

从新世纪初呼唤至今的评价改革依然“犹抱琵琶半遮面”。高中学业水平考、教学质量评价体系似乎都在“难产”。评价的一小步决定了素质教育推进的一大步。

PISA会成为评价改革的引擎吗

文 / 陆璟



2009年上海将首次参加OECD国际学生评估项目(PISA)。该项目测评的是:在接近义务教育末期时,学生是否已经具备了进入社会所需要的知识和技能。PISA评价学生的阅读素养、数学素养、科学素养,但它并非测量学生掌握学校课程的程度,而是展望学生在多大程度上能将学到的东

西推及其他,以及在新的环境中应用他们学到的知识和技能,所以它是一项前瞻性的评价。我们可以从PISA的评价理念和方法中,看到教育评价,尤其是对学生学习质量评价的一些发展趋势。

PISA的“素养”指的是学生应用所学知识和技能的能力,分析、推理

和有效沟通的能力,以及在不同的情境中解决和解释问题的能力。

在此不具体介绍PISA对阅读素养、数学素养、科学素养的定义,只通俗地比较它与传统意义上学校考试内容的区别:PISA阅读素养指的是理解、运用和反思书面文本的能力,它评价的不仅仅是文学阅读能力,还把阅读作为学习新知识、参与社会活动以及完成工作或任务的工具,体现了“大阅读观”;PISA数学素养指的是当学生遇到问题时,能够多大程度上激活其所具有的数学知识和数学能力并解决这些问题,它与传统学校意义上数学概念最显著的区别在于关注“现实生活中的数学”,而非“离开生活的抽象数学”;PISA科学素养并不局限于学生对科学内容的掌握程度,而是重视学生发现科学问题,科学地解释现象以及在生活中遇到涉及科学和技术问题时能够运用科学证据来解释、解决和作出决定的能力。

PISA的素养概念是在不断发展的,从2006年开始,PISA的素养指的不仅仅是能力的范畴,还将学生对学习的兴趣、态度和参与度等非认知因素也纳入了素养评价的框架,更好地体现了终身学习的要求。

为了评价学生应用知识而不是记忆和再现知识的能力,PISA将测评任务融入到多种现实生活情境中,涉及的题材非常广泛,学生只需综合应用已有的能力,无需死记硬背。

这种前瞻性的评价方法能够反映在全球化和信息化时代中现代社会和职场中最有价值的人类技能。麻省理工学院的Levy和哈佛研究院的Murnane分析了过去几十年来美国劳动力市场对工作要求的变化,发现过去十年中需求下降最快的工作不是过去常常提到的手工劳动,而是常规认知工作,即:那些能够很好地用演绎或归纳方式描述的、能用计算机替代的脑力劳动。相反,那些需要复杂交流和人际互动的工作以及要求专家思考的工作则显著增加。这说明在OECD国家中,用计算机可以替代的工作、或低工资国家的工人可以用更低成本完成的工作将不断消失。如果学生仅仅学会记忆或再现知识的话,那么将来能胜任的工作岗位将越来越少,虽然目前大多数中产阶级仍在从事这类工作。

因此,为了全面参与未来的全球经济,学生们要能够解决那些不能依靠清晰的规则来解决的问题,还要能够清晰地、有说服力地交流观点。

质量指的不单单是学生总体的成绩或能力水平,还要看学生的学习兴趣,教育的公平、效率和有效性。

首先是学生成绩和学生对学习有兴趣、态度和参与度等非认知因素的关系。学生成绩好但学生学习兴趣低或不积极投入学习的国家或地区,不能算是高质量的,因为如果没有积极的态度,即使学生掌握了知识和能力,也不会有终身学习的内在动力。PISA2006测试后发现,我国香港、澳门、台湾地区的科学认知测试成绩相对都比较好,但科学态度(对科学学习的兴趣等)都不太理想,这需要引起重视。

其次是学生成绩和教育公平的关系。PISA研究不同社会经济背景的学生能否获得同样的学习机会、达到同样的能力水平。2000年PISA测试结

果公布后,德国的教育受到批评,不仅因为成绩低于OECD平均水平,还因为德国不同社会经济阶层之间教育机会不公平。德国过早分流的教育体系使那些父母社会经济地位低的学生由于缺少早期教育机会而被分配到学业水平比较低的学校中去,但实际上这些学生可能并不缺乏学习的潜力。相反,芬兰的教育却倍受推崇,因为芬兰学生不仅成绩好,而且不同学校之间、不同社会经济地位学生之间的差异小。

第三是教育投入的效益。学生成绩比较好的芬兰、韩国、日本等,教育投入水平在OECD国家中属于中等偏上,而投入很高的美国,其学生成绩却低于OECD的平均水平。说明取得好成绩需要高于平均水平的教育投入,但并不是说教育投入越高越好,高于一定投入水平后,就要看资源配置中的结构和效益了。

第四是教育系统的有效性。通过长期比较,分析教育政策和学校教学组织能在多大程度上提高学生的成绩,以及削弱家庭社会经济地位不利对学生成绩产生的不良影响。

发达国家很重视教育质量监督,但是几乎都不采用全样本统测的方法,而是采用抽样调查来了解教育系统的整体成效,以此减轻对学校和学生个体的压力。PISA、TIMSS、PIRLS、NAEP等都是抽样调查研究。每个国家或地区参加PISA的学生样本一般为4500~5000人。

PISA设计了系统的理论框架,将能力区分为不同的类型和层次,并根据能力框架出题,目的是评定学生不同领域的能力水平,而不是选拔或筛选。

同时PISA在数学和科学评分中,不仅区分对错,还把学生的正确答案和错误答案都区分为不同的类型,研究学生的思考方法,为改进教学提供参考。

OECD将组织对PISA项目的外

部评估,通过问卷调查对5个国家或地区(加拿大、中国香港、挪威、波兰和西班牙)的案例进行研究,从相关性、有效性、可持续性、未预料的影响等4个方面,评价PISA项目对参与国家和地区是否物有所值,该项目在多大程度上达到了参与国家和地区的预期目标和他们的重点要求,在多大程度上影响了教育政策。目前,加拿大已经对近3万名学生进行了长期跟踪调查(青少年过渡期调查,YITS),结果发现学生的PISA成绩与他们接受中学后教育的比例有非常显著的关系,而是否接受中学后教育又与学生未来的收入有非常显著的关系。15岁时PISA阅读能力在1级水平及以下的学生到19岁时只有28%接受了中学后教育,而2级水平的学生这一比例增加到45%,3级水平是65%,4级水平是76%,5级水平是88%。即使是考虑了诸如性别、家长教育水平、母语、家庭收入以及居住地等因素,年青人的阅读水平仍然与接受中学后教育的比例有非常显著的关系。说明PISA在评价教育质量方面是有效的。

基于PISA项目的成功经验,OECD正在拓展这一研究。一是PIAAC,是对成人学习能力的评价,也就是PISA的成人版,将采用以计算机考试为主、纸笔测验为辅的方法,研究成人的个性特征、关键认知能力、教育程度和能力形成的过程、能力在工作和其他方面的应用、与工作收入的关系等,2007~2013年为PIAAC的首个研究周期,有27个国家和地区参加,首次测试将在2011年进行。

二是AHELO项目,是高等教育学习质量评价,也就是PISA的大学版。从2008年起正进行为期3年的可行性研究。

(作者系上海市教育科学研究院普通教育研究所副所长,PISA上海项目组秘书长,副研究员)