

# 美国 IQA 教学质量评估系统的 探索与反思

周九诗 鲍建生

**摘要** 有效的教学评估系统，不仅能够准确评判教学水平，还具有改善性和导向性功能，达到提高教师教学能力的发展性目标。美国 IQA 教学质量评估系统由匹兹堡大学主持研发，包含任务认知需求、解释性话语和教师期望三个评估维度，对教师现场课堂教学和教师提供的教学相关材料进行评估。不同于教学视频录像可供研究者多次反复观看，现场听课具有较强的信息及时性，需要评估者快速准确地捕捉评估要点，因此 IQA 具有较少的评估维度和子维度，且每个子维度的水平区分较为清晰，便于评估者快速、及时地作出恰当判断，提高该工具的信效度。IQA 框架的主要特点体现在以下四个方面：评估要点聚焦化，凸显教学质量本质属性；专业化分科化的评估工具，保证评价效果科学权威性；多元化的评估方式，力求从不同角度真实反映教学质量；评估目标功能化，以深化教师专业发展为评估要旨。IQA 评价系统对我国课堂教学评价研究和教师专业发展都具有一定的启示借鉴作用，包括简化评估指标，注重教学实质、分科化评估工具，保障评估结果专业性、评估方式多元化，力求多角度真实反映教学质量、完善评估理念，加强评估发展性功能。

**关键词** 教学质量 教师专业发展 评估系统 课堂教学

**作者简介** 周九诗，天津师范大学教育学部讲师，教育学博士（天津 300387）；鲍建生，华东师范大学数学科学学院教授，博士生导师（上海 200241）。

中图分类号 G4 文献标识码 A 文章编号 1009-5896(2021)05-0087-13

近年来，随着基础教育改革不断发展，教育评价研究领域成果不断完善，教育过程质量的重要性得到普遍关注。教学评价理念已由单一诊断功能，逐步向改善性和导向功能转变。如何科学系统评价教师教学质量，利用评价过程和结果促进教师专业发展、提升教师教学质量是新一代教学质量评估系统的根本诉求。2019年6月，中共中央办公厅、国务院办公厅印发了《关于深化教育教学改革全面提高义务教育质量的意见》，明确提出健全教育教学质量评价监测体系的要求。

国际上对课堂教学质量的评估始于 Hiller、Fisher 和 Kaess 在 1969 年所做的研究,之后课堂教学评价体系一直处于不断探索和发展的过程。自 20 世纪 80 年代,我国教育测量学和统计学开始逐渐恢复和发展,中小学开始出现各种课堂教学评价量表,以期对课堂教学质量、教师教学水平作出公正、客观、科学的评估。<sup>①</sup>但长期以来,我国尚未形成一套完善的教学质量评价标准,依然存在评估工具操作性不强、指标结构层次混乱,无法深刻准确反映课堂教学活动本质等问题。

我国现有的教学质量评估系统多数不具有学科特色,如陈佑清和陶涛从教学目标、教学内容、教学方法、教学手段和教学效果五个维度评价教师课堂教学的评价指标体系,其中每个维度又分为 2 至 5 个要素。<sup>②</sup>孙亚玲从教学目标、教学活动、教学能力、教学反馈和教学组织与管理五个方面评价课堂教学质量。<sup>③</sup>周建华认为高效的课堂评价指标有六个维度,分别为教师素质、教学过程、教学目标、学习环境、教学方法和学生发展。<sup>④</sup>类似的评估框架看似覆盖了课堂教学的方方面面,却没有抓住有意义教学的本质特点,导致评价工具看似学科普遍性很高,适用范围很广,实则难以把握学科特有的育人价值和核心能力培养,削减了评估结果的专业化价值。西方教学评价框架更关注课堂中的师生互动,侧重于教师的指导性和学生的主体性,如 Danielson (2013) 团队开发的教学评价框架 (A Framework for Teaching, 简称 FFT) 中课堂教学评价部分包括教学环境和教学指导两部分,教学环境评估教师能否布置适于学生学习的物理环境、创建积极融洽的学习文化、良好掌控课堂节奏和合理应对学生表现,教学指导主要评估教师的提问和组织讨论技巧,形成性测试的科学运用和师生交流互动的掌握情况。<sup>⑤</sup>

美国在课堂教学质量评价方面一直走在国际研究前列,美国匹兹堡大学开发研制的教学质量评估系统 (Instructional Quality Assessment, 简称 IQA) 适用多个州和地区,涵盖多个学科领域,针对不同学科都有与之相对应的评价内容和评估标准。IQA 评估系统具有良好信效度、较强可操作性,研究表明 IQA 评估结果与学生学业成就呈正相关。<sup>⑥</sup>美国中等数学教学机构项目 (The Middle School Mathematics and the Institutional Setting of Teaching, 简称 MIST) 采用 IQA 为研究工具,调查 4 座城市共计 114 节数学课,发现课堂教学质量与教师教龄、课程标准使用时间、教师专业发展支持均呈正相关。<sup>⑦</sup>不少研究者以 IQA 作为教师专业发展监控框架,通过 IQA 指标判断教师专业技能变化趋势,以此

① 安桂清,李树培.课堂教学评价:描述取向[J].教育发展研究,2011(02):48-52.

② 陈佑清,陶涛.“以学评教”的课堂教学评价指标设计[J].课程·教材·教法,2016,36(01):45-52.

③ 孙亚玲.课堂教学有效性标准研究[M].北京:教育科学出版社,2008:150.

④ 周建华.中学高效课堂评价标准实证研究——以数学教师和数学课堂为例[J].课程·教材·教法,2013,33(08):47-52.

⑤ Danielson Group. The Framework for teaching[EB/OL].[2020-10-12]. <https://danielsongroup.org/framework/>.

⑥ Boston, M. D., Colby, G. Identifying meaningful aspects of mathematics instruction[R]. Paper presented at the annual meeting of the American Education Research Association, New Orleans, 2011: 18-27.

⑦ Melissa D. Boston, Anne Garrison Wilhelm. Middle School Mathematics Instruction in Instructionally Focused Urban Districts[J].Urban Education, 2017(52): 829-861.

评估教师专业培训效果。<sup>①②③</sup>坎德尔（Candela）和波士顿（Boston）等人的研究均表明，IQA 框架本身可用于指导教师专业发展，教师在理解 IQA 评价理念、评价内容和评价标准后，通过相应的实践培训，教学能力得到显著提升。<sup>④⑤</sup> 本文详细介绍 IQA 的研发基础、系统结构和特点，探讨教学质量评估系统对我国课堂教学研究和教师专业发展的借鉴意义。

## 一、IQA 教学质量评估系统的研发背景与基础

20 世纪 80 年代初，美国国家卓越教育委员会（National Committee on Excellence in Education）颁布《国家危机》（*A Nation at Risk*）法案，强调制定学生学业成就和教师教学质量评价标准的迫切性和重要性，正式拉开美国基于标准（Standards-based）的教育改革运动序幕。随后，美国联邦政府出台的《美国教育：2000 年目标》（*Goals 2000: Educate America*）、《不让一个孩子掉队》（*No Child Left Behind*）法案中均明确指出，教师教学质量是学生能力发展和知识获得的关键因素，需要进一步加强和完善教师专业发展和教学质量评价工作。奥巴马政府十分关注教师教学效能，制定了《力争上游》（*Race To The Top*，简称 RTTP）、《教师激励基金》（*Teacher Incentive Fund*，简称 TIF）等一系列以教师教学评价为主体的政策法案，鼓励高校和科研机构研制开发教学质量评估系统，并对各州教师评价和教师专业发展项目提供大量财政拨款。美国国家研究协会（National Research Council，简称 NRC）认为，不能仅通过某一段时间内学生学业成就上的变化考察教师的教学质量，一些有效的教学指导可能无法体现在标准化测试中，效能高的教师能够对学生学习能力的提升产生深远影响。<sup>⑥</sup> 与此同时，以学生学业成就为导向的教学质量评估模式，由于无法直接解释和改进教学中的不足，开始受到教育界的广泛质疑，<sup>⑦</sup> 以课堂观察为基础的评估系统成为评价教师教学质量和提高教师专业能力发展的重

- 
- ① Boston, M. D., Smith, M. S. A 'task-centric approach' to professional development: Enhancing and sustaining mathematics teachers' ability to implement cognitively challenging mathematical tasks[J]. ZDM: International Journal of Mathematics Teacher Education, 2011(43): 965-977.
- ② Candela, A. G. Mathematics teachers' perspectives on professional development around implementing high cognitive demand tasks. In: C. Martin, D. Polly (Eds.). Handbook of research on teacher education and professional development[M]. Hershey, PA: IGI Global, 2017:538-560.
- ③ Sztajn, P., Wilson, P. H., Edgington, C., Confrey, J. Learning trajectories and key instructional practices. L.R. Wiest, T.d. Lamborg (Eds). Proceedings of the 33rd annual meeting for the North American Chapter for the Psychology of Mathematics Education[M]. Reno: Nevada, 2011: 434-442.
- ④ Candela, A. G. Using the Instructional Quality Assessment observation tool in a professional development capacity. M.B.Wood, E.E.Turner, M Civil, J.A.Eli (Eds). Proceedings of the 38th annual meeting for the North American Chapter for the Psychology of Mathematics Education[M]. Tucson: AZ, 2016: 418.
- ⑤ Melissa D. Boston, Amber G. Candela. The Instructional Quality Assessment as a tool for reflecting on instructional practice[J]. ZDM: International Journal of Mathematics Teacher Education, 2018(50):427-444.
- ⑥ National Research Council. Incentives and test-based accountability in public education [M]. Washington, DC: National Academies Press, 2011: 58-60.
- ⑦ Pianta, R. C., Hamre, B. K. Conceptualization, measurement, and improvement of classroom processes: Standardized observations can leverage capacity. Educational Researcher[J], 2009(38):109-119.

要方式,<sup>①</sup> IQA 在这个大背景下快速发展。

IQA 研发者在研制系统框架时,以美国国家研究协会最新发布的研究报告《人如何学习:大脑、思维、经历和学校》(How People Learn: Brain, Mind, Experience, and School)为准则,<sup>②</sup>并参考前沿的教育教学理论,提出“四个中心”研制基础,分别是“学习者中心(Learner—Centered)”“知识中心(Knowledge—Centered)”“任务中心(Assessment—Centered)”和“共同体中心(Community—Centered)”。学习者中心指教师清楚了解学生已有认知,帮助学生架构新旧知识间的联系,有意识地将元认知教学融入课程,培养学生自我监控和自我决策能力。知识中心要求教师深入全面理解所教学科的内容知识,在教学过程中提供足够实例帮助学生掌握学科核心概念和解决复杂问题。任务中心指教师需要帮助学生明确教学目标,让学生清楚知道自己应该掌握和能够习得的内容,布置的任务和作业不能仅停留“识记”层面,要注重考查学生问题理解和解决能力,任务设计满足学生高认知水平要求。共同体中心要求教师在教学过程中营造学生积极探讨和思考的学习氛围,建立互帮互助、共同进步的学生学习共同体,学生在共同体中充分表达、倾听、交流想法意见,锻炼协同合作与交流能力。在“四个中心”理论基础上,IQA 梳理总结前人的课堂观察和教学评估工具,形成评估系统框架和评估要点。IQA 主要借鉴由美国匹兹堡大学主持的一项国家层面的教育改革项目“定量理解:提高学生学业成就和推理能力(Quantitative Understanding: Amplifying Student Achievement and Reasoning,简称 QUASAR)”。<sup>③</sup>QUASAR 可以被认为是当时最有实证基础的评价体系,具有学科针对性,采用以量化为主的方式记录分析课堂活动,分维度评估课堂教学质量,<sup>④</sup>但仍存在两方面的不足:(1)QUASAR 侧重于对任务本身设计与实施的评估,没有充分体现学习共同体的形成和发展;(2)QUASAR 的教学观偏向以教师为中心的课堂,缺少学生自我监控、决策等能够体现学生主动性的评估指标。因此,IQA 研究者删除 QUASAR 中理念陈旧的观测点,补充有关学生主动性和学习共同体的条目,形成 IQA 课堂观察评估工具。同时,IQA 还增加师生访谈和教学资料的收集,作为教师课堂观察的补充材料,共同作用评估教师教学质量。

## 二、IQA 教学质量评估系统的结构

### (一) 评估要点

IQA 教学质量评估框架主要用于评估者现场听课和对教师提交资料的评价。不同于教学视频录像可供研究者多次反复观看,现场听课具有较强的信息及时性,需要评估者快速

① Stein, M. K., Matsumura, L. C. Measuring instruction for teacher learning[M]. CA: Sage, 2008: 179-205.

② Bransford, J. D., Brown L., Cocking, R. R., (Eds). How people learn: Brain, mind, experience, and school[R]. Washington, DC: The National Academics Press, 2000: 1-374.

③ Melissa Boston. Assessing Instructional Quality in Mathematics[J].The Elementary School Journal, 2012 (01):76-104.

④ Johansson S, Myrberg E. Teacher specialization and student perceived instructional quality: what are the relationships to student reading achievement?[J]. Educational Assessment, Evaluation and Accountability, 2019 (04): 28-41.

准确捕捉评估要点，因此 IQA 具有较少的评估维度和子维度，且每个子维度的水平区分十分清晰，便于打分者快速作出恰当判断，提高工具的信效度。IQA 评估要点分为三个维度：任务认知需求（Cognitive Demand of Task）、解释性话语（Accountable Talk）和教师期望（Teacher Expectations），每个维度包含 2-5 个子维度作为评分要点。表 1<sup>①</sup> 以数学学科为例，列出 IQA 的评估维度、子维度和各个子维度的评估要点，包括对现场课堂教学

表 1 IQA 评估维度及现场课堂教学和教师提交材料评估要点

评估维度	现场课堂教学评估要点	教师提交材料评估要点
任务认知需求 Cognitive Demand of Task	<p>任务潜能（Potential of the Task）：教师设计的课堂教学任务能够满足学生潜在的高认知水平需求。</p> <p>任务实施（Implementation of the Task）：学生能够在解决课堂任务的过程中，锻炼和发展数学思维，提高认知水平。</p>	<p>任务潜能（Potential of the Task）：教师设计的课后作业能够满足学生潜在的高认知水平需求。</p> <p>任务实施（Implementation of the Task）：学生能够在完成课后作业中，锻炼和发展数学思维，提高认知水平。</p>
解释性话语 Accountable Talk	<p>参与（Participation）：学生能够参与课堂活动，发表自己的观点（不包括全班齐答）。</p> <p>教师关联（Teacher's Linking）：教师在鼓励促进学生分析拓展问题，评价反思数学问题和其他学生想法过程中所作出的努力。</p> <p>学生关联（Students' Linking）：学生支持赞同或反对其他同学的观点，并能够给出自己的解释。</p> <p>教师引导（Teacher's Press）：教师鼓励学生解释表达自己的想法，论证自己的观点。</p> <p>学生反馈（Student Responses）：学生能够证明、解释和论述数学解题方法和思想。</p>	无
教师期望 Teacher Expectations	<p>期望严格程度（Rigor of Teacher's Expectations）：教师期望学生完成课堂任务所需要达到的思维和能力水平程度。</p> <p>清晰详细（Clarity and Detail）：教师在教学过程中能够清晰详尽地表述对学生完成任务的期望情况。</p> <p>学生获取途径（Student Access）：学生能够清楚了解教师在课堂任务中对自己的期望。</p>	<p>期望严格程度（Rigor of Teacher's Expectations）：教师期望学生完成课后作业所需要达到的思维和能力水平程度。</p> <p>清晰详细（Clarity and Detail）：教师在作业说明和作业包封面上能够清晰详尽地表述对学生完成任务的期望情况。</p> <p>学生获取途径（Student Access）：学生能够清楚了解课后作业中教师对自己的期望。</p>

① Overview of the Instructional Quality Assessment[EB/OL]. [2020-06-12]. <http://d-scholarship.pitt.edu/26209/>.

学和教师提交材料的评估,其中解释性话语维度只用于评估教师现场课堂教学。

## (二) 评估方法与任务

IQA 以量化评分方式为主,结合质性评分依据,通过多种教学质量相关材料的收集,采用课堂观察、问卷、访谈和作品分析的方法,确保评估者能够准确、高效地进行教学质量评价。评估任务主要包括三项:教师教学现场观察、师生访谈和教师提交资料。其中教师访谈发生在课后,学生访谈发生在课堂进行中,访谈的目的是帮助评估者更好理解课堂活动,这部分内容将作为依据补充,使评估者能够更为合理地对现场教学进行评估。

任务一:教学现场观察。每位评估者需要观察参评教师连续两节课(每节课约45分钟)的现场教学。参评教师在每节课被观察之前,需要完成一份问卷,简要描述教学目标、教学活动和班里学生情况,帮助评估者对教学任务设计进行初步判断。在正式观察过程中,评估者利用 IQA 现场教学评估框架对全部的10个要点进行打分。每个观测点分为四个水平,由低到高分别对应1—4分:水平1为最低水平,表示在该子维度教学表现最不符合评分标准,水平2表示教学在该子维度表现基本合格,水平3表示该子维度教学表现良好,水平4代表在该子维度上教学应该呈现的理想状态。如果课堂中某个子维度观测点未出现或无法判断,则用 N/A (Not Applicable) 表示“不适用”或 N/K (Not Known) 表示“不确定”。评估者除了对教学质量量化打分外,还需要在观测要点旁边简要记录打分依据。

任务二:师生访谈。教师访谈:评估者在课后马上对教师作一个简短访谈,主要是获取教学设计和实施的背景信息,了解教师出于何种原因设计教学目标和教学过程,教学设计和实施间是否存在差异,存在差异的原因是什么。以此帮助评估者更深入地了解教师的课堂教学,给出最为合理正确的评估分数。学生访谈:评估者在上课过程中(一般在学生处理教师布置的教学任务时)对学生进行3—5分钟的访谈,主要考察学生学习管控能力和是否清楚了解教师所布置的任务中对学生完成情况的期望水平。访谈问题包括“你觉得你现在完成的任务还能够进一步完善么”“如果我是班里新来的一个学生,你能告诉我如何顺利解决老师提出的这个问题么”等。师生访谈结束后,评估者给出教师现场教学的最终得分。

任务三:教师提交资料。每位被评估教师提供4份学生作业,其中两份是教师认为对学生有特殊挑战性的(Especially Challenging)作业,另外两份为学生近期完成的作业。教师将每份作业装订成册,并撰写封面,封面包括留该份作业的目的和评分标准,作业对今后的教学有什么帮助,学生完成该作业过程中能够获得怎样的自我学习审视机会,对学生自我决策能力和思维发展有怎样的提高。教师还需要提供学生作业中低、中、高三个不同水平的典型案例各两份。评估者采用 IQA 教师提交材料维度框架对此打分,评估子维度为现场教学评价框架10个子维度中的5个,评估要点的具体内容表述与现场教学版相比稍有不同。

### (三) 评估原则与效果保证

IQA 教学质量评估系统在使用过程中需要遵循四个原则：(1) 评估不仅考核教师教学质量，还要以提高教师教学能力为根本目的；(2) 评价要紧密围绕“四个中心”，设计能够满足学生认知需求的的教学任务，并有效实施，关注学生思维能力的发展，培养学生合作交流和批判创新能力；(3) 被评估教师能够通过评估框架、评估过程与结果清楚了解自身教学存在的优势和不足，明确提升的目标与方向；(4) 评估工具需要具备良好的信效度。

IQA 教学评估系统具有一定的主观性，为了保证评估效果的合理性，研究者设置如下保障机制：多种方法相结合的数据收集方式、系统的评分者培训模式、标准化的评估体系和详细的评分示例。其中第一种效果保障机制已在前文论述，此处主要介绍后三个方面。

#### 1. 评分者的选择与培训

IQA 的评分者可以是中小学一线教师、教研员、教学督导或大学教师，但要求对所评估学科的本体性知识有一定了解，除此之外对评分者教学年限、职称、学历、培训经验等并无严格要求。IQA 评分者要参加为期 4 天左右的资格培训，具体培训流程如下：首先，培训专家组织评分者学习 IQA 开发背景和评估理念，了解 IQA 评估框架、维度与相应评估要点；接着，培训专家通过课堂和作业实例分析各评估要点的得分表现，对水平接近（如水平 3 和水平 4）的实例进行重点分析，解读异同；之后，评分者分以小组为单位观看教学视频，分析教师提交的资料，在组内讨论其得分表现，小组成员讨论后存在异议的地方征询培训专家意见，直至组内所有成员意见达成一致；最后，以个人为单位单独对教学视频和文本材料打分，同一组材料由 3 人或以上进行评估，对于组内评分不一致的维度，若集体讨论后仍无法说服对方，则须由培训专家一起参与讨论。在此过程中，受训者的评分依据由最初的模糊变得逐步清晰，如一位评分者在培训之初评估“教师引导”时，给出的依据既不清晰又缺乏自信，“呃，教师应该是在学生讨论前说了些什么……来引导学生讨论，就在教师读完课文的时候”，而在培训快结束时，该评分者可以给出十分具体有说服力的评分依据：“教师一直鼓励并追问学生‘你是怎么想的？能给我们一个具体例子来说明你的理解么？’以此来引导学生表达论述自己的观点。”<sup>①</sup> 评分者在反复实例操练讨论过程中熟悉评分标准，掌握区分评估要点中不同水平的行为表现，在获得培训专家认可后，正式拥有 IQA 评分资格。

#### 2. 标准化的评估体系

为了最大限度提高 IQA 的信度和效度，IQA 使用手册中详细描述每个评估要点概念，多数评估要点的不同水平表现都有数字化指标加以区分，帮助评估者快速准确判断教学质量在某个子维度所处的相应水平，具有较强的可操作性和客观性。以“解释性话语”

<sup>①</sup> Amy C. Crosson, Melissa Boston, Allison Levison, Lindsay Clare Matsumura, Lauren B. Resnick, and Mikyung Kim Wolf. Beyond Summative Evaluation: The Instructional Quality Assessment as a Professional Development Tool[R]. Annual Meeting of the American Educational Research Association Meeting, San Diego, CA, 2004: 11-12.

维度中的观察要点“学生反馈”为例，其四个水平的标准如下：

水平 4 (4 分)：班里有 3 个或以上的学生能够解释和论证自己的观点，且具有较强的逻辑性。

水平 3 (3 分)：班里有 1—2 个学生能够解释和论证自己的观点，且具有较强的逻辑性。

水平 2 (2 分)：学生尝试解释和论证自己的观点，但表述不恰当且逻辑性不强。

水平 1 (1 分)：学生只能给出某些想法，但无法解释论证自己的观点。

### 3. 详细的评分示例

为了便于评分者准确理解和使用评分标准，IQA 手册中详细列举了不同水平的表现行为，用实例描述的方式具体解释每个评分要点不同水平间的区别。以数学学科“任务需求”维度中的观察要点“任务潜能”为例，其四个水平的详细示例如下：

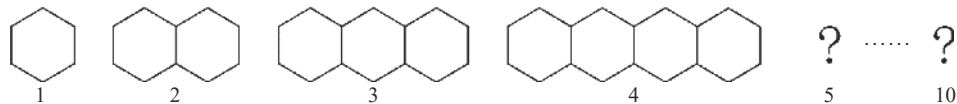
水平 4 (4 分)：设计的任务能够帮助学生理解和探索数学概念、程序和关系的本质，譬如做数学 (Doing Mathematics) 和程序间建立联系 (Procedures with Connections)。做数学指运用复杂的非算术思想解决非常规问题，程序间建立联系指将一般方法应用于具体问题解决。任务中有能明确促进学生理解和思考的表述或指示，如要求学生解释为什么选用某个方法解题；图形找规律并证明规律一般化；用数学语言建立并证明表征、策略与数学概念或程序间的联系等。

水平 3 (3 分)：设计的任务能够引发学生深层次思考，形成建立数学概念、程序间联系的意识，然而该潜在任务没有达到水平 4 是由于：任务中没有明确促进学生理解和思考的表述或指示；任务虽然要求学生做数学或建立数学程序间的联系，但是任务中的数学内容不适合所教学生的学情（如对学生而言太简单或者太困难的题目都无法促进学生高认知水平发展）；要求学生找图形规律，但并未要求学生证明或推广规律至一般化；要求学生用多种策略或表征解决问题，但并未明确要求学生建立所用策略或表征间的联系；只要求学生猜想推测，并未要求学生给出数学解释或理由。

水平 2 (2 分)：潜在任务中没有要求学生建立数学概念或数学解题思想间的联系，任务仅聚焦于正确答案的获得（如解决常规问题、重复练习计算等），而非学生数学思想的形成与发展，或者任务中的数学内容低于学生现阶段能力要求两个年级以上。

水平 1 (1 分)：潜在任务几乎无法促进学生对于数学概念、公式、规则和性质的识记和理解。任务不要求学生建立概念、公式、规则和性质间的联系。

例如题目：下图中每个图形均由边长为 1 的正六边形拼成，计算前四个图形的周长，不画图推断第 5 个图形和第 10 个图形的周长。



分析：该任务要求学生根据图形找规律，依旧停留在对学生具体形象思维的要求阶段，并未要求学生证明或推广规律至一般化，也未要求学生给出数学解释或理由，因此处

于“任务潜能”指标的水平3。若题目要求学生给出第N个图形的周长,并给以证明,加强对学生抽象思维的培养和拓展,则此任务在“任务潜能”方面能达到水平4。

IQA采用科学严格的评分者培训流程和标准化的评估体系,并详细列举不同水平的行为表现和相近水平的区别,使评分者充分理解其评估维度与要点,保证教学质量评估过程中的可信度和有效性。已有研究表明,IQA在使用过程中具有良好的内部一致性,能够反映被测评教师教学的水平。<sup>①②</sup>

### 三、IQA 教学质量评估系统的特点

#### (一) 评估要点聚焦化,凸显教学质量本质属性

美国IQA教学质量评估系统以“四个中心”为基础,从任务认知需求、解释性话语和教师期望三个维度(包括10个观测点)出发评价教师教学质量,聚焦影响教学质量的关键属性。教学发生场所为课堂,包括教师讲授、师生互动、生生互动等多种行为表现,具有复杂的情境性和社会性,包罗要素繁杂。IQA评估框架注重教学实质,剔除形式化的教学表现(如是否使用多媒体技术、是否组织小组活动)为评估切入点,弱化无法直接体现学生思维发展和知识获得的观测点,以学生产出为导向,聚焦少数起决定因素的关键评估要点。“少而精”的评估要点能够帮助评估者更为准确掌握评价标准,在稍纵即逝的课堂教学中快速判断教学水平,提高评估工具的信效度。同时,也能够帮助教师更好地理解教学意义,提升学生能力发展水平。

#### (二) 评估工具专业化,保证评价效果科学权威性

美国IQA教学评估是一个“分科化”的评价系统,它针对不同学科分别研制相应的评分标准,每个学科教学质量的评估都拥有其独有的“学科性”特质,使评估能够更为准确、科学地反映不同学科教师在各自学科的教学情况。IQA评分者要求具有一定的相关学科背景知识,经过严格的培训和资格认定后才能正式成为IQA评估者。IQA手册含有丰富详细的评估实例,用于描述每个水平的表现形式和观测依据、解释相近水平的差异特征。在IQA正式投入使用前,共有124名学科专家对评估框架的合理性和有效性进行审核;超过200名教师作为预研究对象,研究结果显示IQA具有良好的评分者内部一致性,并且能够真实客观反映所评估学科的教学质量。<sup>③</sup>专业化的评估工具和科学严谨的使用流程保证了评分系统的信效度,加强了评价效果的科学性与权威性。

① Lindsay Clare Matsumura, Sharon Cadman Slater, Mikyung Kim Wolf, Amy Crosson, Allison Levison, Maureen Peterson, Lauren Resnick. Using the Instructional Quality Assessment Toolkit to Investigate the Quality of Reading Comprehension Assignments and Student Work[R]. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles, CA.2006: 1-57.

② Anne Garrison Wilhelm, Sungyeun Kim. Generalizing From Observations of Mathematics Teachers' Instructional Practice Using the Instructional Quality Assessment[J]. Journal for Research in Mathematics Education, 2015, 46(3): 270-279.

③ Brian Junker, Yanna Weisberg, Lindsay Clare Matsumura, Amy Crosson, Mikyung Kim Wolf, Allison Levison, Lauren Resnick. Overview of the Instructional Quality Assessment[R]. The Regents of the University of California, 2006: 14-29.

### (三) 评估方式多元化, 力求多角度真实反映教学质量

评估系统将课堂观察、访谈、问卷和作品分析等方法相结合, 利用“三角互证”的思想搜集评估证据, 力求保证评估信息的全面真实性。研究表明, 教师在被随机抽取的某一节课中的表现, 往往难以全方位展现其真正的教学水平, 被评估的课越多, 教师的教学表现越趋于稳定, 但若增加其他相关资料加以辅助评估, 则可在一定程度上降低评估的课时数。<sup>①</sup> IQA 评估参评教师的连续两节课, 同时通过问卷、师生访谈等方式更好了解教学目标、教学任务和教学效果, 帮助评估者快速准确掌握教学活动目的与表现, 了解更为真实的课堂。教师提交的学生作业能够进一步反映学生的思维发展和知识获得, 是对教学效果的深层次检验。教师提交学生作业低、中、高三不同水平的典型案例各两份, 能够反映教师自身对任务所能达到的不同认知需求的理解, 了解教师所认为的高水平思维表现。多元化的评估方式, 多角度的证据累积, 能够提高教学质量评估的真实度与可靠性。

### (四) 评估目标功能化, 深化教师专业发展

IQA 评估系统研发的初衷便是为了解决教育评价界中存在的工具单一诊断性目的, 而面向评估系统带来的改善性和导向性功能。IQA 使用原则中要求评估不仅是对教师教学质量的考核, 应主要基于提高教师教学能力的发展性目的, 被评估教师能够通过评估框架、评估过程与结果清楚了解自身教学存在的优势和不足, 明确提升的目标与方向。教师可通过 IQA 评估要点了解优秀课堂教学的行为表现, 利用丰富的不同维度水平的实例学习, 寻找自身教学中存在的差距, 例如一位语文教师在提问学生回答时总会说“你的答案是什么”, 在学生给出答案后便不再追问, 在学习 IQA 评估手册后, 他发现可以利用进一步的追问如“你为什么这么说”或“关于这个问题还能够多说一些你的想法么”等, 来鼓励学生充分阐释答案背后的想法与思考, 锻炼学生科学化与逻辑性思维。IQA 能够帮助教师明确教学手段和策略, 使有效的教学方式成为一种有意识的教学行为。研究表明, IQA 评估系统在帮助教师自我测评, 提升专业能力发展方面效果显著。<sup>②</sup>

当然, 任何评估系统都会存在自身的局限性, IQA 评估系统在拥有上述优势的同时, 也存在一些不足, 主要体现在: (1) IQA 采用课堂教学现场观察方式进行评估, 与视频录像相比, 虽然前者能够大幅降低设备费用、后期分析成本、保护被评测教师隐私等, 但无法用于某些教学片段的反复推敲, 较为依赖评估者对关键信息的捕捉能力; (2) 目前已有版本适用科目有限, IQA 具有学科针对性, 不同学科的系统维度和评估要点具有一致性, 但是一些评估要点的水平描述会体现其学科属性, IQA 评估学科的广泛性还需进一步拓展; (3) IQA 采用与学生学习表现直接相关的维度作为评估指标, 精要的评估指标提

① Clare, L., Aschbacher, P. Exploring the technical quality of using assignments and student work as indicators of classroom practice[J]. Educational Assessment, 2001 (07): 39-59.

② Dole, J. A. Professional development in reading comprehension instruction. A.P. Sweet, C.E. Snow (Eds.), Rethinking reading comprehension[M]. New York: Guilford Press, 2003: 145-150.

高了该工具的信度和可操作性，但是难以对教学细节做到准确的呈现与刻画，如教师语言是否存在科学性错误、教师板书呈现是否合理、教师提问是否清晰明确等；（4）IQA需要研究者观察参评被评估教师连续至少两节课（每节课约45分钟）的现场教学，再结合访谈、问卷、作品分析等多元化相关资料，呈现被评估教师教学质量的具体分析报告。这导致该评估工具难以适用于教学时长较短（一般介于15—45分钟）的教师技能大赛，在一定程度上限制了IQA评估工具的实际推广度。

#### 四、IQA 教学质量评估系统的启示

##### （一）简化评估指标，注重教学实质

教学质量评估应聚焦于能够反映教学实质的维度和指标，剔除形式化非本质的行为观测。评估要点“少而精”的优势在于，一方面帮助评估者快速精准把握课堂教学质量，提高评估信效度；另一方面帮助被评测教师有针对性地了解自身教学表现，理解“什么是一堂好课”的本质，将评估要点贯穿于教学设计、教学实施和反思之中。心理学研究表明，繁琐复杂的评估指标不利于人类记忆和使用，会在一定程度上降低其推广性和使用成效。<sup>①</sup>目前我国的教学评估系统正处于快速发展中，评估指标依然存在如下问题：一是评价维度基本是对教学过程要素的分解，将“教学目标”“教学内容”“教学方法”“教学过程”等评价指标简单罗列，看似覆盖了课堂教学的方方面面，但是却没有抓住有意义教学的本质特点，不能很好揭示各个指标间的内在联系，难以体现教学的整体效果；二是评价子维度过多且繁杂，有的评估系统甚至包含50个以上评估要点，<sup>②③</sup>容易造成评分者脱离评分具体指标，依靠原有经验打分，失去评分指标体系的意义和价值。我国可以借鉴IQA教学质量评估系统的指标设计，缩减精华评估指标，抓住教学本质，突出教学有效性，提高评估系统的可操作性和应用价值。

##### （二）分科化评估工具，保障评估结果专业性

不同学科的优质课堂教学存在一定共性，如鼓励激发学生深层次思考，培养学生表达能力等，但也都具有其学科的独特属性和核心素养，如数学能够锻炼学生运算能力，语文则需要培养学生的文化传承与理解。分科化的评估工具能够更为精准地把握所评估学科的教学质量，通过对评估要点不同水平的具体学科表现描述，使评估具有更强的专业性和可操作性。我国现有的教学质量评估系统多数不具有学科特色，<sup>④</sup>这就使得评价工具看似学科普遍性很高，适用范围很广，实则难以把握学科特有育人价值和核心能力培养，削减评估结果的专业化价值。除此之外，我国教学水平评价者存在学科专业化程度不高的现

① Alan H. Schoenfeld, Robert Floden, Fady El Chidiac, Dennis Gillingham, Heather Fink, Sihua Hu, Alyssa Sayavedra, Anna Weltman, Anna Zarkh. On Classroom Observations[J]. Journal for STEM Education Research, 2018, (01): 34-59.

② 陈佑清, 陶涛. “以学评教”的课堂教学评价指标设计[J]. 课程·教材·教法, 2016, 36(01): 45-52.

③ 郝志军. 中小学课堂教学评价的反思与建构[J]. 教育研究, 2015(02): 110-115.

④ 韩继伟, 张晓霞. 美国教师教育专业认证及启示[J]. 数学教育学报, 2018, 27(02): 3-9.

象。正如没有人能够教自己不熟悉的知识一样,也没有人能评价自己不了解的学科。<sup>①</sup>因此,需要具有所评学科专业背景的人员,在经过严格的评估培训、对学科知识有深层次把握后,方可合理使用分科化的评估工具,保障评估结果的专业性和权威性。

### (三) 更新评估方式,提高评估信息真实性

我国目前对课堂教学质量的评估方式主要有以下四种:以学生学业成就为导向的结果性评价、教师教学能力水平纸笔测试、学生对教师教学评价的问卷测量和课堂教学观察。<sup>②</sup>学业成就的结果性评价无法反映教学发生的真实过程,难以对教师教学提供有效反馈和指导,且容易让教师为了提高学生分数而采用“应试”教学,不利于学生素养的全面发展;教学能力水平纸笔测试则更多考察知识点和结论性内容,教师对学科内容的熟练掌握和教学理论的清晰解读,与其课堂教学实施能力有时并不具有一致性。课堂教学考察的是教师在实际情境中,对学科内容知识、教学法知识、学习者知识等的综合运用能力,无法脱离真实教学场景的存在;单一学生视角的问卷测评,由于受限于学生自身知识水平、学习经验和主观感受等因素,难以保证对教师教学水平的真实呈现。以课堂观察为基础进行教学评估,辅以专业的评估团队,采用多元化的评价方法,才能保证教师教学水平的真实展现,维持评价的内在价值。可以考虑借鉴 IQA 的评估模式,以课堂观察为主体,同时采用访谈、问卷、作品分析等多元化的方式,从不同角度搜集评估证据,帮助评估者了解更为真实的课堂行为,包括行为背后的原因、行为效果等,结合评估框架对教师教学表现给出准确合理判断。

### (四) 完善评估理念,加强评估发展性功能

美国 IQA 教学质量评估系统着重于提升教师课堂教学实践能力的发展性评价。近年来,教育测评的评估理念逐渐由“对学习的评价 (Assessment of Learning)”,转变为“促进学习的评价 (Assessment for Learning)”<sup>③</sup>和“评价即学习 (Assessment is Learning)”<sup>④</sup>。教学评价不再仅是教师评比、晋升的手段,研究者普遍意识到其改善发展性功能,使之成为教师专业成长的必要助力。<sup>⑤</sup>课堂教学评价标准构建的合理性标准在于对课堂教学实践具有很强的解释力和指导力。<sup>⑥</sup>我国在开发研制教学质量评估系统时应始终贯穿“以评促教,评价即学习”的理念,将提高教师专业发展作为评价的首要目标。首先,评价指标要去形式化,注重教学质量本质,加强对教师专业发展的导向性;其次,评估要点的不同水平要包含丰富详细的实例,在帮助教师利用评估系统进行合理自评的同

① 黄晓,孙丽伟. 小学科学教学设计的规范化和学科化 [J]. 全球教育展望, 2014(04): 111-120.

② 赵轩. 基于多元化理论的中小学教师资格考试数学学科试卷质量分析 [J]. 数学教育学报, 2018, 27(02): 38-40.

③ Assessment Reform Group. Assessment for Learning: 10 Principles-research-based Principles to Guide Classroom Practice[EB/OL]. [2020-02-18]. <http://k1.ioe.ac.uk/trp/arg/CIE3>.

④ Alonso-Tapia, Jesus, Garrido-Hernansaiz, Helena. Assessment for Learning: Assessment of Non-Written Sources Understanding [J]. Electronic Journal of Research in Educational Psychology, 2017, 15(01): 168-182.

⑤ 钟启泉. 课堂评价的挑战 [J]. 全球教育展望, 2012(01): 10-16.

⑥ 裴娣娜. 论我国课堂教学质量评价观的重要转换 [J]. 教育研究, 2018(01): 17-29.

时,指明更高水平的行为表现,作为教师教学改进方向和指导依据;最后,参评教师要在评估者的帮助下理解评估理念与目的,避免“作秀式”课堂,充分利用评价过程剖析自身教学上的优势与不足,进行深度教学反思,提升教师专业发展。

(2019年度教育部人文社会科学青年基金项目“我国中小学课堂教学质量评价模型的构建与应用研究”(项目编号:19YJC880145)。)

## Exploration and Reflections on Instructional Quality Assessment (IQA) System in the U.S.

ZHOU Jiushi<sup>1</sup>, BAO Jiansheng<sup>2</sup>

(<sup>1</sup>Faculty of Education, Tianjin Normal University, Tianjin 300387, China)

(<sup>2</sup>Mathematics Department, East China Normal University, Shanghai 200241, China)

**Abstract:** An effective instructional evaluation system is able to judge teaching level scientifically, and guide teacher to accomplish his or her developmental goal. Instructional Quality Assessment (IQA) is developed by the University of Pittsburgh, comprising with three domains: Cognitive Demand of Task, Accountable Talk, and Teacher Expectations. IQA is used for live classroom observation and student assignment collections. Different from teaching video analysis tools, live lesson assessment entails observers capturing meaningful information timely. As a result, IQA is parsimonious with focus on a small number of domains, and the levels under each dimension are clear to make judgement. The following four characteristics of IQA were identified: (a) emphasizing the teaching nature with a small number of focused domains; (b) discipline-based designing to ensure that the evaluation effect is scientific and authoritative; (c) multiple evaluation methods adopted to reflect teaching quality from different angles; and (d) taking teacher development as essential evaluation aim. IQA evaluation system provides suggestions for teaching evaluation research and teacher development in China, including simplifying evaluation domains to focus on teaching and learning, using content-specific tools to enable better understanding of assessment results, taking multiple methods to grasp real information of classroom teaching, and emphasizing the development function in evaluation.

**Key words:** teaching quality, teacher development, evaluation system, classroom teaching