

为了测量还是为了学习：高校公正评价学生学业内涵探析

□ 刘声涛

摘要：教育部核准并向社会发布的多所大学章程对学生学业评价的要求均是公正评价学生学业。多方面证据表明，目前我国高校行政管理人员、教师、学生主要将公正评价学生学业理解为考试并准确评分。从测量技术角度看，准确评价学生学业难以做到。承认测量技术的局限性是一种基本的科学态度，过于依赖和关注考试不仅得不到想象中的准确分数，还对师生有负面影响。学业评价的首要目标是为了学习而不是测量。公正评价学生学业应该被理解为实施为了学习的评价，使学生获得最适当的教育。高校章程制定及高校以章程为依据办学，都需要促进对公正评价学生学业的正确理解与实践。

关键词：高校；学业评价；公正

中图分类号：G641

文献标识码：A

文章编号：1672-0717(2015)01-0059-05

收稿日期：2014-11-30

基金项目：湖南省教育科学“十二五”规划2012年度项目“高校有效学生学业评价研究”(XJK012CGD002)；湖南省社科基金项目“高校发展性学习评估的国际经验与借鉴”(13YBA077)；全国教科规划教育部重点项目“高校学业评估的国际比较研究”(DIA130302)。本文同时受湖南大学青年教师成长计划经费资助。

作者简介：刘声涛(1976-)，女，江西萍乡人，教育学博士，湖南大学讲师，主要从事高校教学与学习评估研究。

《高等学校章程制定暂行办法》实施以来，教育部核准了多所大学的章程，这些章程对学生学业评价工作的表述高度一致，即：学生公正地获得学业上的评价。这个表述中有两个看似平常实则复杂多义的概念：公正、学业评价。博登海墨(Edgar Bodenheimer)说：“公正有一张普洛透斯似的脸，变幻无常，随时可呈不同形状，并且有极不相同的面

貌。当我们仔细查看这张脸并试图揭开隐藏其表面之后的秘密时，我们往往会深感迷惑。”^[1]在不同的时代与人类活动中，人们对公正的涵义有不同的理解，不同的学者对于公正也有不完全相同的阐述，迄今尚未形成一个统一的公正论。教育领域中的评价是一个正在发展中的概念，其内涵在不断地扩大与丰富中。国际上，学生学业评价近几十年正在产生教育评价历史上从未有过的变革^[2]。当公正与学业评价两个复杂的概念在一起时，有必要了解高校行政管理人员、教师、学生如何理解公正评价学生学业，这种理解是否有偏差，如果有偏差那么更合适的理解又是什么。章程是高等学校依法自主办学、实施管理和履行公正职责的基本准则，探讨上述问题有益于高校章程制定及以章程为依据来办学。

一、对高校公正评价学生学业的实际理解

有三方面的证据表明，当前我国高校学业评价相关群体主要将公正评价学生学业理解为通过以考试为主的方式来给予学生应得的分数或等级的活动。第一方面的证据来自于学生调查。2014年1月，笔者在某高校对大学生进行了学业评价书面调查。调查包括三个问题：(1)大学期间，你的学业有没有得到公正的评价？(2)请给出你或同学的学业被不公正评价的例子；(3)你认为被不公正评价的原因是什么？学生针对这三个问题主要围绕教师如何考及如何评分作答。第二方面的证据来自于高校相关政策文本的分析。有研究者对E大学学生评价管理机构的管理职责、课程学习评价规定、学生实习评定和毕业论文评定的相关规定、毕业资格和学士学位授予资格的政策文本进行分析，发现我国现行高校对学生评价的理解与要求是以专业知识为主要评价内容，以考试为主要手段进行考试成绩鉴定^[3]。第三方面的证据来自于研究文献的内容分析。以“高校、公正、评价”或“大学、公正、评价”为关键词在

CNKI数据库中查询到论述高校公正评价学生学业相关问题的3篇论文,这些论文对公正评价学生学业的界定都是给予学生客观公平的学业成绩。可见,人们对于高校公正评价学生学业的理解与实践主要是围绕学业考试和评分进行,评价即考试,公正即评分准确。

注重考试和评分的学业评价活动发挥了评价的测量功能:收集并量化信息。从测量功能角度理解学生学业评价并不奇怪。我国考试文化盛行,在不少人看来,学业评价即考试是自然而然的。首先,考试在我国有长期的历史。1930年教育评价专家泰勒(R.W.Tyler)开始在教育领域使用评价一词,此前评价学习的术语就是考试和测验,而西方的考试和测验发展的源头正是我国的科举制。其次,我国人口众多,经济尚欠发达,社会诚信体系不成熟,考试依然是对社会有重大影响力的人才选拔手段。

测量意义上的公正评价学生学业要求评分准确。简单地说,公正是在一定社会范围内通过对社会资源的公平合理分配使每一个成员得其所应得^[4]。测量意义上的学生学业评价分配什么?最容易被想到的就是学业评价的直接分配——分数或等级。直接的分配类似于代币,用来获得真正的分配,如保研资格、奖学金、工作机会等社会资源或和社会资源密切相关的东西。分数或等级是学生学业成就客观真实状态的数字表征,公正则意味着数字表征准确或地位表征准确。数字表征准确指数字完全和学生学业成就客观真实状态相符。因为公正主要和分配有关,数字对学生学业成就在其团体中所处位置表征准确也能确保学生在分配时得其所应得,这就是地位表征准确。本文中将数字表征准确或地位表征准确统称为准确。

二、测量视角下公正评价学生学业的挑战

研究者若对测量意义上的学业评价的基本技术有更多的了解和思考,会发现准确评价学生学业在测量的各环节上都面临着挑战。

首先是收集哪些方面的信息。学业评价需要收集学生在预期学习成果上的表现信息。高校中的学业评价主要是课程学业成就评价。为了保护高校教师的学术自由,同时也因为高校课程多且不

断变化,课程应该有怎样的学习成果由教师自行决定。在制定高校学术标准和质量标准方面世界领先的英国高等教育质量保障署(Quality Assurance Agency for Higher Education, QAA)也仅是在学科层次上制定基准,不再细化到专业、课程层次。表层上,教师理解课程、设定合适的教学目标的能力以及个人风格等方面的差异导致课程目标多样化。深层上,课程到底该取得何种学习成果及价值判断,原本就是一个难有统一答案的问题。因而,不同教师授课的同一门课程上被评价的很可能是学生不同的学习成果的表现。

其次是用什么评价方法去收集信息。按学生在评价活动中要做出的反应,测量学意义上的学业评价可分为选择——反应评价和建构——反应评价。选择——反应评价包括选择题、判断题、匹配题等需要学生找出答案的评价形式,建构——反应评价包括论文题、表现性测试、档案袋评价等需要学生自己产生答案的评价形式。每种评价方式都有其适用情形及局限。选择——反应评价能考察学生大范围的知识掌握情况,易于计分,但难以考察高层次能力;建构——反应评价能引发学生在高层次能力上的表现,但难以对学生的真实状况做准确推论,难以计分。迄今为止,从测量技术上没有找到能很好地克服每类评价方式局限性的方法。

再次是收集多少信息。由于用于信息收集的时间有限,实际工作中只能在预期学习成果中选择若干并设计有限的评价任务,因而学业评价只能测量学生特定的、有限的行为样本。测量实质上是对学生的真实状态作出推论,推论基于有限的行为样本,同时推论本身会有或多或少的误差,这些都导致评分有误差。

最后是收集到的信息如何量化。选择——反应评价具有易于计分但难以测量高层次能力的特点。高等教育的重要教学目标,如理解分析综合能力、沟通合作能力、自主学习创新能力等都需要借助建构——反应评价方法来评价。教育测量学中,建构——反应评价的评分者一致性受到高度关注。因为高等教育的重要教学目标难以量化,人们希望至少不同评分者对同一行为样本的评分不要相差太远。但这种对评分者间一致性的期望也面临挑战:评分者的标准设定或对已有标准的理解不尽

相同,同时教育测量学已经证实学业评价中各种心理效应导致的评价偏差客观存在且难以克服。基于评分的困难,波帕姆(W. James Popham)等著名教育评价专家都认为应对论文题或档案袋评价的使用持谨慎态度^{[5] (P139)}。

事实上,对测量技术有更多了解和反思的学者对考试有着冷静而客观的认识。波帕姆认为应该降低教师对分数能反映学生实际成就水平的迷信^{[5] (P28)};沃尔弗德(Barbara E. Walvoord)建议放弃拥有完美、简单的等级评分体系的幻想,接受它会有缺陷和限制的现实^{[6] (P10)};刘易斯(Harry R. Lewis)认为分数决不可能成为准确比较不同学科成绩的标尺,即使在一门课程内,分数也是不准确的^[7];韦墨(Maryellen Weimer)指出分数能够精确地衡量与客观评价学习的认识是错误的^[8]。

可以说,准确评价学生学业是一种误解和想象,但当前我国高校却对它有着过度的依赖和关注,并且高校教和学中的很多负面现象都和这种依赖和关注有关。

尽管多数教师不能从测量技术的角度说清楚给学生准确评价学业有多困难,但是他们会感受到这点,尤其是面对建构——反应题时。有的教师会因没有合理评价学生而内心不安,还有教师担心评分不合理导致学生投诉。为避免评分可能带来的麻烦,高校教师经常使用以下策略:(1)抬高分数。给全体同学更高的分数,尽管这对于优秀的学生不公正,但可以避免因评分过低而被学生投诉;(2)对学生按正态分布评分。正态分布是众所周知的能力分布,按正态分布分配不同分数等级的人数比例,这样使得分数看起来很正常;(3)尽量用选择——反应题。选择——反应题好评分,在理工科的测验中普遍使用;

(4)秘而不宣的标准与任务。若公开标准与任务,教师担心学生通过不同的渠道做准备,导致分数看起来不符合学生水平。

一直以来,大学生学习研究不断地证明着学业评价对于学生学习的影响^[9]。学业评价过于依赖与关注考试对学生的负面影响包括:(1)肤浅学习。选择——反应评价主要考查记忆、再认等低层次的认知能力,这只需要学生的肤浅学习即能应对。即便是形式上的建构——反应评价,若没有对题目及评分进行精心设计,也可能只是考查学生低层次能力。(2)突

击学习。考试发生在学习告一段落后,再加上考察低层次能力,学生完全可以通过突击学习通过考试,部分学生甚至还能获得高分。(3)舞弊。既然评价注重的是分数而不是学习,学生就会想方设法获得分数,包括以舞弊的方式。(4)对课程的误解。事实上,学生是根据学业评价而不是教师的课程介绍来定义课程^[10]。一门只要求死记硬背的课程在学生眼中就是枯燥无味、干巴巴的。不合适的学业评价会让学生对真实的成人智力世界产生误解。(5)失去通过学业评价调整自己学习的机会。考试在学期末进行,没有反馈,而反馈被认为是改进学习的核心要素。

三、高校公正评价学生学业：以改进学习为首要目标

面对测量意义上公正评价学生学业的挑战,自然要提高测量技术,以不断提高对学业的测量能力及测量准确程度。更重要的是,要意识到并承认考试固有的局限性。承认这种局限性不仅是正确使用考试的关键,更是一种基本的科学态度,它告诉我们要改变对考试与分数的过分依赖与关注,正确认识公正评价学生学业的内涵。

首先要正确认识学业评价的内涵。在过去的几十年中,学业评价被不断地审视与反思。早在1930年,泰勒开始在教育领域使用评价概念,提出评价比考试和测验有更多的功能和目的,认为教育评价仅用于检测学生的成就是不够的,同时也要关注课程内容、课程形式、学生活动、学习的多种结果。20世纪60年代初,克龙巴赫(Cronbach, L.J.)建议评价者们重新确定评价的概念——不是根据竞争的成绩,而是作为一个搜集和报告对课程研制有指导意义的信息过程,强调评价的诊断和反馈作用^[11]。可见,在评价的最初发展阶段,学者主要就“什么需要被评价”及“评价起什么作用”扩展了评价的含义,认为评价的功能是信息收集及分析反馈,这一功能不仅学业成就检测上需要,与学习相关的多项工作上都需要。在此阶段,教师被认为是评价的组织实施者及评价信息的分析使用者。1968年,布卢姆(Benjamin Bloom)将形成性评价的概念用于学生学习,认为应在学生学习过程中通过检测诊断学习中的问题。形成性评价主要就“什么时候评价”

扩展了评价的含义。20世纪80年代后,社会经济发展不断对人才培养工作提出更高要求,同时认知心理学等学科揭示了学生在学习中的发挥积极主动性的作用,大量学者开始进一步挖掘评价在学习中所能起的作用。学者主要就“评价的核心功能及其实现途径”、“学生在评价中的角色”扩展评价的含义。评价依然是信息收集及分析反馈工具,但这一工具被认为应该是以改进学习为首要目标。不同学者或研究组织用不同的名称来称呼这种以改进学习为首要目标的评价,如学习导向评价、为了学习的评价等,并且总结了这类评价的特征。英国的评价改革研究小组(Assessment Reform Group, ARG)认为,为了学习的评价应该是:有效的教学设计中的一部分;关注学生如何学习;课堂活动的核心内容;教师需要掌握的关键专业技能;建设性的,因为任何评价都会带来学生情绪、情感的影响;考虑到学生学习动机的重要性;促进对学习目标的理解决,在评价标准上应该达成共识;保证学生得到如何取得进步的指导性指导;发展学生自我评价的能力,促进他们进行反思和自我管理;认可学生各个方面的成就^[12]。不同学者或研究组织总结的为了学习的评价的特征大同小异,从这些特征中可以明显看出为了学习的学业评价从目的到形式完全不同于考试。

其次要正确认识学业评价中公正的内涵。为了学习的学业评价分配什么?分配的是能改进学生学习的信息以及根据这些信息所确定的合适的教育。联合国教科文组织的著名报告《学会生存——教育世界的今天和明天》在谈及教育中的机会平等时指出:“给每个人平等的机会,并不是指名义上的平等,即对每一个人一视同仁,如目前许多人所认为的那样。机会平等是要肯定每一个人都能受到适当的教育,而且这种教育的进度和方法是适合个人的特点的。”^[13]应像理解教育中的机会平等一样理解学业评价中的公正,因为它们有着共同的使命——让学生学会生存,获得发展。

澄清学业评价工作的首要目标不仅是评价内涵扩展和丰富的结果,还是使评价活动有效的要求。选拔、甄别、诊断、鉴定、认证、发现、发展、管理等等都是评价可能具备的目标。研究者认为,包括学业评价在内的评价活动之所以很困难就在于评价可以同时具备多种目标^[14],但难以同等良好

地实现所有目标,得到最重视的目标决定评价的内在结构与外在形式。这也就是为什么在学业评价的研究文献中总是可以看到学者指出对某些评价功能和目标“过于重视”、“过于关注”、“过于依赖”、“过分强调”,认为对某些评价功能和目标应该“突显”、“强调”。评价工作需要确定首要目标,并从这个首要目标出发设计评价,以确保首要目标的达成而不是其它。

人们依赖考试一方面是对考试的局限性了解不够,另一方面是强调了学业评价在人才培养中不合适目标——管理。当前我国高校中的学业评价,与其说是一种教育方式,不如说是一种管理方式。在高校管理中,考试是高效且便利的以考促学的工具。高校中的学业考试是高利害的,因为它和毕业及各种荣誉有关,这样的考试自然能促使学生学习。但我们必须追问,有没有导致正确的学习?不管考试在管理上多么高效和便利,它终究是以牺牲学生的利益为代价。正是对于以管理为目标的学业评价的依赖,人们很难看到它和改进学生学习的关系。只有把学生的利益放在第一位,而不是把管理的效率和方便放在第一位,才能正确理解公正评价学生学业的内涵。

为了学习的学业评价在国外高校人才培养工作中正在发挥着重要作用。首先政府部门和相关组织明确将以改进学习为首要目标的评价写进政策文件中。美国高等教育协会(The American Association for Higher Education, AAHE)制定的良好学生学习评价活动的第一条原则是:评价学生的学习必须从教育的价值开始^{[6] (P171)}。QAA明确要求:评估实践应促进有效学习^[15]。其次学者积极进行理论研究,并和教师一起进行学业评价改革。如,美国学者安吉洛和罗斯(Anglo, T. & Cross, P.)等在20世纪80年代组织的课堂研究项目;澳大利亚学者南丁格尔(Nightingale, P.)等在20世纪90年代中期组织的大学教学促进项目(Committee for the Advancement of University Teaching, CAUT)。这些项目均是由评价专家组织及指导教师进行学业评价改革行动研究,在实践中发现问题,解决问题,形成丰富的学业评价资源。如,学业评价方法、学业评价范例供更多教师使用,并致力于在高校形成以

改进学习为目标的评价文化。相比之下，我国对高校学业评价的认识及实践还停留在考试阶段，高校中盛行的依然是考试文化而不是评价文化。

四、结语

在各项教育改革中，评价均被视为最难点或瓶颈。评价的首要目标选择不当是评价不能取得预期效果的重要原因之一。长期以来，我国高校学业评价均承担着测量者的角色，人们认为公正评价即公正评分。在这种情况下，需要从政策层面对人们的认识和实践进行正确引导。翻看我国《教育法》，可以发现教育部已核准的章程中，对学生学业评价工作的表述基本上和《教育法》中的表述一致。章程要与教育法保持一致无可厚非，但保持一致并不意味着原封不动，其遣词造句若能有利于纠正对高校学业评价即考试的片面认识，引导高校进行有利于改进学生评价即考试的评价则更有益于高校提高人才培养质量。即使对学生学业评价工作沿用和教育法一致的表述，也需要高校相关群体正确认识与实践公正评价学生学业，让学生真正从学业评价活动中受益。

参考文献

- [1] [美]博登海墨.法理学——法哲学及其方法[M].北京:华夏出版社,1987:238.
- [2] 崔允漭.促进学习:学业评价的新范式[J].教育科学研究,2010(03):11-15.
- [3] 沈晓丽.我国普通高校学生评价实践研究——以E大学为个案[D].华东师范大学硕士学位论文,2008:摘要.

- [4] 洋龙.平等与公平、正义、公正之比较[J].文史哲,2004(04):145-151.
- [5] [美]W.James Popham.促进教学的课堂评价[M].北京:中国轻工业出版社,2003.
- [6] [美]Barbara E.Walvoord & Virginia Johnson Anderson.等级评分——学习和评价的有效工具[M].北京:中国轻工业出版社,2004.
- [7] [美]哈瑞.刘易斯.失去灵魂的卓越[M].侯定凯,等译.上海:华东师范大学出版社,2012:117.
- [8] [美]玛丽埃伦.韦默.以学习者为中心的教学——给教学实践带来的五项关键变化[M].洪岗译.杭州:浙江大学出版社,2006:82.
- [9] Nightingale,P.,Te Wiata,I.,Toohey,S.,Ryan,G.,Hughes,C.,and Magin,D.Assessing Learning in Universities.Sydney:University of New South Wales Press,1996.117.
- [10] David Carless, Gordon Joughin, Ngar-Fun Liu. How Assessment Supports Learning: Learning-oriented Assessment in Action[M].Hong Kong University Press.2006:2.
- [11] 王萍,高凌飏.“教育评价”概念变化溯源[J].华南师范大学学报(社会科学版),2009(04):39-43.
- [12] Assessment Reform Group.Assessment for Learning:10 Principles[R].Cambridge:University of Cambridge,2002.
- [13] 联合国教科文组织国际教育发展委员会.学会生存:教育世界的今天和明天[M].北京:教育科学出版社,1996:105.
- [14] Carless,D.Learning-oriented assessment:conceptual bases and practical implications[J].Innovations in Education and Teaching International,2007(01):57-66.
- [15] QAA.Code of practice for the assurance of academic quality and standards in higher education,Section 6:Assessment of students [EB/OL]. <http://www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Code-of-practice-Section-6.aspx>.2014-3-31.

For Measurement or for Learning: Research on the Meaning of Learning Evaluation in Higher Education

LIU Sheng-tao

Abstract: The approved constitutions made by six universities in 2013 require that learning evaluation should be just. There are several evidences which show that learning evaluation is understood as examination and accurate score by the administrative personnel, teachers and students in universities recently. From the perspective of measurement technology, it's hard to evaluate students' learning accurately. Relying too much on and paying too much attention to examination will bring negative influences to teaching and learning. The primary goal of learning evaluation is not measurement, but promoting learning. Impartial learning evaluation should be understood as the implementation of assessment for learning, and it's purpose is to enable students to obtain the most appropriate education. When we make and implement a constitution, we should promote accurate understanding and practice of just learning evaluation.

Key words: higher education; learning evaluation; justice

(责任编辑 黄建新)