

多元·科学·规范：美国高校学生评教制度管窥

□ 徐 丹 熊艳青

摘要：美国高校学生评教已有90余年历史，评教目的由单一走向多元，评教工具形式多样、内容渐成共识，评价体系实施有序，评教数据处理规范、结果反馈合理等特点已成为学生评教工作有效实施的基石。美国高校学生评教经验启示我们：提高参与者对学生评教活动的认识；鼓励评估专家建立评教指标数据库，设计科学的学生评教量表；恰当处理造成评教偏差的潜在因素，提升评教结果的可靠性与准确度；构建高效反馈系统，建立“以评促教”的效能跟踪机制。

关键词：美国高校；学生评教；评教制度；评教流程；教学有效性

中图分类号：G640

文献标识码：A

文章编号：1672-0717(2016)03-0102-06

收稿日期：2016-04-10

基金项目：湖南省教育科学规划2012年基金项目“我国研究型大学本科就读经验现状调查”(XJK012CGD001)；湖南省社科基金2012年项目“中美研究型大学本科就读经验比较”(12YBA063)。

作者简介：徐丹(1977-)，女，湖南汉寿人，教育学博士，湖南大学教育科学研究院副研究员，主要从事高等教育质量评价研究；熊艳青，湖南大学教育科学研究院硕士研究生。

高校学生评教是指高校在一定评教理论的指导下，采用一系列的工具与方法，组织学生对任课教师的教学工作如教学能力、教学组织安排、教学效果等方面进行客观评价，然后由评价专家对数据进行解释分析，最终形成评教报告的过程。学生评教作为教师教学评价和大学教学质量保障体系的重要一环，担负着改进教学、维护和实现学生利益、保证和提升人才培养质量的重要职责。我国高

校学生评教从20世纪80年代陆续开展起来，经过几十年的发展，已取得一些显著进展。但是我国高校学生评教制度与发达国家相比，仍然面临着评教功能与目的不明确、工具与结果的使用不当等问题。为解决目前国内学生评教过程中面临的诸多问题，学习与借鉴美国高校学生评教制度的先进经验十分必要。

一、评教目的由单一走向多元

美国高校学生评教工作最早可以追溯到20世纪20年代，当时的学生评教工作只是零星地在高校开展。直到20世纪60年代至70年代，学生运动在全球迅速展开，顺应时代潮流，美国大学生也参与到了学生权利运动中来。具体到校园中，由于对教学质量、课程设置、考试方法等强烈不满，美国大学生要求高校给予参与学校管理的权利。为了回应学生的要求，各高校纷纷开始收集学生评估教学的信息，目的是帮助教师改进课程教学、提高教学质量，这也是学生评教的初始目的。20世纪80年代以后，由于美国联邦政府对高等教育的资助不断下降以及“学生消费者第一”思潮兴起，管理者不得不依靠更多的评价信息来进行课程设置和艰难的人事决策，管理者成为了这一时期推动学生评教工作的主要力量，学生评教目的相应地演化成为上述事务服务。进入到90年代，高等教育质量滑坡，为了应对更广泛的社会(包括政府、管理者、学生、家长、纳税人等)对高校问责的公共需求，学生评教目的走向多元化，主要包括：帮助教师改进教学，提高教学质量；帮助学生在选择课程及教师时做出合理的选择；帮助管理者在教师选聘、教师薪水、晋升与任期、课程的设置与发展等方面做出更加科学合理的决策。

二、评教工具形式多样，内容渐成共识

1. 多元混合的量表形式

美国高校的学生评教工具通常采用专家设计的量表。由于每个专家的设计理念存在差异，因此，评教工具并未达成共识。20世纪70年代专家设计推荐的量表有四种，分别是整体评教量表(Overall Instructor Rating)、总体或高推断评定量表(Global or High Inference Rating Scales)、低推断量表(Low Inference Scales)，以及围绕学生对课程学习中能力增长的感受的测量而展开的量表。

整体评教量表通常采用李克特五点计量法，邀请学生对授课教师的全面或者一般性教学技能进行评价。此量表的优点是有助于开展总结性评价，进行人事决策，不足是无法反馈给教师详细的教学信息。典型的代表是佛罗里达大学教育学院使用的评教量表。总体或高推断评定量表同样通常采用五点计量法，通过测量学生对教师和课程的感受间接评价教师教学的有效性。此类量表的特点是有助于教师获取学生详细的反馈信息，改进教学。典型代表有普渡大学、密歇根州立大学使用的量表。与高推断评定量表相对应的另外一种量表是低推断评定量表，它直接测量教师行为，而不是通过学生感受推断教师表现。这类量表的功能与高推断量表类似，即有利于获取详细的教学行为，而且相对于高推断量表而言，评价结果更容易解释。第四类量表测量学生对课程学习中能力增长的感受，间接测量教师教学有效性^[1]。

随着专家学者对评教工具的探究走向深入，目前美国高校的学生评教工具绝大多数都具备了以上两种或者多种类型量表的特征。比如Buskist与他的同事2002年设计的教师行为量表(Teacher Behavior Checklist)集聚了低推断量表与整体评教学量的特征^[2]。美国堪萨斯州立大学设计与发展的“教学发展与效果评价”量表(Instruction Development and Effectiveness Assessment, IDEA)是总体评教量表、高推断量表、第四类量表结合的典型代表^[3]。这类综合型量表往往是由大学或者是公司研发设计，并且形成了一个评教系统。除了刚提到的IDEA系统以外，大型的评教系统还

包括普渡大学的PICES系统(Purdue Instructor Course Evaluation Service System)、华盛顿大学的IAS系统(Instructional Assessment System)、密歇根州立大学的ICES系统(Instructor and Course Evaluation System)等。这些评教系统具有一些共同特点：通常会有一个大型的评教指标库，指标库中包括学校核心指标及与课程(内容、目的、材料等)、教师(表达能力、组织能力等)、学生(参与度、学习成就等)相关的指标。相关人员可以根据特定的评教目的按照一定原则在指标库中选取适宜的指标组合成量表，不同大学采用的原则有别。以华盛顿大学为例，该系统按照教学的组织形式将量表细分为15类，这15类量表的结构相似，均由三部分共31个评教指标组成：第一部分是4个整体评价指标；第二部分是若干个解释性指标(比如学生期望获得的成绩)；第三部分是一系列诊断性指标(关于课程的具体信息)；最后一部分的诊断性指标可以由教师自由选择^[4]。

2. 日渐达成共识的评教内容

尽管评教工具风格各异、形式多样，但是，学生评教的本质是一致的，即对教师教学的有效性进行评价。那么，什么能够体现教学的有效性？事实上，美国学者对教学有效性特征的理解是存在争议的。目前能够达成统一观点的是，从教师行为、学生成就、课程要素三个侧面去反映教师教学有效性，每个侧面又表现出各种具体特征^[5](见表1)。在实际的评教量表中，有效教学被视为是以上三种特征的混合体，是多维的。在此基础上，各量表的设计者依据各自的理论模型挑选有效教学特征，并通过因子分析等数据统计方法确定维度，形成多维度评定量表。比如Frey(1978)在西北大学选用教学技能因素(如课程计划、表述清晰能力等)与师生关系因素(比如教师热情)两维度来测量教学有效性^[6]。Marsh(1984)在南加州大学则提出了9大维度^[7]。目前，Centra^[8]和Braskamp and Ory^[9]的研究发现学生评教量表的六个核心维度是：(1)课程组织与设计；(2)清晰明确的沟通技能；(3)师生互动；(4)课程负担与难度；(5)测试与评分；(6)学生认知学习。

表 1 有效教学特征分类^[5]

教师行为	学生成就	课程要素
学科知识	知识与技能	课程的适应性
热情与活力	兴趣和好奇心	书面作业
沟通技能	自我概念	阅读作业
难度、负担	社交技能和态度	教材
节奏	职业技能与态度	考试
组织、明晰		评分
反馈		媒介使用
师生互动		助教
师生关系密切		实验室作业
课堂组织灵活		复习专题

在评教量表中,各维度下面又会有3~4个体现维度特征的子项目。这些子项目的描述具体、易懂、与学生密切相关、在评价时易于把控;每一个项目只测量一种特征,而且测量对象是已经发生或正在发生的现象或行为。例如:“本课程增加了我对该领域的学习兴趣”、“教师鼓励我提问和评论”等。

三、评价活动组织实施有序

美国大学的学生评教工作不是由学校行政管理部门而是由类似教学指导中心的机构实施。如加利福尼亚州圣马科斯大学由院校规划与分析办公室(The Office of Institutional Planning & Analysis)、普渡大学由教学卓越中心(Center For Instructional Excellence)组织实施。

一般来说,这类机构首先就学生评教量表征求教师的意见,针对教师所提的合理建议,完善量表;量表最终确定后,选取一部分学生进行预评价以检验量表可行性;最后,预评价阶段顺利通过以后,根据评教场地的不同分别发放纸质版本与在线版本量表给学生。

在具体的实施过程中,根据学生评教实施时间段的不同,可以将学生评教方式分为三种:一是随堂评价,如教师经常通过课堂评价的方式征询学生的意见和建议,一般不需要正规量表。二是期中评价,既可以是教师向教学指导中心索要问卷进行调查,也可以由教师自发组织比较开放的讨论和问答,还可以是对某一专题或某一方面的反馈。这种评价也称为形成性反馈,该评价目的是促进教师改进教学,帮助教师提高期末评价成绩。Centra曾对400名教师的学生评教成绩进行比较,研究发现:学生在期中时

的评教的反馈意见被教师采纳后,在后续的评价中,该教师的成绩会有一个显著的提升^[10]。三是期末评价,期末教学评价又称为总结性教学评价,一般在课程结束的前两周进行,由教学指导中心或相应的机构发放量表。此阶段评价强调全面性,主要表现在:全体学生都参与评价,所有课程都需进行评价,所有上课的教师都要接受学生的评价。此外,为了消除学生的后顾之忧,评教均采用匿名形式,不允许教师在场,而且教师必须在学生的课程成绩出来以后才能获得评教结果。这些举措在一定程度上鼓励了学生畅所欲言,保证了评教的真实性。

具体的操作流程各校会有所差异,以加利福尼亚州立圣马科斯大学为例:首先,院校规划与分析办公室(IP&A)会根据课程的类型或者特定的评教目的选择评教量表,将之放置在一个特定的信封中,该信封上面有标注教师的姓名、课程编码、名称,以及开课的学期年份;然后,该办公室委派一名工作人员去往教室,在班级随机选取2名同学作为助手,指导评教流程(带领学生读导语、结束后收发问卷等)。所有同学完成答题后,由前面指定的两名学生将答卷统一收好,装在信封中,交给IP&A中心的工作人员,送至专门的统计评价部门。如果实行的是网上评价,那么评教系统会自动发送邮件给学生,学生根据提示进入评教界面,按程序和要求填充评价量表即可^[11]。

四、评教数据处理规范,结果反馈合理

学生评教的数据处理与结果使用是整个学生评教工作中的重要环节。在美国高校中,学生评教的结果必须公开,但不是简单地公布评价排序或是打分,而是通过学校评价统计中心的运算和适当的处理后,按一定模式或要求进行计算和比较,得出综合评价结果。如对于选择题的处理,选择题的统计结果一般会包括分布频次、平均数、标准差、班级规模与回复率四类信息。美国亚利桑那大学在对2001年学生评教数据进行分析时发现:如果某题项数据的标准差大于1.2,学生评教分数可能在高低分数间离散分布,或者均匀分布在不同的反应区间^[12];一定的班级规模会要求相应的回收率,低于该回收率,则该调查结果无效。此外,专家学者通过实证研究

发现,影响学生评教的因素有很多,比如:被评教师的职称、表达能力、学生的学习动机、期望获得的成绩、课程的水平、难度等等。这些因素中,有些是教师可控的,有些则是教师不可控的(比如学生的学习动机、期望获得的成绩)。因此,为了缓解这些因素的影响,统计部门在对数据进行分析时,会适当控制一些变量,以保证结果的可比性。

学生评教的结果一般反馈给三类群体,即教师、学生和管理者,反馈给每类群体的评教结果侧重点不一。每所高校在这个问题上的规定不一样。以华盛顿大学为例,统计部门处理数据后,会呈递三个报告,分别是课程总结报告(Course Summary Reports)、定制的汇总报告(Custom Summary Reports)、高低分报告(High-low Reports)。课程总结报告主要提供教师在各个课程环节的优势与弱势的信息,用于帮助教师改进教学,反馈对象是教师;教师委员会一般会收到定制的汇总报告,该报告对所有的课程评分进行了标准化处理,教师委员会可以参考各课程的评分来识别课程的受欢迎程度,调整院校的课程设置以满足学生的需求;高低分报告主要列出了每学年在总结性评教中的最高分与最低分教师,该报告主要用于识别优秀教师以及需要提供额外支持的教师。高低分报告往往与其它两类报告一起反馈给管理部门,作为进行人事决策的重要依据。学生也可以通过开通的网址浏览部分评教结果(如定制的汇总报告)用于课程与任课教师的选择^[4]。

当评价结果反馈给教师后,为保障教师有便利的解决问题的渠道,学校教学服务指导中心或类似

机构会针对教师关心的问题或者在此次评教中表现不佳的教师给予支持与帮助。当然,如果教师对评教结果存在异议,可以通过高校教师申述机制进行申述,具体流程如上图所示:

五、启示与借鉴

1. 提高参与者对学生评教活动的认识

虽然我国高校开展学生评教已有几十年的历史,但是在实践过程中,仍存在着诸多不足之处。首先,学生与教师的评教态度都有待端正。有学者对湖南某三所高校的学生评教活动调查发现:36.7%的学生表示在评教中没有仔细阅读评教指标,34.6%的学生表示在评教中会参考其他同学的答案;36.9%的教师和42.1%的学生表示教师会因为学生评分高低而“迎合”学生^[14]。其次,管理者对学生评教目的与功能定位不明确,往往无法在人事决策与改进教学之间寻找到一种平衡。这一系列的问题与评价的主客体、实施者对学生评教工作的认识度不够有密切的关联。

因此,需要采用多种渠道宣传,让参与评教的主客体及实施者明白,学生评教工作不仅需要学生的密切配合,教师的理解与支持,也需要管理者的合理引导。具体而言:针对学生,可以通过讲座、小型座谈等方式让学生明确评教活动的目的、作用及其实施过程,唤起学生的权利与责任意识;针对教师,通过开展多种形式的宣传,如介绍国内外成功的学生评教案例,用事实证明学生具有评教的能力、评教结果的合理使用能提高教学等,引导教师调整态度,提高认识;针对管理者,可以通过与评估专家面对面的交流、培训等方式,提高他们对评教工作的认识,使其意识到期末终结性评价的作用是有限的,应以发展的眼光,激发教师自我发展和改进教学的内在动力,引导督促教师自身对课堂进行评价,获取有助于改进教学的学生反馈意见,最大限度地发挥学生评教的作用。

2. 建立评教指标数据库,设计科学的评教量表

学生评教量表的科学性是保证评教工作是否有效的重要一环。美国高校评教量表的设计有两点基本原则值得借鉴:

首先,标准化与个性化相结合,设计弹性化量

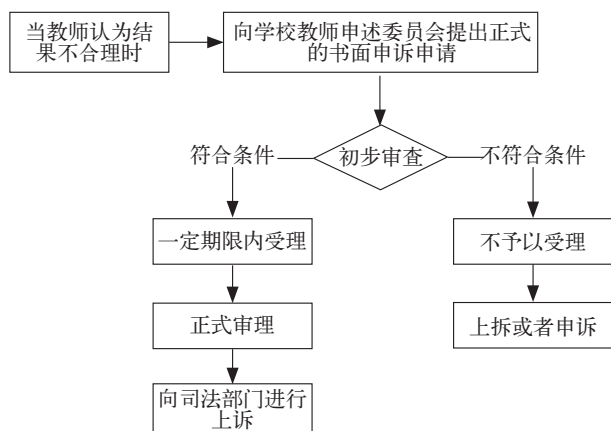


图1 美国高校教师内部申诉流程^[13]

表。目前,我国高校学生评教量表指标体系过于统一,不能满足不同性质、学科课程的评定需求。即使有个别院校意识到了该现状,并尝试根据不同的课程性质设计不同的评教量表,但又出现了各量表之间缺乏共性的问题。如何兼顾评价标准的普遍性与针对性,美国部分高校的做法有一定借鉴价值:先根据有效教学的特征设计评价指标,形成一个庞大的评教指标数据库;然后,按照评教量表结构选择院校指标、个性化指标,形成评教量表。这里的院级指标是标准指标,每个量表共有,具有标准化、可比性特点,它一般是对教师或者是课程的整体评价;个性化指标是各教师根据课程实际情况(教授方式、课程特点等)在评教指标数据库中自由选择的指标,当然也可以是教师自行设计的指标,它们具有个性化的特点^[15]。二者结合于量表中,即保证了共性也兼顾了个性。

其次,设置有关学生学习体验和教师行为表现等学生可以形成直观判断且易于评价的指标。综合国内外研究成果,学生一般可对下述方面做出较好的判断:第一,教师的教学技能;第二,教师的职业道德与伦理规范;第三,学生参与;第四,师生关系;第五,课堂的组织与管理;第六,教学的明晰性、系统性;第七,考试公正性;第八,学生学习收获^[16]。

3.考虑造成评教偏差的潜在因素,提升评教结果的可靠性与准确度

即便在学生评教前进行了充分的准备、也设计了科学合理的评教量表,在实际的评教过程中,仍存在着诸多影响学生评教结果的因素,如学生因素(如选课理由)、教师因素(如教龄)、课程的因素(如课程难度)等。恰当处理这些潜在的因素,才能生成公正科学的评价结果报告。对学校而言,首先需要加大对影响学生评教因素的实证研究,了解各类因素对学生评教结果的实际影响程度。比如:Marsh(1980)曾检验了16项“潜在偏差”(学生、课程、教师背景特征)和SEEQ量表之间的多重关系,发现了学科兴趣最能影响评价的好坏这一重要结论^[17];其次,了解掌控了这些因素后,对原始数据进行分析与处理数据时,采用较高级的数据处理分析方法,适当控制一些变量来解释数据。

4.构建高效反馈系统,建立“以评促教”的效能跟踪机制

为了方便教师、学生及管理者获取学生评教的结果,构建一个具备如下特征的高效反馈系统十分必要:首先,反馈面广,凡与学生评教结果有关的个体、团体、部门都能依据需要获得反馈信息;其次,反馈信息充分,即学生评教反馈的信息应尽量满足不同群体的需求;第三,反馈及时;第四,反馈方式适宜,即学生评教信息反馈的方式能被大家普遍接受,尽量减少学生评教的负面效应^[18]。

为了提高学生评教的实效性,除了建立高效的反馈机制外,还需要建立一套“以评促教”的效能跟踪机制,使优秀者得到褒奖,有问题者得到帮助。例如建立教师评价档案,记录并跟踪教师的成长;建立教师奖励机制,对表现突出的教师给予精神和物质的双重奖励;成立教师教学指导中心,对评教中表现不佳的教师给予帮助,等等。

参考文献

- [1] Greenwood, G. E. et al. (1976). A study of the validity of four types of student ratings of college teaching: Assessed on a criterion of student achievement gains. *Research in higher education* 5: 171-178.
- [2] Buskist, W., Sikorski, J., Buckley, T., & Saville, B. K. (2002). *Elements of master teaching. The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer.* Mahwah, NJ: Erlbaum: 27-33.
- [3] Instruction Development and Effectiveness Assessment Center [EB/OL] <http://ideaedu.org/>
- [4] Office of Educational Assessment in Washington University [EB/OL] http://www.washington.edu/oea/services/course_eval/about.html.
- [5] Richard A. Schwier. (1982). Design and use of student evaluation instruments in instructional development. *Journal of instructional development*. 5, 28-34.
- [6] Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in higher education* 9: 69-91.
- [7] Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*. 76, 707-754.
- [8] Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness.* San Francisco: Jossey-Bass.
- [9] Braskamp, L. A., & Ory, J. C. (1994). *Assessing*

- faculty work; Enhancing individual and institutional performance. San Francisco: Jossey-Bass.
- [10] Centra, J. A. (1979). self-rating of college teachers: a comparison with student ratings. *Journal of education measurement* 10(4): 287-295.
- [11] Course Evaluations in California State University San Marcos [EB/OL] <http://www.csusm.edu/ipa/course-evals/ce-forms.html>.
- [12] University of Arizona, Assessment and Enrollment Research (2001). Guide to student ratings at university of Arizona.
- [13] 欧阳鲁鹏. 我国高校教师校内申诉制度研究[D]. 云南大学, 2015: 26.
- [14] 何云辉, 秦国柱. 高校学生评教结果影响因素研究[J]. *理工高教研究*, 2008(01): 67.
- [15] Center For Instructional Excellence in Purdue University [EB/OL] <http://www.purdue.edu/cie/Website%20CoursEval/courseeval/>.
- [16] 雷敏. 论提高高校学生评教质量的方法和策略[J]. *高教探索*, 2005(01): 52.
- [17] Marsh, H. W. (1980). Students' Evaluation of Instruction: A longitudinal study of their stability. *Journal of Educational Psychology*.
- [18] 陈晓玲. 普通高校学生评教质量控制研究[D]. 华中师范大学, 2003: 32.

Pluralism, Scientificity, and Normalization: An Insight into the System of Student Evaluation of Teaching in American Colleges and Universities

XU Dan XIONG Yan-qing

Abstract: The SET (Student Evaluation of Teaching) system in American colleges and universities has been carried out for more than 90 years. As basis of an effective system, it is characterized by plural purposes, multiple tools with different focus, controlled data analysis, and relevant feedback. For Chinese colleges and universities, it is necessary to improve the understanding of SET by stakeholders, encourage experts in evaluation to establish database of indicators, design a scientific and relevant scales for different purposes, and deal with the potential factors out of faculty control which might effect results, so as to improve the reliability and accuracy, and last but not least, construct feedback system to trace the effectiveness of faculty development in teaching.

Key words: American colleges and universities; student evaluation of teaching; teaching evaluation system; procedure of teaching evaluation; teaching effectiveness

(责任编辑 黄建新)