

技术与伦理:学生评价发展的两条路径*

朱 丽

(上海市教育评估院基础教育评估所,上海 200031)

摘要: 学生评价的发展遵循着两条路径,即技术路径和伦理路径。评价的技术领域在其发展过程中占主流地位,也受到更多关注,而伦理领域则相对处于弱势地位。教育评价的价值属性、教育的道德实践本质以及学生评价中的伦理冲突等决定了有必要考查学生评价的伦理维度。但这两条路径之间并不总是和谐的,甚至有时是冲突的,因此需要加强理论研究,制定相应的伦理规范,将学生评价的目的、手段、方法等置于伦理的框架下来讨论,以此调和两条路径之间的矛盾与冲突,以减少学生评价中的两难困境。

关键词: 技术;伦理;学生评价

中图分类号: G640, G40-058.1 **文献标识码:** A **文章编号:** 2095-3380(2018)04-0011-04

Technology and Ethics: Two Ways of Development of Students' Evaluation

Zhu Li

(Department of Basic Education, Shanghai Educational Evaluation Institute, Shanghai 200031)

Abstract: The development of students' evaluation follows two paths, one is technological, the other is ethical. The technical area of evaluation has dominated the development process, and the field of ethics is in a relatively weak position. Value attributes of educational evaluation, the nature of moral practice in education and the clash of ethics in student evaluation and so on decide that the ethical dimension of student evaluation must be examined. But there are disharmony and even conflicts between them. It needs to strengthen theoretical research, develop corresponding ethical norms, put the purpose, means, and methods of student evaluation into an ethical framework so that the contradiction and conflict can be compromised, the dilemmas in students' evaluation can be lessened.

Keywords: Technology; Ethic; Students' evaluation

美国学者 Susan 认为“评价的两个主要领域,一个称之为技术或科学领域,它是通过运用来自经济、社会科学、统计学等相似学科的定量测量方法来分析活动、目标和对象。但是另一个领域涉及到描述价值、目

标和目的的任务,其与该项目的道德和伦理评价以及其是否已经实现联系更紧”。^[1]可以据此将评价的两个领域分为技术领域和伦理领域。评价的技术领域在其发展过程中占主流地位受到更多关注,而伦理领

* 基金项目: 国家哲学社会科学基金“十二五”规划 2013 年度教育学重点课题“教育现代化评价指标体系及推进路径研究”(AFA130001)子课题“教育现代化进程中的学校评价变革研究”阶段性成果。

收稿日期: 2018-04-02, 修回日期: 2018-05-30

域则相对处于弱势地位。

学生评价作为教学过程中常用的一种管理手段,不可避免地扮演着工具的角色,对学生进行评定,为教学提供反馈,为决策提供信息。学生评价的重要性要求评价具有科学性,能最大程度地反映学生的真实发展水平。追求评价技术的科学性是无可厚非的,但问题是,学生评价首先是为了学生的评价,而不是为了评价的评价。评价技术上的娴熟与精湛,并不意味着其自身的合理性是不证自明的,充其量只是其合理性的前提条件。学生评价的一个重要功能是促进学生发展,这决定了学生评价的价值、道德与伦理是考查学生评价是否合理的另一个维度。在学生评价的发展史上,其遵循着两条路径,即技术路径和伦理路径,两条路径对于完整地认识学生评价是必不可少的,需要调和两条路径之间的矛盾与冲突,以减少学生评价中的两难困境。

一、学生评价技术路径的兴起与发展

古巴和林肯在其著作《第四代评估》中,将评价的发展历程划分为四个阶段:测量、描述、判断和建构。在测量阶段,人们认为任何事物都是有数量的,且均可测量,强调对测试结果进行精确测量。科学性成为其追求的首要目标,是评价获得公信力的重要前提。在评价中,科学性体现为评价的效度和信度。

教育测量的兴起在某种程度上也正是出于对科学性和客观性以及效率的追求。19世纪后半期,美国多采用口试法来考查学生,教师采用不同的问题向学生询问解答,以此判定学生是否掌握知识以及掌握的程度。这种方法的弊端是缺乏统一的评价标准,会掺杂许多评价者个人的感情色彩。同时随着学生数量的增多,口试法在时间上就显得不经济。后来采用笔试来替代口试,采用同样的试题来测试学生,笔试能同时测试大规模的学生,但不足是采用的是主观题(问答题),在评分时易受阅卷者的主观影响,其客观性也难以保证。为了纠正笔试的这种弊端,力求客观化,教育测验运动逐渐兴起。

测量最初用于智力测量。比纳受法国政府的委托编制测量以鉴别能从特殊教学中获益的儿童。经过大量的研究,比纳和他的同事西蒙于1905年发表了一篇文章,在这篇文章中,他们提供了一系列测量儿童心智发展水平的测验。在美国,最早发展起来的是推孟在1916年编制的斯坦福-比纳量表,1937年和

1960年做了修订,它们在心智的测验中起了重要的作用。测量的大量运用直接导致了标准化测验运动的兴起。标准化测验的答案是确定的,便于评卷者评分,这使得标准化测验适于对大规模人群做测试。

至20世纪30年代中期,美国教育测量运动达到鼎盛期。测验的种类更加丰富,不但在技术方面力求正确应用,而且理论方面力求有确证的收获。智力测验方面不断完善,职业指导和升学指导方面出现了性向测验和诊断测验,人格测验开始出现,并力求精密。还成立了专业的测验机构,引进了测验评分机。

20世纪30年代,测验运动进入批判期。人们认为测量或测验并不能测出一个完整的人的全部,如“社会态度、实际技术、创造、兴趣、鉴赏力等都属于教育的重要领域,为教育测验所不能充分把握者,在重视考试与测验的时代,这些重要学力又往往被教育者冷落”。^[2]美国进步主义教育联盟(Progressive Education Association, PEA)认为教育的目的在于生活,在于儿童本位,儿童的个性、趣味、需要等都应受到尊重。测量或测验仅关注学生对知识的学习,更多需要学生进行记忆学习,并不能由此显示学生是否掌握了教师所期望的高级思维能力。“人的主体性都在‘一切皆可测量’的旗帜下不可一世地空前膨胀起来,作为评价主体的人成为了‘近乎上帝的观察者’,评价客体则被当作与自然同质的某种物质加以认识。”^[3]

然而,尽管教育测量遭到了批判,但与口试、面试等相比,测量(如标准化测试)的优势却是有目共睹的。因此,测量仍然得到发展。1955年,美国教育研究协会和全国教育质量标准评议会合作提出了《成就测验的技术标准》。1956年,布卢姆和他的同事出版了著名的《教育目标分类学·第一分册,认知领域》。1964年,《教育目标分类学·第二分册,情感领域》出版。即使在当代,测量依然有其重要地位,作为一种选拔或考核的手段广泛应用于不同领域。

当前,我国进行的考试改革,依然是以测验为基础的改革。随着时代的发展,测量或考试的技术手段也在不断提高,如融入信息技术,许多标准化考试已实现电脑答题、阅卷,建立了题库,越来越科学。此外,人们逐渐意识到事物的发展变化不仅体现在量的变化上,同样也体现在质的发展变化方面,教育效果尤其如此。因此,教育测量越来越与其他评价手段和方法相结合,以更全面地获得评价对象的信息,以便做出客观的评价。例如,近年来,为了更科学地评价学

生,学生评价无论是理念还是实践操作越来越呈现多元化发展,除了标准化测试的新技术不断得到运用外,档案袋评价、真实性评价、表现性评价等评价方式逐步运用到学生评价中,使学生评价的手段越来越丰富,对学生的发展也起到越来越大的促进作用。

二、学生评价的伦理诉求

评价起源于心理测量,其理论基础更多地来源于心理学、测量学、教育学、工艺学等,总的来说都是科学技术性理论,^[2]但“科学能回答是什么和曾经是什么的问题,但回答不了应当是什么和应当做什么的问题”。^[4]应当的问题需要借助伦理学来进行思考,要回答的是“什么样的学生评价是恰当的”。学生评价的另一条路径便属于伦理的,既源于对测量的批判,也源于对教育本质的认识。

1. 教育评价的价值属性决定了学生评价的伦理属性

在测量受到批判的同时,教育评价兴起了。1933-1940年,PEA会长艾金(Aikin,W.M.)领导开展了历史上有名的“八年研究”。“八年研究”的主要目标是实验核心课程(core curriculum)。为达成新课程的目标,符合新教育精神,极需要有一套符合新课程目标的考查法。为此组成了以泰勒(Tyler,R.W.)为中心的评价委员会,他们以全面发展人的才能为主要目标,从事研究教育成绩考查的方法。“八年研究”促成了一种新的评价学生的方式,不再用测量与测验时期的流行观点来进行评价。被称作“教育评价之父”的泰勒在“八年研究”报告(《史密斯-泰勒报告》)中,首次提出了“教育评价”这一概念,他明确提出:“评价过程本质上是一个确定课程与教学计划实际达到教育目标的程度的过程。但教育目标本质上是指人的行为变化,因此,评价是一个确定行为发生实际变化的程度的过程。”该概念不限于教育测量的范畴,通过评价发现学生学了什么以及这些东西的价值,从而与测量相区别。

从词源学上来说,“评价”(evaluation)一词是源于“价值”(value)这个词的,是一种基于事实基础之上的价值判断,正如美国学者格朗兰德将教育评价描述为教育评价=量的记述+价值判断或质的记述+价值判断。从评价的实践来看,评价总是关涉人的,其不可能是价值无涉的,本身就具有价值属性。

2. 教育的道德实践本质决定了学生评价应该合

乎伦理

学生评价是在教育中为学生的发展而进行的评价,教育的伦理道德品性决定了学生评价的职责与使命。教育从其本源与本质上来说,是一种道德实践活动;从其终极目的来说,是人们对善的不断追求,善与教育实践联系紧密,如教师对学生的关爱、学生对知识美德的追求等,可以说教育问题始终伴随着伦理问题。《说文解字》中有“教,上所施下所效也”,“育,养子使作善也。”它指出教育的目的是使人为善。在西方,众多教育家对教育的道德性做了阐述。如英国哲学家彼得斯经过系统分析后发现,“教育”无论作为成就词(achievement word)还是作为任务词(task word)来使用,在日常生活语境或理论陈述中都带有规范性含义,“它意指正在或曾经以某种道德上可以接受的方式有意识地传授某种有价值的东西”。一旦活动或过程被某人称作“教育”,就意味着他在伦理价值上已经认可了它的内容和方式;如果他认为它的内容没有价值甚至有害,或者认为它的方式在道德上不可接受,他就不能称之为“教育”。否则在逻辑上就自相矛盾。^[4]一种活动或影响若被称作“教育”,必须同时满足两个规范标准:其一,它必须传递有价值的内容;其二,它必须采取道德上可以接受的传递方式。^[5]第一条标准鲜见疑义。但对于第二条标准是不是“教育”的必要标准是有争议的。

教育是一种道德实践活动已经无需再过多证明了。教育评价属于教育过程的一个重要环节,教育的道德性决定了教育评价必须合乎道德,遵循一定的伦理规范。从根本上说,学生评价应当居于教育评价的核心,其根本目的在于引导和促进学生发展,更需要伦理规范的调适。但评价的作用兼具积极与消极两重性。积极效应表现在:①促进确认(confirm)的效应。促进师生分别确认指导的成果和学习的成绩。②激发动机的效应。③调整努力和时间分配的效应。④强化成功经验的效应。⑤消退失败经验的效应。⑥调整儿童学习态度的效应。⑦赋予安全感。而消极效应则表现在:①赋予不安全感。②对学生形成自我概念的消极效应。③助长个人主义的学习态度。^[4]例如,将标准化测验置于学生评价的主导位置,以分数作为评价学生的唯一指标,教育本应是育人的,而被异化成育分,分数比成长重要,教育被扭曲了,评价成了“帮凶”。同时,由于标准化测验过度信奉无法被测量的事物是不真实的,导致学生的发展中很多因素,如品行、

兴趣、学习态度等无法得到有效检测,也就成为学生评价中被忽略掉的重要因素,唯智主义的教育导致了学生的片面发展。鉴于学生评价存在可能给学生发展造成消极影响,评价者在对学生评价时必须遵守一定的伦理规则,以避免或降低可能对学生造成的伤害风险。

学生评价作为教育领域的一种日常实践活动,公正合理的评价不仅能促进学生的发展,而且也教会了学生能以客观公正的态度去评价社会。因此,学生评价的道德本性要求不仅关注评价的技术改进,更要关注评价的价值和伦理规范,不仅要问“如何”,更要问“为何”,也决定了学生评价除了求真(科学性)之外,还需对其本体价值进行反思。

3. 学生评价中的伦理冲突需要用伦理规范来调适

从理论上来说,评价研究的两条路径应该并行不悖,二者的结合构成完整的评价认识。然而,现实中却存在科学的未必是合乎伦理的(或者反之)的现象,两条路径时有冲突。在教育评价实践中,“评价工作者在伦理方面所遇到的挑战,要远远大于在评价技术上面面临的困难,他们经常对‘什么是最恰当的评价行为’感到疑惑”。^[6]很多时候,评价者即使掌握了评价技术,但面对真实而复杂的评价情境时会难以做出抉择。

学生评价不是一项纯技术的活动,教师等人在评价学生的过程中可能会遭遇各种冲突和对立,这种冲突有时候很难用简单的“对与错”来评判,而是让评价者陷入一种道德困境或者道德两难。道德困境是指道德主体需要在两种乃至以上的正价值观和行动过程中做出艰难的抉择。很多时候,道德主体不是善与恶之间进行选择,而是在善与善之间做出抉择,即不是简单地在“对与错”“是与非”“善与恶”之间进行选择,而是在相互竞争的对、是与善之间选择,这样的选择无疑是困难的。

比如,学生评价中教师面临的角色冲突,当学生的学业成绩作为学校考核教师的一项重要指标,此时教师是以学生的发展为目标呢,还是以加班加点牺牲学生休息时间为代价来提高学生的分数呢?这些冲突、对立等是无法完全用技术思维来解决的,只能通过伦理规则来调适,在不同的也许都是合理的不同冲突、对立和困境中选择对学生最有利的行为,规避对学生发展不宜且可能有害的行为。这种道德困境在真实的学生评价中是很常见的。从根本上来说,道德困

境迫使教师将学生作为人来深思,从学生的立场来考虑利害得失,并思考自己的行为对学生可能产生的显在的或潜在的影响。

三、学生评价的技术与伦理两条路径的调和

在学生评价研究中,技术路径与伦理路径不是抑此扬彼的,而是学生评价发展的两翼,不可或缺。如果要致力于提升学生评价的品性,必然要调和两条路径。然而,在科学主义范式的主导下,当前学生评价的技术的和测量的路径无论在理论层面还是实践层面乃至国家政策话语体系中处于主流位置,如《国家中长期教育改革和发展规划纲要(2010—2020年)》中明确提出“改革考试评价制度和学校考核办法,建立科学的教育质量评价体系,改进教育教学评价”,科学性依然是教育评价中的主导话语。

相比于技术的测量的路径,伦理研究因其缺乏刚性的、公认的规则与规范以及评价标准,仅是一种道德合约和道德应当,无疑处于“弱势”地位。如何才能调和学生评价的技术与伦理两条路径呢?

1. 加强理论研究,制定相应的伦理规范

从已有研究来看,学生评价伦理理论研究依然薄弱,仍未形成一套公认的伦理规范。美国国家标准研究会(ANSI)于2002年颁布了《学生评价标准》,该标准针对学生评价的四个不同方面提出了四条标准,即适宜性标准(Propriety Standards)、应用性标准(Utility Standards)、可行性标准(Feasibility Standards)、准确性标准(Accuracy Standards),包括28条具体标准。从美国《学生评价标准》的具体内容来看,其涵盖的范围很广,是对学生评价的目的、程序等的规范,其中不少是伦理规范,如学生的隐私保护、学生权利和权益的保护、学生评价中应没有偏见,保证结论是公平的,等等。该标准可以看作是学生评价标准制定中的典范,走在世界的前列,对什么是好的学生评价做出规定性的界定,为教师及其他评价者提供行为指南,能够全面提高学生评价的质量。

当然,学生评价伦理研究的相对“弱势”只是说明了当前它的处境,但并非否定其重要性和必要性,“弱势”的处境更增添了对重视的吁求。

2. 将学生评价的目的、手段、方法等置于伦理的框架下来讨论

技术的和科学的学生评价探讨的是“如何有效”

(下转第19页)

导师制改革要打破终身制现象,推进导师的动态调整,实现“能者上,平者让,劣者下”的导师进退机制。导师制进退机制实施的难点在于不合格导师如何处理,直接取消不合格导师的资格影响较大,牵涉到导师、学生等多方因素,政策实施的难度不小,可以考虑采取选聘分离机制,将导师遴选、聘任以及招生等环节结合起来,在保留导师资格的前提下,通过停止聘任、限制招生等措施,实现导师实际意义上的进入退出。要真正把学术水平高、责任心强又有指导意愿的青年教师选入导师队伍,优化导师的结构层次,提升导师队伍的科研水平,充分发挥青年导师在研究生培养全过程的积极作用。

5. 完善导师的管理机制,调动导师的积极主动性

要改变行政主导的导师管理方式,为导师创造宽松的教书育人氛围。要赋予导师相应的职责权限,从研究生的招生、培养到学位授予,使之真正成为研究生教育的主导者和实施者,切实调动导师的积极性和主动性。要充分发挥高校学术委员会、学位委员会等学术团体的作用,通过整合职能,明晰导师、培养单位、职能部门和学术组织等各方职责,在招生录取、教学培养、学位授予及就业规划等研究生教育过程中,各司其职,确保学位与研究生教育的各项政策得到有效的贯彻执行,不走过场。^[11]要规范学校、学院和导师的三方教育质量分担机制,严格落实教育质量问责机制。

参考文献

- [1] 杜智萍.19世纪以来牛津大学导师制发展研究[D].保定:河北大学,2008.
- [2] 马晓霞,尹逊震.我国研究生导师聘任制的困境及其改革路径[J].中国成人教育,2013(2):41-42.
- [3] 李海生.我国研究生院高校导师队伍现状及思考[J].学位与研究生教育,2015(9):14-19.
- [4] 读研目的调查[EB/OL].中国教育在线,(2014-09-08).<http://www.kaoyan365.cn/dongtai/1617.html>.
- [5] 周文辉,张爱秀,刘俊起.我国高校研究生与导师关系现状调查[J].学位与研究生教育,2010(9):7-14.
- [6] 李春根,陈文美.导师与研究生命运共同体:理念与路径构建[J].学位与研究生教育,2016(4):55-59.
- [7] 程星.世界一流大学的管理之道[M].北京:北京大学出版社,2011:21.
- [8] 杨斌.做合格导师需从心从德从范[N].人民日报,2017-04-20(17).
- [9] 国兆亮,王楠.关于研究生导师评价的几点思考[J].中国高教研究,2012(1):56-58.
- [10] 教育部关于全面落实研究生导师立德树人职责的意见[EB/OL].教育部官网,(2018-1-18).http://www.moe.gov.cn/srcsite/A22/s7065/201802/t20180209_327164.html.
- [11] 李增森.我国研究生教育转型发展及应对策略分析[J].上海教育评估研究,2016(12):11-14.

(上接第14页)

的问题,有效更多是对效率的追求,如对标准化测试中试题科学性的研制就是提高评价的有效性。然而试题的科学性并不能完全回应“如何公正”的问题,对客观存在的学校差异、城乡差异以及个体差异,标准化测试的试题并不能完全兼顾不同利益群体的差异,否则很难说是公正的。在进行学生评价之前,对学生评价的目的、手段、方法等应置于伦理的框架下来讨论“是否公平、公正”及“如何公平、公正”等问题,而不能仅仅关注“如何客观、科学”“如何有效”等问题,应将二者结合起来。

总的来说,对于什么是最恰当的学生评价实践需要从科学性和道德性两个方面来考虑,既需要完善评价的技术,提高评价的信度,也需要从道德的角度进行质询,讨论其必须遵守的伦理规范,为评价者与被

评价者提供道德规约。

参考文献

- [1] Susan Salasin. The Evaluator as an Agent of Change [J]. New Directions for Program Evaluation, 1980(7): 2.
- [2] 瞿葆奎.教育评价[M].北京:人民教育出版社,1989: 65,156.
- [3] 刘宗南.知识论视野下教育评价的科学性与伦理性[J].咸宁学院学报,2008(2):110-113.
- [4] 陈桂生,范国睿,丁静.教育理论的行质与研究取向[M].上海:华东师范大学,2006:351;340;145-146.
- [5] R. Peters, Ethics and Education [M]. London: Allen and Univer-sity, 1967.
- [6] 蔡敏.美国教育评价的伦理规范建设及其启示[J].外国教育研究,2005(6):18-21.