒森严问题突出。更有甚者，各种教育数据统计口径不统一、结构化程度低，难以支持实施常态化监测和专门评估。

二、中国特色高等教育评价体系的底层逻辑

当前我国的高等教育评价体系，基本上是在学习借鉴西方经验的基础上建构出来的。比如，审核性评估理念来自英国的“卓越教学框架”，工程专业认证来自“华盛顿协议”。建构中国特色高等教育评价体系，应在中国特色上下功夫，从中国高等教育独特的价值理念、治理模式、改革方法论中，寻找建构高等教育评价体系的底层逻辑。

1. 价值逻辑：立德树人、政学结合的高等教育理念

价值逻辑植根中国独特的文化，文化的内核是价值观。以“大学之道”建构评价体系，可使大学回归初心、回归育人本位。从这个意义上说，立德树人、政学结合的高等教育理念是中国对全球高等教育的最大贡献，能为其他国家文明提供实质性启示。立德树人、政学结合，是建构中国特色高等教育评价体系的逻辑。

2. 历史逻辑：有为政府、有效市场的高等教育治理模式

历史逻辑植根于中国独特的历史，强调政府在高等教育治理中的主导作用是中国的文化传统。推进高等教育治理体系和治理能力现代化，为我们从历史维度理解与认识新时代高等教育评价改革提供了新的视角。

进入新时代，高等教育治理需要新理论新观念新模式。推进高等教育治理现代化必须坚持守正创新。一方面，需要整合科层式治理和市场式治理的优势，克服高等教育治理中的政府失灵和市场失灵。另一方面，需要紧跟时代步伐，以新的理论指导新的实践。多元共治理念可以成为新时代中国高等教育治理的新模式。多元共治强调发挥科层式治理和市场式治理的优点，克服两者的缺点，推动政府问责向社会问责转变。以此为指引，中国特色高等教育评价体系的基本框架应该包括三个方面：政府评价、社会第三方评价、高校自主评价。

3. 实践逻辑：问题导向、系统观念的改革方法论

实践逻辑植根中国独特的国情。党的二十大报告提出，实现中国式现代化必须坚持问题导向和系统观念。

以问题导向、系统观念的改革方法论为指引，要求我们把握好中国式高等教育现代化的新特征新要求，突出系统化改革思维，整体性再造中国高等教育评价体系的流程，从评价模式、体制机制、结构类型、评价标准、评价方式等各方面进行建构，以高质量的高等教育评价引领高质量的高等教育体系建设，全面推进中国高等教育治理现代化。



三、中国特色高等教育评价体系的行动框架

价值逻辑、历史逻辑和实践逻辑之间具有内在统一性和高度关联性，共同构成了高等教育改革的逻辑。中国逻辑为建构中国特色的高等教育评价体系提供实施路径和行动方案。

1. 基本原则

第一，坚持价值引领。建立完善以立德树人成效为根本标准的高等教育评价指标体系。立德树人、政学结合，既是中国高等教育的优秀传统，也是中国高等教育评价改革的底层逻辑。在立德树人的旗帜下，高等教育评价改革要破以科研指标为主的评价，立以人才培养指标为主的评价；要破以结果评价为主的方式，立以过程评价和增值评价为主的方式；要破以“五唯”

指标为主的教师评价，立以师德师风为主的教师评价；要破唯分数的不良做法，立全面发展的育人要求。

第二，坚持分类评价。分类评价是高校评价的基本原则。2023年，教育部印发的《关于深化新时代高等学校评估改革方案》（以下简称《评估改革方案》）明确提出，注重引领和规范相结合，对不同类型高校开展不同导向评估。当务之急是针对不同类型高校研制科学合理的质量标准，引导不同类型高校科学定位，办出特色和水平。

第三，坚持问题导向。《评价总体方案》指出，要坚持问题导向。建构中国特色高等教育评价体系，必须着力解决高等教育领域制约性、瓶颈性、关键性问题。一方面，要推动高校坚持“四个面向”，加快实现高水平科技自立自强；另一方面，要引导高校“为党育人、为国育才，全面提高人才自主培养质量，造就拔尖创新人才”。



第四，坚持守正创新。一方面，我们要全面总结改革开放以来我国高等教育评价的成功经验，充分吸收借鉴国外高等教育评价的最佳实践。另一方面，要紧跟时代步伐，顺应实践发展，以满腔热忱对待一切新生事物，不断拓展认识的广度和深度，敢于创新，以新的评价理论指导新的评价实践。

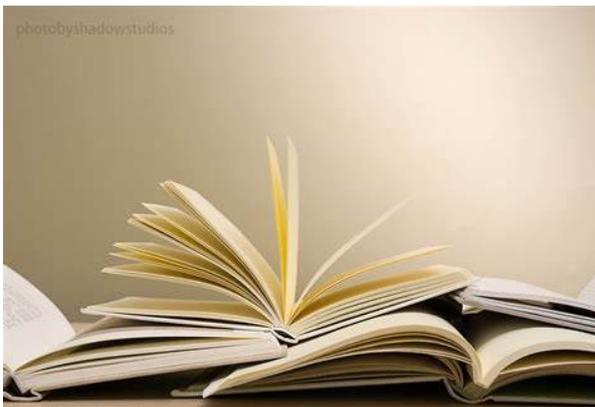
第五，坚持数字化转型。数据是高等教育评价的基础。一是要建立“覆盖全国、统一标准、上下联动、资源共享”的高校基本状态数据库；二是要积极运用动态监测数据，客观反映高校办学治校现状、组织特征和发展趋势，开展同型比较、问题诊断、质量预警，督促学校持续改进；三是要强化社会问责，对公众关注的办学数据须在一定范围向社会发布，充分发挥基本状态数据在政府问责、社会监督中的作用；四是要强化业务培训，提升高等教育评价从业人员数据素养，为高等教育评价数字化转型培养专业队伍。

2. 建构三大评价体系

中国特色高等教育评价体系有三大支柱：政府评价、社会评价和高校自我评价。推进治理现代化，高等教育评价改革须进行战略性转换，要从评价体系的修修补补转向全面重构，从政府问责转向社会问责，从外部评价为主转向内部评价为主，从单项评价转向综合评价。

(1) 优化政府评价

事实上，构建以政府评价为主、多元参与的高等教育督导体系，一直是中国特色高等教育评价体系的核心部分。然而，当前政府评价存在多头评估、重复评估、同时评估等问题。优化政府评价的关键是坚持“法不授权不可为”原则，即政府组织的评估项目如果没有明确的法律授权，应该合并或取消。根据此法治原则，政府组织的评估应该以周期性综合评价为主。作为兜底性评价，所有高校每五年必须接受一次政府问责。依据新时代教育评价改



革总体目标，落实高校分类指导和分类评价要求，应将高水平大学综合评估、本科高校审核性评估、新建高校合格评估作为主要的评估类型。

高水平大学综合评估主要是国家教育行政部门面向“双一流”建设高校和其他高水平研究型大学开展的综合性评估。这类高校是国家战略科技力量的重要组成部分，是支撑高等教育强国建设的一流大学集群。对此类高校开展综合评价，旨在打破科学研究与人才培养割裂，实现本科教育与研究生教育的贯通与一体化。为此，应将学科评估、学位点合格评估、专业学位水平评估、本科教育教学评估、“双一流”建设成效评价、教育部重点实验室认定、国家人文社会科学基地认定等评价活动整合。评估内容聚焦坚持党的全面领导、拔尖创新人才自主培养能力、高水平科技自立自强、世界一流可比水平、服务世界人才中心和创新高地建设、高等教育综合改革进展。

（2）规范社会评价

多元共治是高等教育评价改革趋势，社会参与是多元共治的精髓。通过社会参与来改善政府问责、让专业的人做专业的事，是高等教育评价改革的重要内容。“管办评分离、放管服结合”为社会评价提供基本遵循。社会评价遵循“法不禁止即自由”原则，为社会组织和人员参与政府和高校评估提供有效渠道，让社会组织和人员参与政府评估的标准制定、过程监督、体系完善之中。规范社会评价的关键是区分公益性的社会评价和营利性的市场评价，培育有公信力的第三方评价机构。

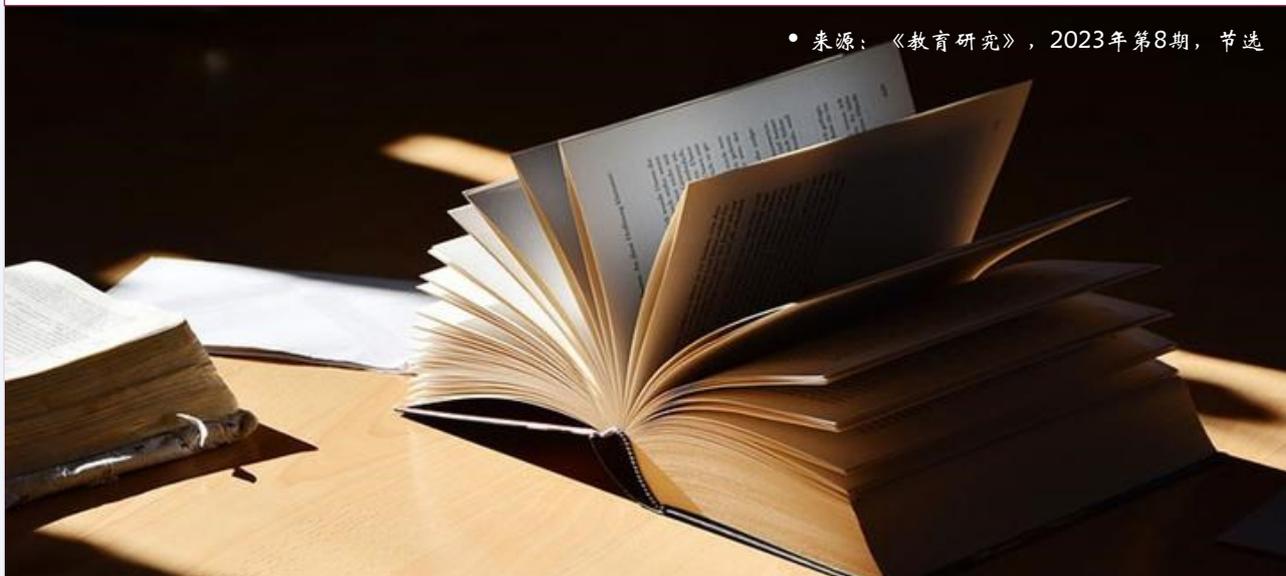
社会评价以专项评价为主，不同于政府的综合性评价。狭义的社会评价特指公益性第三方专业机构开展的高等教育认证和调查。教育行政部门的专项评估项目也可以委托第三方专业机构开展。培育有公信力的第三方专业机构，坚持正确导向，对各类民间机构参与高等育评价进行有效监管。

（3）强化高校自主评价

建立高校内部质量保障体系的关键是强化高校自主评价，围绕培养质量、学科专业建设、条件保障进行常态化自我评估和总结。坚持以学生为中心、以立德树人成效为中心、以持续改进为中心的理念，统筹学校、学院和学系（教研室）三个层面的系统配合，贯通目标、实施、评价三个环节的相互促进，建设教学共同体，提升学校整体的教育质量。

总之，中国特色高等教育评价新体系顺应高等教育质量保障体系由外部转向内部的全球趋势，体现了“以学生为中心、产出导向、持续改进”的世界教育认证理念，充分彰显“以立德树人成效作为评价学校一切工作根本标准”的中国逻辑。

• 来源：《教育研究》，2023年第8期，节选



数智时代我国高等教育

评估体系的转型与重构

文 | 王战军 李旖旎, 北京理工大学

数智时代我国高等教育的转型呼唤高等教育评估的变革。从内涵维度透视, 高等教育评估的目的由单一价值判断转向多元智能优化, 评估的尺度由单视域及时反馈转向多视域实时呈现, 评估的模式由数据集中密集转向区块互联分布。从功能维度透视, 全息评估支撑多元主体从认识价值上升为“共创价值”, 人工智能驱动主管部门从被动应对决策上升为主动“预警预测”, 人机互动牵引教育教学质量从持续改进上升为“提质增效”。建构新型高等教育评估体系须秉持“以评促强、争创一流”的发展理念, 探索“平台化”协同慧治的治理模式, 创造“教育评估智能体”的超级生态, 持续系统提升高等教育质量, 服务于教育强国建设。

关键词: 高等教育评估; 数智时代; 教育强国; 教育评估智能体

一、引言

自改革开放以来, 我国高等教育评估已走过40多年历程, 高等教育评估体制机制日臻完善, 高等教育评估的价值取向实现了由“保障高等学校质量”到“提高人才培养质量”, 再到“建设高等教育强国”的“三级跳”, 为2035年如期建成教育强国、科技强国、人才强国提供了有力支撑。近年来, 5G通信、区块链、人工智能、移动互联网、虚拟现实等前沿技术的快速迭代和应用, 推动人类社会迈入“数智时代”。随着数智时代的到来, 高等教育领域也正在经历一场“大变局”。在教学端, “远程教学”成为高等教育领域新常态, 上课下课、校内校外、区域之间的时空界限被打破, 线上与线下教学实现无缝衔接。在学习端, 在线教学拓宽了大学学生的学习方式, 养成了大学生“碎片化”学习习惯, 增强了大学生掌握学习时间的灵活度和学习内容的针对性, 满足了大学生的个性化知识需求。高等教育逐步与移动互联网的深度联结, 推动了高等教育的主体内容数字化、交互形式网络化、课堂界限模糊化、保障资源智能化。

数智时代高等教育的转型同样助推高等教育评估的变革, 如何运用一系列数智技术赋能高等教育评估已成为备受关注的核心议题。现有研究多聚焦整体性教育评估, 重点关注数智技术对教育评估方法和手段的影响, 却忽视了未来教育评估的思维变革, 对高等教育评估数智化转型支撑有限。因此, 站在数智时代的起点上, 探析新时期高等教育评估的内涵、功能和体系重构, 对于顺应时代发展变化、把握历史机遇, 依托高等教育评估变革牵引高等教育在育人方式、办学模式、管理体制、保障机制等方面的综合改革具有重要意义。

二、数智时代高等教育评估的新内涵

进入数智时代，5G通信、区块链、移动互联网、人工智能等前沿技术持续推动高等教育评估的改进和迭代，在根本目的、状态监测和基本模式等方面为高等教育评估注入新的内涵。

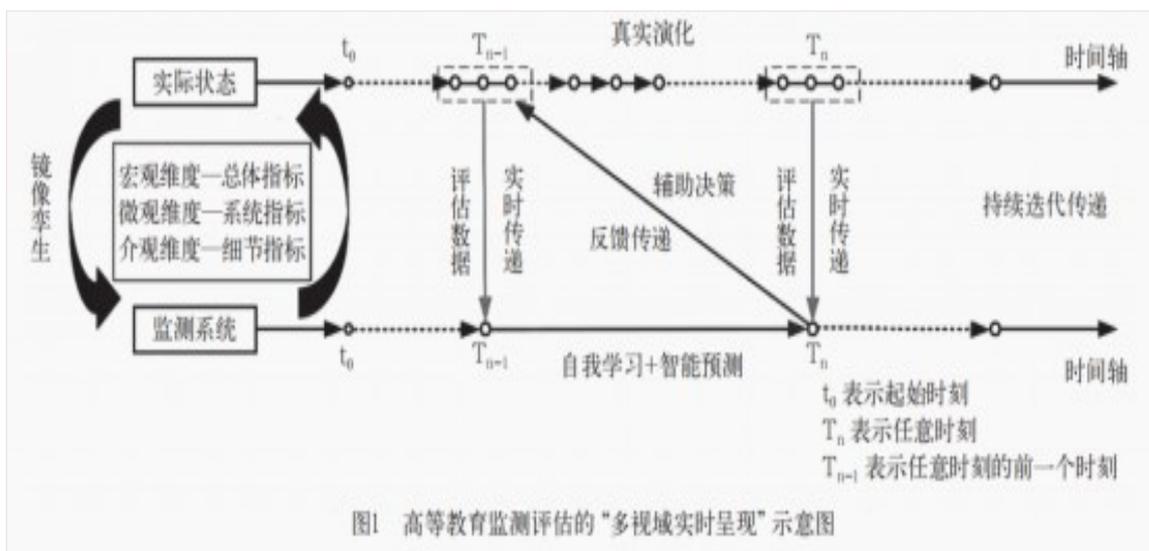
1. 高等教育评估根本目的：单一价值判断转向多元智能优化

进入数智时代，在人才培养需求变化和各项前沿技术发展的双向助推下，高等教育将逐渐呈现知识迭代快速化、课程更新频繁化、教学资源海量化、培养方案动态化、学习模式自主化等特点。传统高等教育评估的单一维度价值判断结果将难以满足数智时代高等教育变革和人才素质提升的快节奏发展需求。与此同时，将大数据、区块链、人工智能、辅助决策、专家系统、高性能计算、数字孪生等数智技术赋能高等教育评估，可实现在“优”“良”“合格”等传统单一评价维度的基础上，更进一步在“教学内容”“教学方式”“知识资源”“培养方案”等多元维度上自适应生成若干优化方案，通过在虚拟空间进行多方案平行数值模拟，快速迭代计算出最优的评估量化结果，并提出可行的优化方案或组合方案建议，直接供评估主体和评估对象遴选与决策，从而支撑高等教育质量持续提升。因此，数智时代高等教育评估的根本目的必将从单一的价值判断转向多元的自适应智能优化。



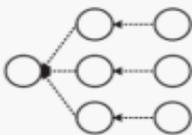
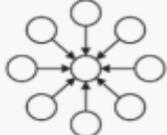
2. 高等教育监测评估尺度转型：单视域及时反馈转向多视域实时呈现

状态监测评估是高等教育评估的重要组成部分。随着高等教育数字化转型，以及区块链、5G网络通信、移动互联网、人工智能等前沿技术的深入应用，各维度间多尺度耦合的技术门槛降低，“多尺度耦合视域”成为数智时代状态监测技术未来的发展方向之一。宏观维度的评估可通过对国家人才、科技、就业、区域等与高等教育相关的公开数据进行数据深度挖掘和整合来实现；微观维度的评估可通过对高校、院系的实时监测动态数据来实现；介观维度的评估可通过对MOOC、线上课堂、网络会议等与师生强捆绑的移动APP的动态数据来实现。而且，各维度的监测数据能够共用共享、深度融合，评估者对个体APP数据进行“积分”处理就可以支撑宏观趋势判断，对宏观数据进行“微分”处理便能支撑微观和介观的机理分析。因此，数智时代高等教育状态监测评估的空间尺度将呈现“多尺度耦合视域”的趋势，使状态监测具备根据评估需求进行空间尺度“变焦”的能力，如图1所示。



3.高等教育评估基本模式转型：数据密集集中转向区块互联分布

我国高等教育评估模式大体可划分为三个阶段：从1985年至2010年为第一阶段，形成了“经验驱动”评估模式；从2011年至2019年为第二阶段，形成了“数据密集集中”评估模式；2020年之后开始进入第三阶段的数智时代，有望形成“区块互联分布”评估模式。三个阶段评估模式的显著特点和变化可以从信息域、数据域、技术域等层面来呈现，如表1所示。

	第一阶段 “经验驱动”评估模式	第二阶段 “数据密集集中”评估模式	第三阶段（数智时代） “区块互联分布”评估模式
信息域			
	点对点单向离散式 (内部呈现)	中心集中单向连续式 (局部公开)	区块分布双向平行式 (跨域共享)
数据域	专家驱动的 小样本数据	模型驱动的 密集型数据	人工智能驱动的 海量数据
技术域	现场调研 人工采集 传统评估	互联网+ 大数据 数据挖掘 4G通信	5G通信 人工智能 区块链 辅助决策 智能运维 高性能计算 健康管理 数字孪生 多目标优化

从信息域上看，第一阶段的评估信息流为“点对点单向离散式”，信息流架构呈现多维度多层次递阶特点，评估信息覆盖不全，信息传递效率低，各高校、各层级之间信息隔离，且信息只在教育管理部门或第三方评估机构内部呈现，仅能初步反映高等教育的整体状态，难以快速反映整体状态的短期内变化，对评估工作的支撑力度有限。第二阶段的评估信息流为“中心集中单向连续式”，信息流架构呈现扁平化“行星状”特点，评估信息覆盖较为全面，信息层级得到压缩，可基本反映高校教育教学的多维度状态，但各高校之间信息仍然隔离，评估信息也只能做到局部公开。在数智时代的第三阶段，评估信息流为“区块分布双向平行式”，信息流架构呈现“区块互联分布”特点，动态监测数据平台对高校教育状态进行全自动化的信息获取，信息数据可双向传递，各节点之间信息互联互通，评估信息全面跨域共享，各高校教育状态信息实时呈现，可全面支撑高等教育评估，并可实现各高校间的相互促进与交叉借鉴。

从数据域来看，第一阶段获取的评估数据属于“专家驱动的小样本数据”，一般由教育管理部门或第三方评估机构定期组织评估专家赴高校现场，以专家的理性、知识、经验对高校教育教学状态进行主观判断。第二阶段获取的评估数据属于“模型驱动的密集型数据”，在定期专家评估的基础上增设了教学工作质量常态监测机制，通过远程技术定期采集各高校教育教学的状态数据，以获取更为密集的数据，最后利用评估模型对高校教育教学状态作出客观判断。在数智时代的第三阶段，获取的评估数据属于“人工智能驱动的海量数据”，在密集型数据的基础上，通过数据的自我学习、自我生成和自我反演，得到时间维度和空间维度更多未知状态点的大量预测数据，“自适应”建立起更为精细的评估模型，并通过与各高校实际教育教学状态信息进行多轮迭代比对验证，数据与模型持续优化，依托先进算法对高校教育教学状态作出更为精准的客观判断与趋势预测，科学指导高等教育管理决策。此阶段评估数据形式多样、体量巨大，可基本替代传统的专家判断模式，数据与智能成为评估的主导。

从技术域来看，第一阶段高等教育评估的支撑技术较为基础且单一，主要依靠现场调研、人工采集、传统评估等方式。进入第二阶段，“互联网+”理念的兴起与普及使高等教育状态常态监测成为可能，并通过大数据等技术实现对数据的清洗、整理、提取与分析，推动该阶段形成了“数据

密集集中”的高等教育评估模式。在数智时代的第三阶段，去中心化的区块链技术广泛应用为分布式互联的信息流架构提供了理念借鉴；5G通信技术支撑更高效的数据传递和更海量数据的跨区域互联；人工智能技术推动数据的不断自我学习；高性能计算技术使得对海量数据的分析处理与复杂运算得以实现；多目标优化技术可通过反复迭代得出最优的改进方案；健康管理技术可以更好地监测各高校教育教学状态信息；数字孪生技术促进评估对象数据信息实时平行传递的实现；辅助决策、运行支持及智能运维等技术则可以为高等教育评估不断优化提供坚实基础。这些先进技术的深度融合塑造了数智时代高等教育评估的“区块互联分布”新模式。

三、数智时代高等教育评估的功能转型

立足新时代，高等教育评估的重心已由“针对”评估对象转向“服务”评估对象，评估的功能迫切需要随之优化升级。数智时代高等教育评估在功能定位上更加追求多元主体“价值共创”、主管部门“预警—预测—决策”一体化，以及教育教学质量的“提质增效”。

1. 全息评估支撑多元主体从“认识价值”上升为“共创价值”

数智时代，先进技术有望推动高等教育“全息评估”的实现，使得多元评估主体从“局中人”转向“当事人”，进而为“创造价值”提供支撑。

数智时代的高等教育评估将呈现系统性的特征。虚拟与现实空间的交叠引起多源、异构、海量数据的汇聚，消除了基于“小样本数据”的传统高等教育评估数据采集分析所面临的量与质难以兼顾的技术瓶颈。数智技术带来的“高涵盖性”评估数据进一步提升了高等教育评估环境的透明度，使数据来源不再局限于高等教育系统内部，加强了高等教育数据与国际、区域、行业需求等数据的结合、交换和共建共享，从而催生更多数量和种类的新型高等教育数据，有助于多元评估主体对高等教育发展现状进行全面把握。对评估主体而言，只有把握评估过程的全要素，充分释放评估数据“生产力”，才能兼顾“局中人”和“当事人”的双重身份，才能以系统思维和战略眼光谋划高等教育发展之路，将马克思所说的“解释世界”和“改造世界”有机统一起来，推动高等教育评估的多元主体在协调互动过程中实现价值共创。



2. 人工智能驱动教育主管部门从被动“应对决策”上升为主动“预警预测”

从治理的角度看，教育评估是提高教育决策科学性的重要手段。高等教育决策的科学性体现为促使高等教育活动达到合规律性与合目的性的统一^[27]。合规律性是指高等教育决策要遵循教育自身的发展规律，合目的性是指高等教育决策要凸显教育对人和社会的价值。合规律性是为了合目的性，高等教育评估主体不仅要探索教育规律，更要揭示教育价值，体现高等教育对人和社会的意义。

大数据和人工智能技术为高等教育决策优化提供了新的视角和途径。基于大数据背景下的人工智能，未来高等教育决策将更加趋向“预警—预测—决策”一体化。例如，在未来的院校评估中，评估专家可以借助大数据和人工智能技术开展线上评估，以便对高校的发展状态进行持续性跟踪，并对高校运行过程中存在的异常和风险进行提前警示。这不仅减少了专家进校考察的时间和成本，

还促进了评估过程的科学性。在评估结果反馈阶段，评估专家可以利用数智化手段对高校的发展趋势进行科学推测，以便更好地帮助高校绘制发展图景。预警功能和预测功能的有机结合，有助于高等教育评估主体从表象判断深入到系统推测和预判，提升高等教育决策的洞察力和前瞻性，进而推动高等教育与经济社会发展需求的无缝对接，从而真正实现合规律性与合目的性的统一。

3. 人机互动助推高等教育教学质量从“持续改进”上升为“提质增强”

为教学服务是高等教育评估的核心诉求。数智时代高等教育教学质量的改进将由“持续改进”上升为“提质增强”，关注的焦点是提升教学质量、增强竞争力。在大数据的驱动下，高等教育评估者在与机器智能的交流过程中推动了知识的不断扩展、共享和创新。可以说，面向知识的评估不仅为高校和教师及时改进教学提供了内在动力，更为高等教育教学质量持续提升提供了有效路径。

面向知识评估的可能性和有效性取决于信息反馈的及时性。信息的及时反馈可以借助有效的虚拟现实可视化技术来实现。当可视化技术将大规模、高维度、多来源、动态演化的高等教育评估数据以图形展现出来时，评估者往往能够直观、快速、精准地洞悉数据中隐含的规律知识和价值，以达到“深入浅出”的评估效果。



信息的及时反馈增强了评估主体与智能设备的虚拟互动，大幅缩短了评估结果的反馈周期，加速了评估信息的整合迭代优化，有助于高等教育教学质量的系统提升。基于虚拟现实可视化技术的交互式展示和高密度信息动态展示功能，评估者不仅能获得教学活动的整体情况，还能对不同教师的教学状态实时跟踪，为教师构建具有个人特色的教学知识体系提供保障，实现细节与整体的同步认知，推动高等教育教学质量的系统化提升。

及时反馈过程中还会不断生成新的数据，且随着数据的迭代优化，教学管理者对教学状态的认知逐渐清晰。教学管理者可以围绕需要改进的教学“问题清单”，凭借数智技术将改进问题系统化为科学问题，进而提出“一对一”最优化解决方案，以促进高校合理定位、分类发展、相互协调、有序竞争，提升高校培养人才、服务社会的核心竞争力，形成“以评促强”新局面。

四、新型高等教育评估体系的建构

深化新时期高等教育评价改革，需要重构高等教育评估体系。建构新型高等教育评估体系应以“瞄准突出问题、遵循教育规律、顺应发展趋势、回应时代需求”为主线，秉持“以评促强、争创一流”的发展理念，探索“平台化”协同慧治的治理模式，创造“教育评估智能体”的超级生态，在评估者与技术之间建立起合乎数字与智能生态伦理准则的良性秩序，不断激发高等教育评估的牵引力，统筹推动高等教育领域综合改革，为建设教育强国开辟新赛道、塑造新动能。

1. 发展理念：秉持“以评促强、争创一流”发展理念，为建设教育强国提供强大动能

首先，应回归高等教育评估本真，推进落实立德树人根本任务。传统的高等教育评估过于专注教育主体的功利性需要，对育人的本体价值重视不够。“唯分数”“唯论文”的评估结果具有一定的片面性，忽视了高等教育作为实践活动的复杂性。因此，高等教育评估应从纯粹的学术标准走向立德树人的育人标准，从学术维度回归到人的维度。高等教育的根本任务是立德树人，这里的

“人”是指马克思所说的“全面发展的人”。影响学生全面发展的因素包括个人特征、院校环境、知识技能发展情况、个人社会化程度等。传统的高等教育评估受限于数智技术发展水平，无法对动态变化的质性数据或隐蔽性数据进行大规模采集。在数智技术驱动下，高等教育评估有望实现全方位、全过程、全动态的立体化数据采集，可有效解决传统高等教育评估中信息量小、失真和结构缺失等问题，有助于对评估对象进行全面价值观照。

其次，需凝聚全链路数字化共识，推动多元评估主体联动协作。信息化建设成为提升评估效果的核心动能，数智技术更是打通了高等教育活动各个环节之间的壁垒，实现了全链路数字化，简化了高等教育活动中信息沟通、流程，节省了传递时间，增强了评估者之间、评估者与评估对象之间的协同效率。同时，数智技术也为高等教育评估活动的参与者搭建了“沉浸式”的评估场景：通过全链路数据的沉淀，高等教育管理者能够及时发现问题，洞察教育变革未来趋势。因此，数智时代的高等教育评估迫切需要达成数字化共识，须打通评估数据全链路，持续拓展高等教育评估的数智化深度和广度，推动数智技术与高等教育评估全系统深度融合，探索运用虚拟现实、数字孪生等技术和资源创设评估场域，在虚拟与现实交互空间中产生提升高等教育评估效果的数智化驱动力。

再次，要坚守科学导向的底线价值，增强高等教育评估牵引力。数智技术催生了“数据模型+人工智能”驱动的高等教育评估范式，这意味着高等教育评估不再受先验假设的限制，也不局限于因果关系探究，而是借助算法算力对“数据洪流”进行分析，从而揭示高等教育数据背后隐含的关联性和规律性，破解了“经验主义”“机械方法”“单一数据”的桎梏。这里的规律性，不仅要体现教育发展规律，还要符合经济社会发展规律，也就是说，要做到合规律性与合目的性的统一。因此，数智技术赋能高等教育评估，是为了更好地挖掘教育价值、实现教育价值、促进教育价值共享，充分发挥高等教育在教育强国建设新征程中的龙头作用。

2.治理模式：探索“平台化”协同慧治，为建设教育强国提供组织保障

首先，需构建平台化组织形式的评估主体，彰显高等教育评估服务导向。传统的高等教育评估积极倡导评估主体的多元参与，但评估主体的结构仍具有科层制的属性。在数智时代，基于数字化枢纽平台，评估主体的纵向层级结构被压缩为纵向联通、横向联动的平台化结构，这就有效解决了各部门之间存在的数字孤岛、协作壁垒、系统割裂等问题。平台化组织呈现出“分布式”“自适应性”“有序化”的特点。在平台化组织中，各评估主体都将是一个独立的网络节点，通过数字平台将自身的职责、义务、需求转化为代码形式的“智能合约”，以形成大规模的社会化协同。在这样的背景下，作为高等教育评估组织者的政府部门将部分权力让渡给社会组织，给予评估实施者“自我增权”，确保了高等教育管理组织的运转更加协调、有序和高效，使得高等教育评估在弱化行政色彩的同时强化了服务功能。



其次，应激发评估主体多维潜能，提升高等教育治理效能。数智技术的涌现和广泛应用为高等教育评估主体认知方式的改变带来了契机。借助数智技术，评估主体不再需要对评估对象的所有信息进行记忆和存储，也不需要依赖个人经验挖掘和判断评估对象行为背后的模式、规则，而是需要掌握数智技术的使用方法来帮助自己突破认知极限、驾驭复杂评估情景、加速知识积累、提高思维效率，从而高效地处理高等教育评估事务、增强高等教育治理效能。

再次，须提升评估主体智能素养，走好高等教育数字治理之路。目前而言，我国高等教育评估主体缺乏智能素养，关键在于数字化评估人才的培养机制不够健全。为此，要加大高等教育数字化评估人才队伍建设，建立支撑高等教育评估事业和学科发展的人才梯队和创新团队，以评估主体智能素养支撑数字化治理能力，最大限度释放数字在高等教育治理中的效能，加快推进高等教育治理现代化。

3. 超级生态：创造“教育评估智能体”，为建设教育强国提供技术支撑

超级生态是一种不同于自然生态的数智生态系统，不仅包含了人类主体，也包含了可以参与行动的智能机器，每一个行动元都将在数字网络中留下大量数据，高等教育评估者被整合到数字生态下的物体系中，成为一个被高度编码和数字描绘的“行动元”。评估者通过数字化的中介，与智能机器和其他评估参与者进行交流、沟通、传递、互动，与人工智能形成以数据交换和智能运算为基础的“教育评估智能体”。

教育评估智能体是在人工智能、虚拟现实、大数据、云计算等数智技术驱动下，由同行专家、教育管理者、教师、学生等多元化利益实体与智能机器相互结合构成的一个立体感知、全域协同、开放共享的智能评估系统。教育评估智能体的构建应以智能平台为支撑，以数据安全为保障，以“人主机辅”为根本，须做到数据开放安全与数据安全防护并重，并坚持“以人为本”的核心立场。虽然数智生态中包含多种行动元，但“人”仍然是高等教育评估的主体，人工智能扮演的只是“私人助理”（Agent）的角色。数智赋能的高等教育评估应遵循教育规律，并立足于经济社会发展对高层次复合型人才培养的现实需求，坚持用“以人为本”的价值理念引领高等教育评估数智化发展，让人工智能通过机器学习、深度学习等方法不断提升自身的可信度、可控性、可解释性和伦理性，促进评估者和人工智能在万物互联的数智生态中不断地交互，使评估者和人工智能的智慧融合在一起，最终形成有效的联合决策。

参考文献略



• 来源：《大学教育科学》，2024年第二期

新时代大学教师教学评价：反思与重构

文 | 赵 锋，华中科技大学

在长期历史中形成的大学教师教学评价制度存在教师教学评价主体不重视、评价标准反科学以及评价效能难以发挥等一系列问题。这些问题的出现与我国大学教师教学评价利益主体构筑的权力格局、科技时代所塑造的组织行为偏好和大学追求的效率最大化路径及其固化密切相关。为适应新时代教育评价改革和教育教学高质量发展需要，我国大学教师教学评价需要从理念到制度进行深度变革，实现基于“主体-目的”框架的发展性评价与基于“教学-学术”框架的系统性评价的制度重构。

关键词：教师教学评价；发展性评价；“主体目的”框架；“教学学术”框架

党的二十大报告提出了“人才是第一资源”的论断，这对大学的人才培养工作提出了更高要求。大学人才培养工作的核心问题在于教学，关键在于形成对教师教学具有积极引导意义的评价体系。2020年10月，中共中央、国务院印发了《深化新时代教育评价改革总体方案》（以下简称《方案》），为新时代教育评价改革提供了根本遵循。《方案》强调，有什么样的评价指向，就会形成什么样的办学导向。因此，我们必须关注新时代大学教师教学评价问题，如此才能将大学教育教学导入高质量发展的轨道。在新的政策与现实背景下，已有研究针对大学教师教学评价问题或进行历史梳理，或总结具体问题，或思考应对策略，均存在系统性不足且缺乏分析深度等问题，针对大学教师教学评价开展制度分析的研究尚不多见，需要我们进行更深入的探讨。

透过现象看本质：大学教师教学评价的制度症结

我国大学教师教学评价已经发展成为包括管理者、同行及学生等多种评价主体在内的系统化制度体系。在历史脉络中逐渐发展起来的大学教师教学评价制度，实际上存在着诸多问题。

1. 知行冲突：教师教学评价主体的不重视

从大学教学主体对教师教学及其评价的基本认知来看，很难说相关主体不重视教学及其评价工作，但由他们的现实行动与实质性投入可知，这种“重视”是值得怀疑的。

大学管理者重绩效、轻教学。大学管理者虽然始终强调教学工作的重要性并制定了相应制度，但这种“重视”未能得到社会的普遍认同，教师教学及其评价活动在其领导的高校建设实践中居于次要地位已成事实。一方面，在“科研至上”的发展取向之下，科研及其转化成果更能表征大学获得的成就和在评估体系中的地位，大学管理者进行业绩评定时所关注的主要是科研与服务指标，而不是教学质量。另一方面，大学教师教学评价缺乏明确的刚性标准，管理者很难真正保障其对教学的资源投入。诚如德雷克·博克（Derek Bok）所言，没人能知道一所大学在培养学生方面与其他大学相比孰优孰劣，更没人知道一所大学在这些方面是否进行了最大的投入。因此，教师教学质量难以测量的事实为管理者利用绩效杠杆进行资源分配提供了可操作空间，他们事实上普遍倚

重绩效显示度更高的职能活动。再者，即使某些大学管理者展现出推动教学评价改革的热情，也有可能遭遇来自一线教师的底部“梗阻”，故而管理者不愿再为此付出成本，更倾向于简单维持现有的教师教学评价制度。

大学教师重科研、轻教学。大学教师普遍在保障教学基本投入的情况下，利用大部分时间沉浸于科研锦标赛中。一方面，大学管理者的考核重点是科研成果，科研表现成为大学教师赢得教职、晋升职称和获取报酬的核心依据，于是教师便将精力更多地投入科研工作中。此外，科研在学术共同体中的认可度日益高于教学，教学甚至被视为学术人员从业过程的“阻碍”。另一方面，缺乏确切的标准表征大学对教学及其评价的重视程度，教师的“重视”很可能浮于表面。在日常工作中，只要大学教师不刻意制造问题，即便在教学方面“不作为”，其后果也并不严重。例如课时数是大学主要的教师教学通用考核标准，很多教师虽然能够达到基本要求，但不排除是以“出工不出力”的敷衍方式完成的。

大学生重学习、轻评教。大学生虽然认同教师教学及其评价对自身发展的意义，但普遍参与的表现仍无法掩盖其积极性不足的问题。一方面，大学生主要关心自身学习任务的完成情况，对教师教学评价缺乏强烈的需求和敏感性，这在一定程度上导致了学生评教具有强制性施行的特点，也说明很多大学生尚未形成重视教师教学评价的意识。另一方面，大学生在接受大学教育前极少对教师进行评分排名，故而即使对教师教学质量有所感知或有意参与教师教学评价，也还未养成正式评教的习惯。再者，学生评教的强制性方式及其效能不佳使大学生的参评意愿趋于冷淡，从而仅将其作为程式化任务来完成。

2.体系脱嵌：教师教学评价标准的反科学

从大学教师教学评价制度的发展历程来看，评价体系虽然始终追求科学化标准，但在无形之中却偏离了引导教学高质量发展的方向。

表1 B大学、H大学不同类型课程的教师教学评价标准

	通用课程教学	实验课程教学
教学态度	教师精神饱满	爱岗敬业,仪表整洁,严格公正
教学实施	讲解思路清晰,重点突出	讲解清楚,认真指导,操作规范
教学方法	合理利用辅助教学手段	利用多种手段和方法
教学内容	准备充分,内容科学实用	科学性,启发性,探索性,实用性
教学效果	效果良好,学生收获大	加深理解,受到训练

大学教师教学评价标准重统一、轻多元。统一的大学教师教学评价标准虽然更具客观性与高效性，但与具体教学实际并不契合。一方面，大学教师教学评价标准千篇一律，忽视了不同类型的教学特点。在“学术漂移”效应之下，层级较低的大学表现出模仿、复制研究型大学教师教学及其评价制度的行为，研究型大学教师的理论化教学及与之相应的评价标准使得应用型院校、职业院校偏离了符合各自人才培养特点的教师教学评价方向。另一方面，统一的大学教师教学评价标准无法准确评价不同的专业和课程。大学制定了教师教学评价制度及其标准，评价主体便拿着统一的评价表对不同类型课程进行评价，缺乏切实的针对性。再者，即便大学针对不同类型的课程制定了不同的教师教学评价标准，但有限的分类标准也不能完全体现实质性区别。本研究整理并分析了B大学、H大学教师课堂教学评价的相关制度文本，其不同类型课程的教师教学评价标准如表1所示。

通用课程教学实验课程教学教学态度教师精神饱满爱岗敬业，仪表整洁，严格公正教学实施讲解思路清晰，重点突出讲解清楚，认真指导，操作规范教学方法合理利用辅助教学手段利用多种

手段和方法教学内容准备充分,内容科学实用科学性,启发性,探索性,实用性教学效果效果良好,学生收获大加深理解,受到训练由表1可知,两所大学课程类型划分并不细致,而且通用课程教学与实验课程教学评价标准虽略有不同,但并未体现它们在教学内容与成果形式、学生主导性发挥等方面的根本性差异。

大学教师教学评价标准重教课、轻教学。目前,大学教师教学评价标准侧重观测教师的教学行为信息,效率虽高但难以完全体现优质教学要求。一方面,重教课、轻教学的教学评价标准强调的主要是教师教的视角之下教学结构设计、教学方法使用等内容,没有触及学生学的主观感知层面。如此标准之下,教师可能教得“很好”,但学生却少有收获。另一方面,大学强制要求大学生去评价教师专业化的教学设计工作,他们可能在无法理解评价内容具体内涵的情况下盲目应对。错位视角下所得到的教师教学评价结果不仅无法呈现以教促学的效果,而且容易引发教师群体的抵触情绪。再者,片面强调“教师教”而忽视与之关联的“学生学”,这样割裂式的教师教学评价标准无法形成革新教学、满足学生需要的反馈作用力,教师教学质量的切实改进更是无从谈起。

表2 B大学、H大学教师教学评价制度中的量化与质性评价标准

类型	院校	具体制度	量化评价标准	质性评价标准
人事考核	B大学	教学系列教师聘任	有任教经历,每学年教学工作量不少于192学时或相当12学分或独立承担4门课程	思想素质良好,教学效果良好,无教学事故
		教师教学奖评选	教龄达到要求,评教分数优异,承担教改项目、出版教材、发表论文等优先考虑	政治立场坚定,师德师风优良,教学效果优异
	H大学	教师年度教学考核	教学总学时、课堂教学学时、本科教学学时及评教分数,承担教改项目等	教学及其改革工作的质量
常态评价	B大学	教师教学奖评选	本年度授课不低于96学时,综合评教成绩居于前列,督导评教成绩不低于90分等	热爱教学,严守纪律,效果优异
		学生评教	评价等级分为A、B、C、D四级,评分区间依次为 $A \geq 90, 85 \leq B < 90, 75 \leq C < 85, D < 75$	对本课程的意见和建议
	H大学	学生评教	评分等级分为优、良、可、差或其他等级形式	对本课程的意见和建议

大学教师教学评价标准重量化、轻质性。目前,大学教师教学评价制度主要是对评价内容进行量化赋分。一方面,量化评价标准在保证客观性、科学性的同时,忽视了大学教师教学及其质量的质性特征。本研究基于B大学、H大学涉及教师教学评价的教师聘任、考核、评奖及评课等制度文本进行提炼和总结,整理得到大学教师教学评价的量化与质性评价标准(见表2)。由表2可知,两所大学教师教学评价标准均主要表现为数字或等级形式;质性指标内容较少,主观性强且模糊不清,在量化评价体系中缺乏可操作性与可比性。这表明教师教学成绩仍是“客观证据”,但这样的结果在表征复杂生动的教学实践时具有很大的局限性。另一方面,强调量化数据比较的评价方式不仅进一步加剧了大学教师重科研、轻教学的倾向,而且因量化评价标准将焦点置于教师教学产生的结果端,忽视了质量真正发生的教学过程本身,也导致了大学教师教学实践评价与教学研究评价的失衡。

3. 执行异化: 教师教学评价效能难以发挥

大学教师教学评价制度在取得一定成效的同时,出现了扭曲制度本意、桎梏群体和个人发展的制度异化问题。

“管理”异化为监督。大学教师教学评价标准及相关管理制度的初衷是确保教学秩序的稳定,但现如今却逐步成为束缚教师教学的工具。一方面,很多大学的教师教学评价制度均强调对教师教学加以规范,使学生获得良好的学习体验。但是评价主体在施行制度时,往往会机械地利用管理标

准框定教师的教学行为，制度的弹性空间在执行中被不断压缩。另一方面，大学教师教学评价理应最大化规范与改进教师教学活动，但各教学主体的协同共管关系在教学督导的教学评价之中被异化为对教师教学信息和行为的“督察”。由此，自上而下的严格监督在使教师自主权受到挤压的同时，也造成了管理者与教师的信任危机。

“服务”异化为管理。大学教师教学评价的人本化服务属性在执行中被弱化，评价制度成为管理部门刺激竞争、问责教师的工具。一方面，量化评价制度与问责制的结合使得为大学教师职业生涯发展服务的聘任、晋升制度等异化为大学管理教师的手段和工具。在某种程度上，教师考评标准成为工作目标，教师专业成长被绩效增长所替代。另一方面，大学教师教学评价制度的重要目标是改进教学服务，但在真正落实过程中，评价结果主要作为教师奖惩的依据而存在。以学生评教为例，其评价结果往往被功利地应用到教师的人事管理之中，而其服务教师教学改进的功能则难以发挥。从完善教师教学到危及教师职业安全的大学教师教学评价功能异化，其对教师的教学热情造成了极大的负面影响。

“互惠”异化为共谋。评价者行使评价权利，被评价对象改进教学，这种本可以改善大学教师教与学生学的“互惠”之举却产生了异化。一方面，就同行教学评价而言，大学试图通过同行听课的方式来实现教师的取长补短，最终促成教师教学水平和人才培养能力的共同提升。但教师群体将打分式评课视为关乎奖惩且不得不应付的流程式任务，基于利害心理而彼此“关照”，同行之间的教师教学评价由此成为互打高分的

“人情往来”。另一方面，具备合理性与科学性的学生评教制度在执行中出现了异化问题。具言之，学生通过评定任课教师的教学成绩对教师形成“制约”，教师进而通过给学生高分或优秀等级予以“回应”。如此，奖惩逻辑之下的学生评教演化成为一种利害攸关、充满博弈的“交易”活动，由“利益交换”带来的“教学相涨”显然无法体现真实的课堂教学质量。



二

主体、偏好与路径：大学教师教学评价的制度分析

W·理查德·斯科特 (W. Richard Scott) 认为，制度包含“为社会生活提供稳定性和意义的规制性、规范性和文化认知性要素，以及相关的活动与资源”。这三类制度要素对应着大学教师教学评价及其相关政策制度体系中的不同内容，是历史制度主义、理性选择制度主义和社会学制度主义等流派的主要分析对象。基于此，在将各类制度要素涵盖的三个新制度主义流派的交流过程中，逐渐形成了一种新的制度分析取向：“利益驱动下的理性、文化形态中的观念和作为历史集装器的制度，在政治过程中都有自己的位置”。这种整合性的分析框架能够更全面地解释制度的形成与变迁问题。

1. 主体利益与权力关系

多主体的利益诉求。行动者具备进行择优选择的理性计算能力，他们通过表明偏好并进行比较来运用自己的理性。一方面，对大学管理者而言，教学的现实影响力很难超越教室的边界，无法为大学赢得广泛的认可，而以科研为基础的学术声誉积累才是利益所系，极力追求科研绩效便成为管理者的理性策略。大学教师的逐利倾向与此同理，更具“流通与交换”价值的科研活动能为他们

带来更高的地位与荣誉。大学生的理性计算主要表现在对课业实用性的衡量以及教师教学评价权利的行使上，很多大学生认为评价教师教学是正当行为，且应使评价结果发挥改进教师教学、提升学习质量的功效。另一方面，在“手段目的”的理性计算思维影响下，理性人会尽可能降低使用制度工具的成本来达成利益最大化的目标。精力有限的管理者、教师和大学生不仅会避免教学及其评价对科研资源与时间的占用，而且会设计和接受统一而简便的评价工具，从而降低自己的参评成本。再者，基于不同个体需要的逐利行为可能会造成集体行动困境，此时的制度选择往往由掌握最高权力的主体作出，并通过组织的层级体系——促进上级决策与下级行为一致性的稳固链条，将其上升至整体利益的层面。作为大学代表的管理者会将大学教师教学评价的制度要求逐级下压，在制度的强制执行中整合教师与学生的利益诉求。

等级观念中的认同。社会学制度主义更关注文化因素的作用，强调为人们所共享的、理所应当的观念赋予人类行为的意义，即通过合法性机制获取认可。这里的合法性机制是指诱导或迫使组织采纳具有合法性的制度结构的观念力量，合法性不仅包括正式规则的作用，而且包含文化观念、社会期待等对组织行为的影响。一方面，基于我国长久存在的“官本位”思想，组织及其成员普遍认可和信任官方部门。在这种观念的影响下，大学管理者严格贯彻教育管理部门的政策指令要求，制定并施行相应的教师教学评价制度，处于基层的师生基于共享的“唯权威部门是从”的文化心理，不仅重视由行政部门供给的制度本身，而且在制度执行中表现出对官方指示的强烈依赖。另一方面，传统的师者权威观念影响至今，合乎这种观念要求的教师便难以接受被学生评价的行为，学生也较少主动地对教师进行评价或一律予以“好评”。再者，师生虽然会在形式与流程上回应上级的强制性评价要求，但共享的人情观念使他们将大学教师教学评价作为“交换人情”的行为。基于此，他们便能在人情网络中维持良好的人际关系，大学生也会在某种程度上将评价教学作为向教师“示好”或“回馈”的社交行为。



稳固的权力格局。制度化组织实质上是一种权力体系，其运行与发展的每个方面和环节都表现着权力的作用，并且深受社会结构的影响。大学教师教学评价制度中蕴含的权力关系在我国宏观管理体制的制约下具有明显的不平衡特征。一方面，国家举办与政府管理大学是我国高等教育的宏观体制背景。在这种管理体制内，各级政府部门是教育资源的主要分配者，并在很大程度上主导学术生产的全过程。大学内部为承接任务而设置相应的职能部门，形成了行政性较强的科层制组织结构，基层教师普遍要接受各行政部门的领导与考核。另一方面，当行动者之间的关系被嵌入一种由权力关系及多种依存关系所构成的不平等领域时，行动逻辑各异的他们若要达到各自的目的，就会对彼此施放自己的权力。在进行大学教师教学评价制度的选择时，管理者在权力博弈中始终处于主导地位，对能够最大化满足其利益需要的制度，“从这种权力系统中受益的、成熟的制度建立者，通常会付出相当大的努力来维持他们的支配权”。于是，大部分师生在一定程度上便成为管理者意志的“代言人”，在大学教师教学评价制度的运行中将管理者的意图予以贯彻。

2.时代风尚与行为趋向

复杂的选择。有限理性的存在，使行动者的逐利行为被赋予更多的解释。一方面，在有限理性使大学管理者和教师无法完美协调教学与科研关系时，他们所能做的就是关心那些并不遥远且能够

把握的利益，至少不会为损害这些利益行事。在这种逻辑下，较为明确的利益所在往往是科研而非教学。此外，人们无法获得目前以及未来的所有信息，很可能基于信息或地位优势去要挟或利用他人。学生评教的“教学相涨”与同行评价的“人情往来”都是如此，师生看似理性的利益计算中充满着非理性的风险。以异化的学生评教为例，大学生虽然获得了短期的显性收益，但无法完全预测发展前景的他们或许会在未来的学习与工作中面临教育质量危机。另一方面，行动者追求的最大化效用不仅体现在物质利益方面，也体现在利他、自愿担负等非物质利益方面。这种拓展的理性观认为，行动者决策理性中的各种计算推理并未排除无知、直觉、诸种情感与情绪的影响。教师群体对职业道德的认同使其不惜以超负荷的支出来保障教学工作量，而不是基于利益计算完全投身于科研活动。部分大学教师之所以表现出明显的科研偏好，可能只是出于对科研活动的热爱，这种无需深思熟虑的选择能满足其心理需要。此外，偏向科研的部分大学教师也可能不完全是出于利益计算，不为物质利益所动的他们始终将教书育人作为核心使命，但不科学的大学教师教学评价制度引起了他们的反感，进而只能在科研导向的评价制度与履行教学职责之间谋求平衡。

科研崇拜中的认同。科研职能在大学发展壮大过程也是人们形成“以科研为尊”的共享价值观念的过程。一方面，社会成员对科技文化的崇尚使其对大学的科研表现愈发关注。这种重视科研的观念在大学排名浪潮的裹挟下被传播和强化到极致，社会上具有广泛影响力的排行机构利用无孔不入的排名榜单信息塑造并强化着科技时代的价值观念：研究型大学是最好的大学，科研是最重要的标准。另一方面，社会评判大学的观念在大学内部得到了强烈回应，大学教师对科研文化的认同深植于学术共同体之中。在他们的认知框架内，科研是大学专业性最强的活动，是展现专业水准和获得学术荣誉的主要途径。这种认知在作为大学教师预备役的博士生群体那里实现了传递，发表文章、参与课题等是其日常话语体系中的高频词汇并主导着他们的职业生涯规划。与科研的发展速度和专业化程度相比，由于“教学是个人化的实践”“教室是私密领地”等大学教师所沿袭和共享的保守观念的存在，作为大学最古老职能的教学不仅没有产生根本性的变革，还因其缺乏组织化、制度化的专业积累机制而备受轻视。

制度环境的影响。历史上形成的社会宏观制度结构会制约集体或个人的选择，对行动者行为偏好的形成具有重要影响。一方面，改革开放以来，1985年在《中共中央关于科学技术体制改革的决定》《中共中央关于教育体制改革的决定》等政策文件出台的背景下，被赋予建设教学与科研“双中心”重要使命的大学，在科技事业发展中的地位得到了加强。在随后的科教兴国战略、人才强国战略与创新驱动发展战略的实施中，科研职能在大学的地位得到了极大提升，科技与学术人才日益成为最重要的发展资源。在这种背景下，处于体制改革浪潮中的大学依靠市场化手段围绕科研项目与学术人才展开了激烈竞争，由此引发了学术劳动力市场秩序的重建，使得高级科研人才的重要性与稀缺性愈发明显。另一方面，国家的科教发展政策直接促成了大学组织制度的变革，大学科研制度化推动了组织用人标准与教师偏好的变化。大学广泛设置了研究所、研究中心等专门机构开展研究工作，并设置专门的科研职能部门进行管理，与之形成鲜明对比的则是教研组织的式微。同时，大学对人才项目的争夺日趋“白热化”，人才的学术等级直接与利益挂钩。大学对在学术劳动力市场中具备竞争优势的科研人才予以高额激励，不仅远远超过了项目本身的奖励额度，而且基于人才



项目的符号效应，高层次科研人才的酬劳远高于普通教师。这种评聘奖励制度成为影响教学与科研关系的最大因素。由此，不同学术等级所造成的落差促使大学教师纷纷投身科研，通过积累学术资本来获取学术荣誉。

3. 结构要素与制度惯性

理性组织的设计。基于明确的组织目标，大学设计了最大化提升组织效能的结构与规则。一方面，组织内部的正式结构形成了自上而下逐级制约的行动秩序。结构正式化就是对约束行为的规则进行说明，确立各角色之间关系的规范，通过标准化要求和管理手段提高行为可测度。大学教师教学评价制度的设计与执行就是基于这样的正式化组织结构，实现了自上而下的问责。这样的正式化结构很大程度上排除了情感和爱好的影响，科层制管理使得一切组织活动都变成可测、可控的，根植于其中的理性主义赋予组织制度以标准化、效率化的鲜明特质。另一方面，明确的效率指标体系是理性组织开展评价的首要选择。科层结构具有超越其他结构形式的纯技术优势，它以效率追求为核心，通过计算来设计路径以最有效达成目的。在近乎狂热的效率崇拜之下，管理者的生产效率分析——对完成生产任务所需资源进行计算与比较以及对所完成任务的测量与评价，成为常态化行为。这也决定了大学教师教学评价制度基于标准化统计手段对教师进行问责的特征。

有效性中的认同。组织及其成员通过满足合法性要求获得资源支持，这种满足程度需要通过有效性来说明。一方面，在大学的职能活动中，能够得到有效说明和积累的部分成为评价大学教师的重点。管理者与教师普遍认同教学与科研都是大学的重要活动的观念，其中教学更是大学确证组织合法性的根基。但教学在表明自身“实效”性方面远不如科研，重视科研逐渐成为追求效益的大学及教师的行动逻辑，科研激励制度的效益合法性在挤压教学道义合法性的同时，也在不断塑造新的社会认同。与此同理，在行动者的共享观念中，教学活动既有共性也有个性，教师教与学生学密不可分，教学活动兼具质性特征和量化属性，然而外显、易于比较和测算的客观指标成为说明和积累效率的重点，量化考评由此成为主导性的大学教师教学评价方式。另一方面，基于组织制度趋同效应，我国大学普遍采用量化评价与排名的教师教学评价制度。所谓组织制度趋同，是各组织为了达到合法性要求，通过强制趋同、模仿趋同与社会规范趋同等机制普遍采用相同或类似的结构与制度。具体而言，我国大学教师教学评价制度既是在政府强力推行大学教师教学评价政策的要求下形成的，也受到同类组织教师教学评价制度的示范性影响，还与大学教师教学评价的专业化、制度化相关。无论哪种趋同机制，其相应制度均主要表现为基于问责的量化评价特征，使得量化评价得以逐级贯彻、广泛施用和趋向专业化。

制度的自我强化。在宏观管理体制环境中建立起来的问责性、量化的大学教师教学评价制度由于路径依赖的存在，在回报递增的过程中不断固化。一方面，在市场化体制改革背景下，我国大学逐渐形成了类似于企业的强化竞争与谋求效益的特点。这种强调显性指标成绩的倾向在20世纪80年代以来蓬勃发展的高等教育质量保障运动中被无限强化，集中表现为“审计型社会”以评价、认证与排名等手段对高等教育进行问责。面对这一挑战，大学管理者也不得不运用从工商业领域引入的指标测量法来开展频繁的考评，由其主导的大学量化教师教学评价体制由此形成。另一方面，大学



教师教学评价制度的正反馈作用使量化评价长期存在。具言之，占据优势地位的决策者要求执行者落实量化评价制度，并且形成了多种具体制度密集交织、协调并行的绩效考核体系。同时，相较于质性的建构式评价，量化评价具有简捷高效等特点，高昂的制度成本使得量化评价难以被完全取代。此外，制度学习效应与调适空间的存在使得评价主体对现有的大学教师教学量化评价制度有所适应并认可其优势，进一步强化了评价主体对这种制度的包容度。

三

从理念到实践：大学教师教学评价的制度变革

进入新时代，高质量发展成为大学最重要的利益追求。大学必须重新思考与确立人才培养在大学中的中心地位，切实重视教育教学及其评价工作，将教师群体的教学活力充分释放出来。面对由主体权力失衡与技术化治理所引发的大学教师教学评价制度困境，在新时代深化教育评价改革理念的指导下，对现行制度进行重构显得极为迫切且重要。

1. 建构基于“主体目的”框架的发展性大学教师教学评价制度

基于我国制度改革的历史经验及成本考量，大学教师教学评价制度变革宜在合理利用现有制度资源的基础上，按照自上而下的强制性变迁与自下而上的诱致性变迁相结合的方式进行，并且遵循注重制度延续性的渐进性变迁思路。

从权力收放走向主体自觉。放权式改革会带来难以避免的“放乱收死”困境，其根本原因在于各级各类主体的主体性权责不明确。因此，我们需要避免权责配置的行政化倾向，遵从法治逻辑来实现行为主体的权责一致。大学教师教学评价制度的主体主要包括作为关键主体的管理者与作为核心主体的教师和学生，他们在评价立场、目的、动机、方式及功能取向等方面均表现出不同特点，其教学评价类型主要体现为管理者主导的问责性评价和师生群体主张的改进型评价。不同主体在教师教学评价中具有不同地位，作为核心主体的师生群体对教学质量的影响最为根本，也是大学教师教学评价中最为重要的主体。如此，大学管理者与教师和学生基于权力主体的主体性重构，激活各主体关于教学的理性自觉，围绕教学评价积极行使权利和履行相应义务，合理表达各自的诉求并为之负责。

从教学二元对立走向以教促学。

无论是传统的“教师中心”观，还是新近的“学生中心”论，在话语表述及与之关联的实践中都可能出现一方压制或遮蔽另一方主体性的弊病。在教学活动的历史与现实中，教师教与学生学从来都是相融的一体化关系，学生真正意义上的学习要建立在教师引导的基础之上，其属性表述为“生为本，师为要”较为适当。基于此，

一方面大学要注重对学生学的评价，这是显示教学质量的根本标志与最终依据。另一方面，由于课堂教学中的大学生学习难以被具体把握和观测，大学教师教学评价制度若要有效推行，就必须关注教师教的方面，尤其要关注教师以教促学的教学表现，使其在教师教与学生学的一致性联结上得到真实的体现，最终促成教学质量的根本提升。

从基于问责的奖惩性评价走向基于合作的发展性评价。大学要在明确各主体权责及地位的基础上，使教学个体的创造性真正得以发挥。一方面，大学要使发展性评价成为主流。大学应在遵循核



心主体（教师与学生）、关键主体（管理者）的主体次序，确保作为课堂教学当事人的师生、同行在得到充分交流的基础上使教学问题得到反馈和解决，实现教师的专业成长。发展性评价应该在大学或院系明确基本规范的基础上，由每位教师在各自课程的教学学期、学年中针对性地开展，并对相关内容进行记录，以供必要的绩效考核之用。这也符合《方案》中“强化过程评价”的要求，如此方能突破统一的量化标准束缚，在尊重多元化教学实践的同时及时改进教学。另一方面，在注意问责限度和频率的前提下，管理者理应对教学活动加以规范，充分考虑教学自主空间内的教学行为复杂性，始终以是否有利于立德树人作为根本的评价标准。无论是教学评比，还是定期考核，大学都既要以资源分配者主导的公平评判为主，又要重视大学教师的接受能力与表达权利。《方案》提出“探索增值评价”，这要求大学在对教师进行考核时，需要将教师任教过程中的教学表现与评价记录进行比较分析，形成肯定其进步的发展性评价结果，使带有奖惩性的教师教学评价真正发挥其发展性功能，将大学教师教学评价活动全部纳入促成教师职业生涯健康发展的轨道。

需要注意的是，目前大学教师教学评价制度所反映的教学价值与教学地位需要进行重新评估与科学定位。扭转教学价值观的关键既在于认识并尊重教学及其功能的专业性与独特性，又在于将教学激励提升到应有的水平。就全球主要高等教育强国的学术聘任状况而言，保持聘任标准的“高门槛”与维持体面生活的“弱激励”是其主要特征，功利化的学术竞争与利益驱动下的高额科研奖励在促成高质量学术成果产出方面有害而无益，从事学术职业的教师或许更加重视创造性成就而非最大化的经济利益。这些客观事实为我国大学重新调整失衡的教学科研资源分配方案提供了可能。我国大学理应在加大资源投入总量的同时，将科研激励资源向教学领域适当倾斜，实现对大学教师教学及其评价的实质性重视。

2. 建构基于“教学学术”框架的系统性大学教师教学评价制度

从经验教学走向学术教学。仅凭个体固有认知与实践经验开展的大学教学，往往顺应教师个人主观意愿而忽视大学生的实际发展需要，致使高质量的创生型教学无法实现。对此，我国大学教师应该避免以主观立场为中心、以自我情绪为主导的经验主义教学倾向，利用学术教学的理念和模式变革和改造保守而低效的经验教学。教师要将大学教学过程视为师生共同体合力创造和探索的过程，使学生在富有高阶性、挑战性的教学活动中积极投入，快速进入深度学习状态，从而最大化提升学习效能。作为一种创新性教学实践，学术教学的开展并不局限于以科研学术著称的研究型大学，应用型院校有应用型学术及其教学模式，其关键在于增加学术厚度与教学深度，使高质量学习得以真正发生。开展学术教学不仅符合大学阶段的教学特点，而且能够切实展现与提升大学教师的教学质量，这对大学教师的教学素养与教学投入是一种考验，能够检验大学教师在真实教学实践中的专业水平，理应成为大学教师教学评价制度的重要内容。



从教学研究走向教学学术。传统教学研究在偏好理论建构与奖评取向的同时忽略了与教学实践的密切联系，致使很多成果的实践价值备受质疑，而基于教学实践的**教学学术**则实现了对一般教学研究的超越。新时代大学教师教学高质量发展的前提是对教师教学工作进行科学定位与评价。为此，大学一方面要重新认识并拓展有效性评价的内涵，将真正生成高质量教学的教学实践作为大学教师教学评价的重要内容。另一方面，大学要积极探索提升教学效果显示度的新理念和新方法，使

教学的有效性积累得到更为确切的说明。按照《方案》中“健全综合评价”的要求，无论是偏向质性的学术教学实践评价，还是易于量化的教学学术成果评价，都是高质量大学教师教学评价的重要方面。就大学教师教学评价现状而言，教学学术成果评价就是“改进结果评价”的体现，需要发挥和强化大学教师教学学术评价对系统化大学教师教学评价的带动作用。在渐进式变迁的制度变革背景下，基于证据的量化评价仍然是目前最具说服力的评比方式，代表教师教学专业化水平的可流通、高标准学术成果能使大学教师教学评价“有迹可循”。同时，高质量的教学学术成果在某种程度上能够代表相应的实践效果，可以作为系统性大学教师教学评价制度的高效度指标。

从偏向量化的片面性评价走向整合教学的系统性评价。大学将学术教学与教学学术联结起来的评价，能从根本上体现其教师优质教学中以生为本、深刻高效的质量要求。一方面，在针对学术教学实践的评价中，大学教师教学评价主体着重考察的是教师教学学术成果在实践中的运用效果。在内容方面，课堂教学实践从根本上决定着教学质量，也是践行立德树人使命的重要方式。因此，评价主体既要根据以学生发展为本的要求对教学内容、方法等开展评议交流，又要对隐性的教学要素进行捕捉、描述和分析。对此，大学可以借鉴国外表现性评价、真实性评价的经验，使大学教师在真实教学情境中的立体化表现得到全景式记录。在此过程中，评价主体要始终根据“以教促学”的教学目标统摄一切要素，尤其要强调对学生深度学习的促成。需要注意的是，大学教师教学评价需要规避技术主义倾向，将教师作为完整的教育实践主体来看待和分析，将其对学生在学业、道德人格等方面的全面影响作为重要的衡量标准。另一方面，大学教师教学评价主体开展教学学术评价，首要考察的是教研成果的问题来源与实践成效，要始终坚持以应用价值作为第一评价标准。在内容方面，评价主体要注意教学学术作为行动研究的属性，其研究选题、程序、方法等都应作为评价分析的对象，以便对学术研究的科学性进行判断。需要说明的是，教学学术目前仍没有公认的确切评价标准，无法严格限制的口径以及教学活动的复杂性也使得教学学术评价对象种类繁多。其成果既包括与科研学术类似的著作、论文、课题项目等，又包括教学总结、优质课程及课件等文本及音像资料，这些对象均可作为教学学术成果加以评价。

需要注意的是，大学教师教学评价制度的重构与实施是一项系统性工程，需要开展与之匹配的制度环境建设工作。通过组织及其文化变革、配套性制度完善与智能技术赋能等举措，大学要构建全面的支撑体系与保障机制，最大化减少评价主体之间的利益冲突并降低其参与成本，从而实现评价主体的深度参与和通力合作，避免出现执行偏差及制度异化等问题。

来源：《高校教育管理》，2023，17(4)：114-124。



破局、重构与协同：

生成式人工智能视域下的“破五唯”审思

文 | 肖 璐，曲阜师范大学教育学院

以ChatGPT为代表的生成式人工智能以其革命性的技术创新对当下以“破五唯”为旨归的高校学术评价改革产生了诸多影响。如何充分利用人工智能技术带来的便利推动当下的“破五唯”改革，以及如何从制度、技术和观念等领域加以改进，规避人工智能带来的风险与隐患？通过梳理生成式人工智能的技术变革特质，从“破”与“立”两方面剖析其作用于高校学术评价改革的机制原理，并在此前提下探讨学术评价改革与最新智能工具协同发展的优化理路。

关键词：“破五唯”；生成式人工智能；学术评价改革；ChatGPT；高校

一、“破局”：生成式人工智能赋能“破五唯”的逻辑理路

2018年9月，习近平总书记在全国教育大会上作出“破五唯”指示后，《关于深化项目评审、人才评价、机构评估改革的意见》《关于破除科技评价中“唯论文”不良导向的若干措施（试行）》《关于完善科技成果评价机制的指导意见》等文件相继出台。2020年10月，《深化新时代教育评价改革总体方案》正式出台，将“破五唯”写入文件，明确了评价改革的内容和方向，拉开了新一轮高校学术评价改革的大幕。各地通过开设“绿色通道”、设置“破格”条件、推行代表性成果评价，降低论文在职称评审中的比例，将职称评审向教书育人实绩突出的教师倾斜。2022年12月，深化新时代教育评价改革工作推进会召开，再次对“破五唯”的具体内容和方式进行细化。

然而时至今日，高校评聘环节中的“五唯”痼疾依旧难以根治。归根究底，对竞争性学术资源的争夺使得学术研究与学术评价的同一性被割裂，学术研究与利益分配直接挂钩导致高校和教师群体困囿于对短期个体利益的迷思，无法超越功利化的工具理性，过多关注外在价值而忽视了内在价值。在资源分配和权力分割的博弈中，科学性、权威性和简明性便成为分配和管理资源的唯一准绳。“五唯”评价的数量化、外显性、高效率能充分发挥评价的工具效用，而分类评价、代表性成果评价、长周期评价等改革新政由于评价数据容量、时效和分析手段的限制，难以在高校学术评价改革实践中得到切实运用。“破”“立”两面受阻导致“五唯”之风未止，甚至有愈演愈烈之势。因此，“破五唯”首要任务是“破”，其次是“立”。生成式人工智能依托大数据、高算力、优算法所衍生的技术变革特质改变了原有的科研产出节奏，加快了学术成果的产出效率，迫使学术评价体系修正学术成果的认定标准，回归“以人为本”的价值取向。尤其是高拟真度的机器生成内容消解了“论文”作为高校学术评价重要指标的权威性与适用性。正是这些技术变革将学术评价从标准化计量困境中解放出来，进而形成对“五唯”体系的破解之势。生成式人工智能赋能“破五唯”的逻辑理路见图1。

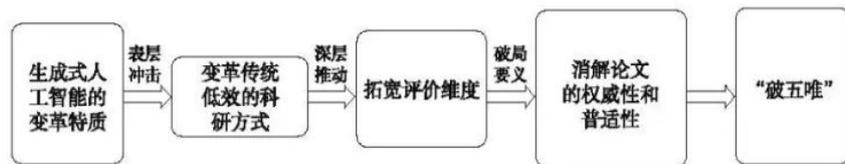


图1 生成式人工智能赋能“破五唯”的逻辑理路

1.表层冲击：人机交互模式与机器深度学习改进了传统低效的科研方式

生成式人工智能在人机交互模式和机器深度学习上的重大变革改变了传统低效的知识采集和科研方式，提升了科研工作者的工作效率和产出质量，改变了“五唯”赖以生存的学术土壤。人们不再需要输入代码或通过浏览器来获取某一学科领域的知识内容或某个管理系统的数据分析结果，只需通过对话框提出问题，智能机器便能根据算法推荐和用户需求调取并整合全网大数据，不断调整答案直至给出更为准确和全面的结果。

生成式人工智能不仅开创了人机交互的新格局，在知识学习和迭代能力上也远超传统人工智能水平。与此同时，基于人工神经网络的发展，研究人员开发出生成式预训练转换模型和利用人类反馈的强化学习训练方式，推动智能机器在与人的互动过程中不断沟通和学习，扩充数据、提升性能、改善应答结果。生成式人工智能自学习、自成长的知识积累模式使自身的学习能力和自我更新能力得到不断强化，知识积累和更迭速度更是达到人类难以企及的高度，知识生产将不限于教师、科学家、工程师等群体和特殊职业者。当人类遭遇人工智能在知识学习能力上的全面碾压时，人类之于智能机器的优势何在，何为高质量学术成果，成为学术评价改革亟待回答的问题。这个问题的回答建立在对人工智能时代学术评价内容维度的重新考量上。

2.深层推动：学术产出鸿沟拓宽了评价维度

无论是创造性思维还是学术责任，都属于量化评价难以简单测量的范畴。“五唯”所体现的绩效主义和科学主义在这两个重要维度的计算和呈现上无计可施。评价维度的与时俱进使“破五唯”有了可以为之的理由，而对论文这一高校学术评价权威指标的消解则将“破五唯”立于不得不为的境地。



目前，人工智能撰写的论文不仅能够鱼目混珠，还可以修饰文本语言，生成妙笔生花、文采斐然的文字，甚至能在未来全面取代人类进行论文的“无性繁殖”。生成式人工智能在消解学术论文权威的同时，能够取代人类从事部分学科尤其是人文社会科学专业的论文写作。面对新一轮数字智能技术的革新之变，“唯论文”势必导致部分学科式微，思辨类、综述类的学术论文和科研成果将成为机器写手泛滥的重灾区，这会扰乱现有的学术生态体系，引发一系列学术伦理问题。

在这场技术革命的洪流中，作为高等教育改革“指挥棒”的学术评价肩负着助力国家在人工智能研发竞赛中实现弯道超车的重任。然而，在“五唯”体系的裹挟下，论文发表、课题申报成为高校教师就业和升迁的唯一通路。“五唯”构建的学术评价空间无法公正准确地测评数字人的创新能力、学术责任履行情况和人机协作沟通能力，也无法使人类在与机器比拼知识学习和学术创作的竞赛中获得

先机，亦无法推动人类创造力的发展与自我价值的实现、恢复学术评价与学术研究的同一性，因此势必被时代和国家所淘汰。

3.破局要义：高质量人工智能生成内容消解了论文的评价权威

无论是创造性思维还是学术责任，都属于量化评价难以简单测量的范畴。“五唯”所体现的绩效主义和科学主义在这两个重要维度的计算和呈现上无计可施。评价维度的与时俱进使“破五唯”有了可以为之的理由，而对论文这一高校学术评价权威指标的消解则将“破五唯”立于不得不为的境地。

目前，人工智能撰写的论文不仅能够鱼目混珠，还可以修饰文本语言，生成妙笔生花、文采斐然的文字，甚至能在未来全面取代人类进行论文的“无性繁殖”。生成式人工智能在消解学术论文权威的同时，能够取代人类从事部分学科尤其是人文社会科学专业的论文写作。面对新一轮数字智能技术的革新之变，“唯论文”势必导致部分学科式微，思辨类、综述类的学术论文和科研成果将成为机器写手泛滥的重灾区，这会扰乱现有的学术生态体系，引发一系列学术伦理问题。

在这场技术革命的洪流中，作为高等教育改革“指挥棒”的学术评价肩负着助力国家在人工智能研发竞赛中实现弯道超车的重任。然而，在“五唯”体系的裹挟下，论文发表、课题申报成为高校教师就业和升迁的唯一通路。“五唯”构建的学术评价空间无法公正准确地测评数字人的创新能力、学术责任履行情况和人机协作沟通能力，也无法使人类在与机器比拼知识学习和学术创作的竞赛中获得先机，亦无法推动人类创造力的发展与自我价值的实现、恢复学术评价与学术研究的同一性，因此势必被时代和国家所淘汰。

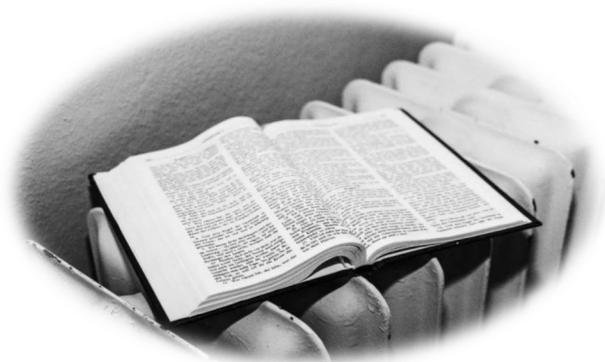
二、“协同”：人机协同“破五唯”的发展向度

生成式人工智能技术在“破五唯”上潜力初现，但是在制度、技术和应用环节上存在的风险和不足使其全面介入评价领域备受业界质疑，目前的学术评价体系还无法适应新一代人工智能的迅猛发展。制度的缺失、机器写手辨识工具的落后以及评价参与方工具使用能力参差不齐将极大制约新技术在学术评价领域的广泛使用，甚至有可能加剧学术造假、伦理失范等评价乱象的泛滥。为了进一步提升生成式人工智能赋能学术评价改革的成效，相关管理部门、高校、评价机构和智能科技研发公司需要在新技术与旧制度磨合变革的过程中与智能机器协同共进、提质增效，不断调整战略和政策，攻克技术难关，推动结果性评价向数字化评价、人机协同评价转变。

1.制度协同：加强政策引导和机构监管，规范智能工具使用行为

智能科技的高速发展呼唤国家出台相关法律法规或规章制度，规范人工智能的研发内容、发展速度和应用范围。就高校学术评价而言，目前可知的知识侵权、学术不端、学术悖德等弊端初现端倪，未来可能还会产生更多与现今的法律和政策相悖离的问题；人工智能引发的意识形态、伦理道德和权力转移等问题同样需要在政策层面进行管控和引导。主管部门亟须加强政策引导和机构监管，规范生成式人工智能在评价体系中的引入机制和运用边界，促进多方协作的学术共同体的建设。

在宏观指导层面，教育和评价主管部门需要制定高校学术评价视阈下生成式人工智能相关法律和管理制度，让智能机器促进评价公平，而



非滋生权力腐败，规避其背后暗含的政治风险、法律风险和伦理风险，但总体来看还只是纲领性指导文件，尚未细化到生成式人工智能监察和认证程序、人工智能产出内容指导规范等操作环节，涉及高校学术评价相关领域的政策和法律文件还有待完善。国家必须从制度管理层面制定数据和信息聚类、分级和评估的管理办法，加强对科技公司的信息监管和研发规制，防范因数据扩充所引发的知识侵权和信息失真，或因地域、国家、文化差异导致的价值观错误；制定学术不端界定标准和惩罚机制、人工智能撰写论文使用指南，引导学术圈层规范使用生成式人工智能，规避可能的违法行为、学术失范和伦理危机；制定生成式人工智能评价系统的使用机制、预警措施和管理办法，协调机器评价与人工评价的权力争夺和操作冲突，确保智能机器能够真正长期服务于评价工作而非成为不可控因素。

在职能协同层面，加强各种各样的教育和学术共同体建设是提升政策效能的一项重要举措，因此，高校学术评价应当坚持以政府牵头、高校主导、出版行业引领三方协作的模式开展评价活动。具体而言，教育评价机构需要尽快设立专职部门负责人工智能造假或侵权事件的处理和防范工作，通过资金、项目等资源的合理配置引导人工智能相关学科规划、发展，鼓励其他学科积极参与数智社会的科研创新和人才培养。高校的学术评价改革应致力于重塑其主体地位，积极响应数智时代的国家发展需求，团结各种合作主体，通过有组织科研发挥决策咨询和学术引领效用，以引领学术科研与数智时代同频共振。出版行业应探索生成式人工智能视域下思想库、智库型媒体的转型之路，通过革新论文发表形式和内容，助力高校教师挣脱“科研KPI”“学术GDP”等数字规训，重新界定与时代相呼应的知识素养和技能水平，助力当下的学术评价改革。

2.技术协同：推进基础建设和学科发展，提升智能工具研发水平

生成式人工智能反制技术和基础设施缺陷制约了其进一步发展，只有依托国家、高校和科技企业共举共建，提升国家自主研发实力，才能推动形成融合数智技术的学术评价体系。目前，生成式人工智能的监督和反制软件尚处于研发阶段，Turnitin、Illuminart、Smodin、

Copyleakes等任何一款人工智能检测软件都无法保证检测精准有效。智能机器反控制技术的缺失势必造成学术造假泛滥、学术审查虚弱、学术公信力丧失等严重后果。因此，加快研发有效的

检测方法和工具已不容置喙。在政府定标准、搭平台，企业做产品、保运维，学校买服务、建资源的总体思路下，国家应当鼓励信息化企业参与评价信息化生态系统建设，为高校学术评价提供建设、维护、安全管理等方面的支持。中国知网、玛格泰克等论文查重机构、投稿软件开发商必须加快并加大对机器写手捕获软件的研发力度，在学术不端检测系统、审稿系统上增设人工智能撰写风险提示选项，助力评价机构和行业专家增强辨识能力、严把质量关。与此同时，相关部门要规范和协调科技公司的商业运作和竞争行为，促进生成式人工智能科技在政府指导下有序、健康、持续发展，提升人工智能技术研发的协同度和专业化水平。

算力发展所依赖的芯片、计算处理器亦是技术层面的难关。随着人工智能预训练大模型规模的持续增加，算力基础设施建设也成为国际竞争的关键要素。为了满足发展人工智能和机器深度学习的硬件需求，国家和地方政府应当积极推进智能计算中心建设。这不仅需要国家在财政上给予支持，更需要从学科发展规划角度加大对光子计算处理器、硅光人工智能芯片的研发力度，大力支持硅基光电子技术学科发展，推进与光电神经网络相匹配的算法的演进；建立统一的教育数据、技术、接口、认证等标准体系，实现数据归集和技术整合，并在标准规范的指引下推进评价数据治理，实现



“一数一源”；加大对相关学科的支持力度，打破学科壁垒，推动材料科学、信息技术、光电技术等领域深度融合；增设人工智能研发学科，提升高校人工智能科研创新能力，培养人工智能研发人才，并制定相关政策以留住塔尖人才。

3.应用协同：改进思维模式和工作方法，增强智能工具运用能力

未来社会对人才和技能要求的变化，尤其是在机器对人的可替代性增强的情况下，人类需要培养自身的创造能力，储备跨界技能。简而言之，评价工作者必须改变传统单一的工作方式和思维模式，提升高校学术评价主体的智能工具应用水平，促进评价主体的能力培养和专业发展。

评价主管部门应落实数字素养培训计划，提升智能技术专业化水平。相关管理部门应当通过人机交互式的主题培训、专题讲座和考评考试等方式帮助评价主体获取人机对话的学习经历和体验，掌握与机器对话的路径和策略；从数字化意识、数字技术知识与技能、数字化应用、数字社会责任、专业发展五个维度加深对数智技术的理解、学习、运用和研究，明晰生成式人工智能技术在教育领域的使用价值，提升评价主体的生成式人工智能知识专业化水平和运用研发等数字能力，最终获得思维能力、专业能力与应用能力的延展与跃迁。数字素养再教育的具体落实还需注意平衡地区差异，相关部门通过改进培训方式、扩大培训范围、增设实操性内容等方式提升数字素养教育培训的可操作性和针对性。

高校学术评价主体应根据业务条线需要动态调整评价策略和方法。学术期刊尤其是高水平期刊要善于辨识投稿作品是否为人工智能机器所生成。当前，生成式人工智能仅能依照人类固有的学术写作逻辑和框架进行论文构思和撰写，尚未能给出引文的正确出处。布兰科冈萨雷斯（Alexandre Blanco-Gonzalez）等发现，ChatGPT撰写的论文正确引用仅占6%，这就给甄别学术造假提供了契机。期刊编辑要筑牢专业技能地基，以踏实敬业的态度考据引文出处以提升稿件识别能力，突破以往固定的思维模式，以内容的创新度和贡献度作为论文的评价标准；评价组织机构要注意对大数据和人工智能的运用不能流于表面，要充分发掘数据背后的价值，通过对数据的聚类和分级改进评审流程，依托新型人工智能技术发挥新评价方法在融合量化指标和质性分析方面的优势。

结语：生成式人工智能时代的诸多机遇和挑战要求我们在当下应促进多元主体有序协调，形成制度变迁合力；加强教育评价制度内外部互动，把握改革时机；培养能独立思考，并具有正确价值判断能力的人。我们不能简单将生成式人工智能等同于一种新型便捷的教育评价手段，作为对原有评价模式的补充，而应当将人工智能教育范式融入教育理念、教育教学体系中。因此，在生成式人工智能的图景中，高校学术评价变革并非单纯的技术应用工作，必须预先研究、整体规划、统筹实施、因地制宜、扬长避短。面对新技术带来的教育挑战与潜在风险，改革高等教育评价体系的应对之举既是时代发展的必然选择，也是历史和逻辑的应然理路。

来源：《高校教育管理》，2024，18(2)：53-64.

