

# JCSEE 元评估标准中“评估问责”属性的探析

刘苹苹

(上海市教育评估院,上海 200031)

**摘要:**元评估是对评估的评估,是评估设计和评估进程中的重要环节,“评估问责”是2011年美国教育评估标准联合委员会(JCSEE)修订出版的《方案评估标准》,即元评估标准的第五个属性。文章从分析中国教育评估和教育元评估实践的现状入手,提出了中国教育评估需要教育元评估的紧迫性;分析了JCSEE元评估标准的“评估问责”属性及其对应的三项标准的内涵;并在此基础上结合实际,思考如何开展适合中国国情的教育元评估,以供教育评估机构和教育评估者参考。

**关键词:**元评估;元评估标准;“评估问责”属性及标准

**中图分类号:**G647 **文献标识码:**A **文章编号:**2095-3380(2020)02-0081-06

## Exploration on the Attribute of Evaluation Accountability Presented by JCSEE

Liu Pingping

(Shanghai Education Evaluation Institute, Shanghai 200031)

**Abstract:** The meta-evaluation is defined as the evaluation of evaluation, and is a key step in evaluation procedure. Evaluation accountability is the fifth attribute of meta-evaluation quality according to The Program Evaluation Standards: A guide for evaluators and evaluation users which was edited by JCSEE in 2011. By referring to the evaluation accountability standards and considering the actual situation in China, this paper points out the need of meta-evaluation in China's education evaluation, deals with the principles and standards in education meta-evaluation, and identifies the key points in the implementation of education meta-evaluation. These may provide references for evaluators and institutes of education evaluation.

**Keywords:** Meta-evaluation; Standards of meta-evaluation; Attribute and standards of evaluation accountability

在现代管理理论中,“评估”始终被作为保障质量、推动发展的重要手段。评估本质上是对评估对象进行一种价值判断,它能使投资人明晓它投入的效益,让被评估者发现自身的位势和发展方向,让社会各界把握信息、获得决策的参考。自20世纪80年代以来,随着高等教育走向大众化,教育质量保障问题愈发得到重视,教育评估因此而得到迅猛

发展;90年代,作为改革开放的成果,我国开始了教育评估活动,建立了教育评估机构,培育了一批教育评估人才,教育评估成为推进中国教育事业进步不可或缺的动力。

### 一、教育评估需要教育元评估

评估作为一种社会活动,它本身也具有价值,

因此也应该对它进行价值判断,即评估本身也需要评估,这种评估被称为元评估。据考证“元”这个字的释义来自希腊文“meta”,相当于英语的“post”,这里的“元”是标志研究层面转变的一种哲学语言。<sup>[1]</sup>教育元评估就是对教育评估的批判与反省,通过教育元评估这一特殊的教育评估活动,实现对教育评估活动的监督和规范。教育元评估是要对进行中或已经完成的教育评估活动进行价值判断,对方案设计、过程实施、结果运用等要素进行稽查,寻找评估偏差,向公众反馈信息,从而进一步保障教育评估的规范有序进行;同时通过对教育评估活动的系统评价,帮助教育评估主体检测和发现问题,了解教育评估的局限和不足,以达到促进教育评估质量提升的目的。教育元评估可以是发生在评估实施的进程中,时刻对评估项目的进展给出评判,即形成性元评估;也可以是发生在评估实施完成后,对评估项目进行总结和判断,即总结性元评估。

教育元评估的发展滞后于教育评估,是教育评估发展到一定阶段的必然产物,也是教育评估专业化程度的一种体现。

教育元评估诞生在美国。美国的现代教育评估起源于二战后,20世纪60年代,各种形式的教育评估活动层出不穷,评估结果也是五彩纷呈。当时,在教育内部,无论是学校管理人员还是教师学生,都或多或少地对教育评估有所非议,同时社会上的投资人、家长和企业也对形式各异的评估结果不甚满意,教育评估需要改革,教育元评估应运而生。首先是西密歇根大学的研究者提出要对教育评估本身进行评估,在他们的倡导下,1975年“美国教育评估标准联合委员会”(JCSEE)成立,<sup>[2]</sup>美国的很多教育组织,如美国教育研究协会、美国全国教育测量学会、美国评估协会等,都是该委员会的成员;之后,加拿大的教育团体,如加拿大教育研究协会、加拿大评估协会等也加入了该委员会。JCSEE制定教育评估活动的标准,或叫准则,供一切与教育评估项目有关人士对评估做出评判。JCSEE的成立是教育元评估正式进入社会视野的标志,毋庸置疑,它的诞生是出于教育评估发展的需要。

我国的教育元评估至今还处于学术研究期,还未有专门机构从事教育元评估活动,也未颁布过元

评估标准,但社会已经产生了对教育元评估的需求。例如,我国首次“全国本科教学水平评估”后,社会上质疑声不绝,无论对评估方法、评估指标、评估结论还是评估绩效都广有意见,<sup>[3]</sup>这些不足不能不归结于教育元评估的缺位。又如“双一流”建设评选结果发布之后,社会上出现了一些质疑:评估方法是否得当?评估过程是否公正?评估结果是否合适用于决策?也有学者分析指出,我国第四轮学科评估的局限性主要体现在评估方案中缺乏元评估环节。<sup>[4]</sup>

如果说美国教育评估的发展已经推动了教育元评估的盛行,那么中国教育评估日益显露出来的问题正在敦促教育元评估的诞生,教育评估需要教育元评估。

## 二、JCSEE元评估标准及其“评估问责”属性

### 1. JCSEE和元评估标准

美国教育评估标准联合委员会(JCSEE)的职责是研制对教育评估进行评估的标准,即元评估标准。JCSEE实行团体成员制,团体成员派出的代表和少量委员会聘请专家共同制定教育元评估标准。1981年,JCSEE在西密歇根大学研究人员提出开展教育元评估的基础上,提出了由有效性(utility)、可行性(feasibility)、适切性(propriety)、精确性(accuracy)四个衡量评估质量维度所构成的教育元评估框架,并用质量属性(attribute of quality)来称呼维度,每个质量属性均设置了若干个标准(standard)。这个标准制定后立即得到美国国家标准学会(ANSI)的认定,并被视为全球第一个相对完整的教育元评估标准。1994年,JCSEE对标准做了修订,仍保留原有的四个属性,但对各属性所含的具体标准做了大量的甄别和补充,有效性、可行性、适切性和精确性四个属性各有7个、3个、8个和12个标准。<sup>[5]</sup>JCSEE的第二版标准在美国和加拿大得到广泛应用,对规范教育评估活动、提高教育评估质量起到了决定性作用,且很快在欧洲和全球其他地区流行。

2011年JCSEE颁布了第三版标准。JCSEE这样描述定稿过程:“有50多个参与者和用户在定稿之前试用了此标准并提出建议,2008年由独立的专

家组审阅了这个标准终稿,之后经过45天的公开听取意见阶段,最后提交美国国家标准学会(ANSI)审查通过。”<sup>[6]</sup>与第二版相比较,第三版标准作了很大变动,虽然仍保留了原来四个属性,但各属性所含具体标准分别调整为8个、4个、7个和8个,且对每个标准所要求考核的内容及应达到的程度做了详尽说明,更重要的是第三版增加了第五个质量属性“评估问责(evaluation accountability)”。调整方案是将第二版中第四个属性“精确性”中“文档”和“元评估”两个标准抽出,纳入新增的第五个属性“评估问责”,其对应的三个标准分别是“评估文档(evaluation documentation)”、“内部元评估(internal metaevaluation)”和“外部元评估(external metaevaluation)”。第三版标准要求将“评估问责”属性贯穿在其他四个属性的实现过程中,其强调的内涵是,没有问责就无法真正保障评估质量。需要说明的是问责的核心是讨论而非责问。<sup>[7]</sup>

## 2. “评估问责”属性及其三项标准

在第三版标准中,JCSEE对“评估问责”属性的三项标准做了如下诠释。<sup>[8]</sup>

**评估文档:**评估要全面系统记录和保存经协商确立的目标和实施方案、操作流程、评估数据及评估结果。

**内部元评估:**评估者应使用本标准或其他可用标准对评估设计、实施过程、信息采集和评估成果问责。

**外部元评估:**方案评估赞助者、委托方、评估者和其他利益相关者应积极推行使用本标准或其他可用标准开展外部元评估。

那么,JCSEE为何在第三版元评估标准中新增“评估问责”属性,并将其内涵调整为以上三个标准呢?

首先是来自社会发展的要求。在西方政治发展中,作为现代民主治理的标志和公共管理的核心概念,“问责”一直是当代西方学术界关注的焦点。1986年,美国第二架航天飞机“挑战者号”失事,启动了行政问责,问责引发了群众性的关注,最终导致失责官员的下台。<sup>[9]</sup>此事件之后,理论界出现了“对‘问责’功能、方式的研究,实践派则将‘问责’作为重要监督手段引入包括高等教育在内的各个领

域,高等教育改革的政策和文献中频繁出现“问责”以及与“问责”密切相关的“质量”“绩效”“结果”等术语。斯克里文曾指出,元评估是对评估系统或工具的评估,旨在服务决策和问责,<sup>[10]</sup>实际上反映了将“问责”引入元评估是顺应社会发展适时的举措。

其次是来自教育评估经验。评估所涉及的利益相关者总有着不同的需求。例如说到有效性,投资人和校方对于资金使用很可能有不一致的看法,从而导致对某个评估项目的目的、方法和评估结论的接受程度不一样,这种分歧会影响评估结论的落实,从而影响评估的价值,因此需要在评估设计时组织各方进行协商,以便在最大程度上达成一致。这种协商可以运用到“评估问责”的外部元评估中。除了有效性,其他元评估标准的落实也会产生外部协商问题,因此外部元评估非常重要,它被列为元评估的一项内涵,成为“评估问责”中的独立标准。

由于文化、教育、习惯等主观意识形态会影响评估者对评估目标、方式和结论表述的认识,而个体的主观形态是多样化的,价值观也可能有差异,这必然导致评估者对同一事实的不同看法,有时个体差异会对评估结果产生很大影响。为了加强评估结果的客观性,内部交流是十分必要的,统一认识才能够使评估准确和有效。因此内部元评估也被列为元评估的另一项内涵,也成为“评估问责”中的独立标准。

在ISO标准中,文档是第一位的,质量保障的第一步是将所见所闻、所想所说、所作所为记录下来。文档不只是一个记录,也是众多相关者行动的准则,更是评估问责的依据。JCSEE在第三版标准中明确指出,评估文档要记录每个具体标准是如何被选择并实施的,为达到有效性和高效率是如何平衡协商的,对评估中可能遇到的不可避免的一些局限性是如何沟通的等等。虽然专业评估者对于“评估问责”有特殊的职责,但来自利益相关者的共同努力才能保障高质量的评估,就需要加强落实评估文档工作。因此,JCSEE将评估文档作为“评估问责”属性的第一个标准,说明文档的规范和完整在该属性中的首要地位。

此外是为了适应教育的发展。教育的大众化和全球化使得教育相关团体越来越多,教育评估内

容更多元化,成员构成更细化,对教育评估的看法也更多样化。因此,加强教育评估的质量保障,推动教育元评估的发展,评估问责成为有效手段。

### 三、探索适合中国国情的教育元评估

JCSEE强调其元评估标准主要适用于美国和加拿大,建议其他国家或地区的评估者应慎重考虑当地文化对它的可接受性和有效性。这表达了对客观实际的尊重,体现出一种科学态度。那么,把JCSEE元评估标准与新时代中国特色社会主义的教育结合、与文化背景相融合,建立扎根中国大地的元评估标准,是一个迫在眉睫的课题。

#### 1. 中国迫切需要开展教育元评估

推进教育治理体系和治理能力的现代化需要教育元评估。党的十九届四中全会明确了推进国家治理体系和治理能力现代化的总体目标。教育部部长陈宝生发表了署名文章《推进教育治理体系和治理能力现代化》,强调教育治理体系是国家治理体系不可分割的重要组成部分。国际经验告诉我们,开展问责是增强社会治理能力的有效手段,美国从问责制度兴起到将“评估问责”引进教育元评估标准就是有力的证明。因此,要实现教育治理能力现代化,落实和推进中央的部署,教育评估是重要手段之一,而要提高教育评估质量,教育元评估不可或缺。

从国内组织开展教育评估实践的现状看,需要开展教育元评估,建立教育元评估标准。纵观三十年来我国教育评估实践和经验,评估一般只针对评估客体,忽略了对评估主体的评估,缺乏对评估本身的价值考量。因此,开展教育元评估理论研究,建立教育评估问责制度是中国教育评估事业与国际接轨,赶超国际教育发展的重要举措。

从教育评估机构专业化发展的角度看,需要建立教育元评估制度。教育评估机构作为教育评估的组织和实施者,改革的使命责无旁贷。从加强自身专业性发展的角度出发,迫切需要国际教育元评估的成功理论和实践经验为我所用,致力于构建适合中国国情的教育元评估制度,比如可借鉴当今国际高校“自愿问责”机制,<sup>[11]</sup>对教育评估机构的评估行为开展内部质量保障的探索。

#### 2. 探索具有中国特色的教育元评估

新时代新挑战,当面对中国教育现代化的目标、立德树人落实机制、教育评价改革等重大课题,需要结合国家教育方针、政策和教育评估实际,探索具有中国特色的教育元评估,发挥元评估功能,提高评估的可靠性,确保评估结果公信力,改进提高评估质量,办好让人民满意的教育。

一是要围绕立德树人这个教育的根本任务。党的十九届四中全会《决定》中明确指出要完善立德树人的体制机制,因此,我们要将立德树人理念渗透到教育工作的各个方面各个环节。因此,实施教育元评估要立足基本国情,遵循教育规律,对教育评估进行评估,确保培养合格的德智体美劳全面发展的社会主义建设者和接班人。

二是要建立正确的价值引领机制,要树立正确的价值取向,将“立德树人成效”作为评估的根本标准。<sup>[12]</sup>教育元评估是教育评估的一种特殊形式,要确保价值取向没有偏差。新时代教育评估价值取向要落在促进“人的全面发展”的教育初心上,要设计将正确的价值观“内化于心,外化于行”的具体内容、路径和方法,要有考量价值导向达成度的评价标准。培育和弘扬社会主义核心价值观,才能落实好元评估的根本任务。

三是要在破除“五唯”问题上打硬仗。教育部部长陈宝生在2019年全国教育工作会议上指出,“唯分数、唯升学、唯文凭、唯论文、唯帽子”这“五唯”是当前教育评价指挥棒方面存在的根本问题。因此,需要开展教育元评估,对教育评估进行评估,以全面梳理、分析、评估教育评估的现状、问题与原因,分层分类研究提出改革的思路举措。要创新思路,从教育规律和人才成长规律出发,以立促破、破立结合,要在“唯”与“不唯”之间找到平衡,设计科学的评估指标体系。

四是要理论联系实际坚持知行合一。首先要顾及现状,中国的教育评估起步较晚,教育元评估几乎还未正式开展过,尽管零星出现过一些论文,比如近年来有学者从元评估的视角对审核评估和学科评估进行探讨,但研究还不成体系。因此,我们必须学习国际经验,尤其是要借鉴美国在教育元评估方面的成果。其次要坚守中华文化,发扬尊师

重教、教学相长、学以致用等教育传承,并融合到教育元评估对评估目的、评估设计和评估结果的价值判断中,形成自己的特色。

### 3. “评估问责”属性在中国教育元评估中的应用思考

关于JCSEE元评估标准中的“评估问责”属性及其三个标准,邱均平和欧玉芳在《美国〈教育项目评价标准〉的制定及启示》一文中评价,“它是一种平衡效用性、可行性、適切性和精确性四大评价质量属性的综合规定,它大量讨论的内外部元评价为提高和保障评价质量提供了方法论”。因与西方存在政治及教育等体制差异,我国教育评估机构的独立程度、教育评估利益相关者对评估的参与程度等与西方有所不同,因此,要将JCSEE元评估标准,尤其是“评估问责”属性所对应的三个标准,结合中国实际探索运用。

(1)建立健全评估文档。我国教育评估大多由教育主管部门委托第三方评估机构实施,多以政府发布文件的形式启动教育评估,文件一般有附件说明评估的目的、内容、形式、时间等。因评估的涉及面较广,所以文件及其附件一般是比较原则性的内容。从建立健全评估文档的角度出发,建议对院校开展评估时,可由专家组会同被评院校,甚至可以邀请教师和学生代表,根据文件要求共同制定评估方案,并记录文档,作为实施评估的基准。要对座谈、访谈、会议、查询等按照规范格式予以书面记录。评估结论可与被评院校、教师和学生代表讨论沟通产生,这些做法与JCSEE对于评估文档工作过程中利益相关各方共同参与并协商的宗旨是一致的。鉴于某类评估可能涉及很多单位,因此可设计专用的文档格式,供评估人员使用。此外,对于相关专家和行政部门领导对评估活动予以现场指导或总结,亦可做好文档记录,归档留存。

(2)开展内部元评估。我国教育评估活动中,评估项目组的内部会议大多是讨论做什么和研究怎样做,很少研究做得怎样和如何改进,一个重要原因是我国的教育评估者大多不是专职人员,一般是根据评估的目的和性质,临时从各单位遴选并抽调出来的专家。一旦评估项目完成,专家即返回原单位,客观上也无法组织探讨评估中出现的问题及

今后的改进方案。因此,开展内部元评估的一个重要保障是组织专职元评估队伍,这在其他一些国家是常规动作。对专业评估者来说,JCSEE还要求研究和改进评估本身就是评估者的重要职责。

(3)开展外部元评估。我国教育评估活动中,外部元评估是相对薄弱的环节,评估项目组除了与被评单位及其内部教育质量保障部门商讨评估安排之外,与被评单位其他人员的交流一般是为了获取评估需要的信息,很少会涉及探讨怎样的评估活动是更有益或更有效的。JCSEE的外部元评估要求在评估活动开展前和实施进程中要与一切利益相关者协商讨论评估应该怎样进行,其中重要的环节是学生参与,在欧美被评单位一定会有学生代表参加,我国的评估也需要学生的参与。

本文探讨了特殊的教育评估——教育元评估,讨论重点是JCSEE元评估标准中的“评估问责”属性和它对应的三项标准的内涵。问责不是责问,“问责过程的核心就是讨论阶段,因为它不仅提供了数据来设定预定的或紧急的问责规范,而且提供了一个问责主客体双方沟通和交流的平台”。<sup>[7]</sup>纵观该属性的三个标准,其蕴含的共性特点之一是要在协商和交流的基础上达成目标,保障教育评估的质量,应对社会对教育的拷问。在我国,教育评估的主体和客体对于教育发展的愿望是一致的,因此协商是有基础的,教育元评估的根本目的是促进在评估过程中利益相关各方加强协商,加强沟通,提高教育评估的质量,提升教育评估的效能,推动教育事业的发展。

### 参考文献

- [1] 刘森林.元语用论概述[J].解放军外国语学院,2001(4):6-10.  
Liu Senlin, A Review of Metapragmatics [J]. Journal of PLA University of Foreign Languages, 2001(4): 6-10.
- [2] 张荣娟,徐魁鸿.美国高等教育元评估制度探析——以高等教育认证委员会为例[J].高教探索,2018(2):65-69.  
Zhang Rongjuan, Xu Kuihong. A study on Meta-evaluation in Higher Education in America - Taking CHEA for Example [J]. Higher Education Exploration, 2018(2): 65-69.

- [3] 李延保. 从首轮本科教学评估的社会效应看评估理论研究的学术文化责任[J]. 高教发展与评估, 2009, 2(5): 38-42.  
Li Yanbao. Academic responsibility of evaluation theoretical study from the perspective of social influence after the first round of undergraduate teaching evaluation [J]. Higher Education Development and Evaluation, 2009, 2(5): 38-42.
- [4] 赵立莹. 问责与改进: 我国第四轮学科评估元评估[J]. 学位与研究生教育, 2018(2): 32-38.  
Zhao Liying. Accountability and Improvement: Meta-evaluation for the Fourth Round of Subject Evaluation in China [J]. Academic Degrees and Graduate Education [J]. 2018(2): 32-38.
- [5] Joint Committee on Standards for Educational Evaluation. The Program Evaluation Standards —A Guide for Evaluators and Evaluation Users [M]. 2nd ed. Thousand Oaks: Sage Publications, 1994.
- [6] Yarbrough D B, Shula L M, Hopson R K, et al. The Program Evaluation Standards: A guide for evaluators and evaluation users [M]. 3rd ed. Thousand Oaks: Corwin Press, 2010.
- [7] 谷志军. 西方问责领域的定量研究及理论发展[J]. 国外社会科学, 2015(4): 41-48.  
Gu Zhijun. Quantitative Analysis and Theoretical Development of Accountability in the West [J]. Social Sciences Abroad, 2015(4): 41-48.
- [8] Program Evaluation Standards [EB/OL]. [2020-02-10]. <https://evaluationstandards.org/program/>.
- [9] 谷志军. 当代西方问责研究: 理论框架、热点主题及其启示[J]. 中国行政管理, 2017(7): 143-147.  
Gu Zhijun. Accountability Research in Contemporary Western Countries: Theoretical Framework and Hot Themes [J]. Chinese Public Administration, 2017(7): 143-147.
- [10] Scriven M. An introduction to Meta Evaluation [J]. Educational Products Report, 1969(2): 36-38.
- [11] 张明广. 自愿问责: 我国公立高校推行问责制的新途径[J]. 山东高等教育, 2018, 6(1): 43-50.  
Zhang Mingguang. Voluntary Accountability: A New Pathway for Public Colleges and Universities in China to Implement the Accountability System [J]. Review of Higher Education, 2018, 6(1): 43-50.
- [12] 冯晖. 教育评估现代化的内涵特征与推进策略[J]. 上海教育评估研究, 2019, 6(3): 1-4.  
Feng Hui. Connotation Characteristics and Promoting Strategy of Modernization of Education Evaluation [J]. Shanghai Journal of Educational Evaluation, 2019, 6(3): 1-4.

(上接第69页)

- mechanism of affecting performance [J]. Advances in Psychological Science, 2003, 11(4): 452-456.
- [2] 陶西平. 教育评价辞典[M]. 北京: 北京师范大学出版社, 1998: 123.  
Tao Xiping. Dictionary of educational evaluation [M]. Beijing: Beijing Normal University Publishing Group, 1998: 123.
- [3] 陈玉琨. 教育评价学[M]. 北京: 人民教育出版社, 2007: 54.  
Chen Yukun. Evaluation of education [M]. Beijing: People's Education Press, 2007: 54.
- [4] 测试结果反馈[EB/OL]. [2019-02-20]. <https://baike.so.com/doc/3731115-3920290.html>.  
Test result feedback [EB/OL]. [20190220]. <https://baike.so.com/doc/3731115-3920290.html>.
- [5] 龙君伟. 反馈干预的内部机制研究[J]. 心理科学, 2005, 28(1): 241-243.  
Long Junwei. Internal mechanism of feedback intervention [J]. Journal of Psychological Science, 2005, 28(1): 241-243.
- [6] 田霖, 沈钢. 反馈理论与自学考试分数报告的改进[J]. 评价与测量, 2011, 3(3): 14-20.  
Tian Lin, Shen Gang. Feedback theory and improvement of self-study test score report [J]. Evaluation and Measurement, 2011, 3(3): 14-20.
- [7] Zhou J. Feedback value, feedback style, task autonomy, and achievement orientation: interaction effects on creativity performance [J]. Journal of Applied Psychology, 1998, 83(2): 261-276.