

我们的教育评价能促进学生发展吗?(上)

◆夏正江

摘要:教育评价具有强大的导向功能,有什么样的教育评价,就有什么样的教育实践及学生发展。基于这种立场,本文对现今中小学教育评价系统中存在的妨碍学生发展的三个基本问题即进行了系统反思,并提出了相应的改进建议。问题之一是,“该考评的与不该考评的发生错位,导致教育过程的扭曲与变形”;问题之二是,“评价的标准、方式不当导致评价结果的失真与无效”;问题之三是,“对评价结果的解释与使用不合理导致评价功能的窄化与异化”。

关键词:教育评价;学生发展

“现在的考试,用对付敌人的办法,搞突然袭击,出一些怪题、偏题,整学生。这是一种考八股文的办法,我不赞成,要完全改变。我主张题目公开,由学生研究、看书去做。例如,出二十个题,学生能答出十题,答得好,其中有的答得很好,有创见,可以给一百分;二十道都答了,也对,但是平平淡淡,没有创见的,给五十分、六十分。考试可以交头接耳,无非自己不懂,问了别人懂了。懂了就有收获,为什么要死记硬背呢?人家做了,我抄一遍也好。”^[1]

上面这段话出自毛泽东 1964 年的春节谈话。看了这段话,许多人可能会吃惊,感到有些“不可思议”,因为其中包含的、传递的教育考试观念,同我们当中绝大多数人习以为常的或当下教育界主流的、正统的教育考试观念相比,似乎太过于“离经叛道”了!考题公开、即便题目没有做完只要答得有创见可以得满分、考试可以交头接耳、可以“问”、可以“抄”,诸如此类观点,即便是今天看来也显得太过“大胆”和“超前”!但仔细想想,这里面似乎也有让我们对如今的教育评价系统进行深刻反思的成份!在我看来,现今的教育评价系统主要存在以下三大问题:

问题之一:该考评的与不该考评的发生错位,导致教育过程的扭曲与变形

该考评的与不该考评的发生“错位”主要体现在:有些重要的、该考评的东西未纳入中小学现行教

育考试与评价范围,或即便纳入其中、有所涉及,也处于被忽视的边缘地位。与此相反,有些不该考的东西却反复出现在某些学科的教育考试与评价中。

一、在评价范围的分布上,偏重认知类教育目标的评估

1.表现及证据

同培养健全的人格、良好的个性心理品质相比,我们的教育评价系统更加看重的是文化知识的掌握和认知能力的发展。说我国中小学存在重“知”(智)轻“德”、重“理”轻“情”(情感、态度)的倾向,多少还是切中时弊的。因为我们很容易从学校教什么、考什么、评什么中得出这种印象:课堂上教师一味地教学生知识,考试中一味地考知识,评价中一味地根据学生考得如何来评价学生,这是大多数学校普遍存在的现象。尽管教育官员、学者们、教师们都承认培养学生良好的品格、对学生进行情感教育很重要,但流行的教育实践尤其是我们的教育评价系统却告诉我们:在学校教育中,真正重要的是掌握书本上、教材上、大纲中规定的知识、技能,通过大量训练获得做题、应试的技巧,其他东西虽说也很重要,但只停留在口头上,很少被纳入学校教育正式的考评系统之中,更无法转换、折算为学生在升学考试中的考分,因此,它们的重要性是第二位的,只能作为“软”指标供有关部门参考。

2. 危害及后果

衡量一种教学有效与否、优劣与否,不能仅从认知角度去观察、去判断,还需要道德教化和审美体验的角度去观察、去判断。从某种意义上讲,学生在情感、态度、价值观、行为习惯等方面的学习与发展,比认知方面的学习与发展来得更为重要。众所周知,教育的目的不仅仅在于让学生掌握知识、发展智力,更重要的是使学生养成良好的品格。那种仅仅注重知识传授,不注重学生德性养成、情感发展的教育,是丧失“灵魂”的教育,这种教育只能培养出没有灵魂、空心的机器人。“知识”只能教学生“正确地做事”(doing things right),却不能教会学生“做正确的事”(doing the right things),唯有正常的情感和品格,才能告诉我们哪些事情是应当去做的。英国学者、夏山学校的创始人尼尔指出,现代教育最大的问题在于知识与情感的分离,即过分抬高书本知识教育的地位,忽视了儿童的情感世界和性格的正常发展。他极力主张,学校应尽量去适应儿童,而不是让儿童去适应学校;儿童应该按他们自己的意愿去过他们自己的生活,而不是按望子成龙的父母或自以为是的教育家的看法去生活;所谓成功,就是“能快乐地工作、积极地生活”,“学校与其培养一个有神经病的学者,还不如培养一个快乐的清洁工”。夏山学校最大的成功就是,培养出了未被“恐惧”与“仇恨”摧毁的,对人友善、自信、有创造力的,健康而自由的孩子。^[2] 尼尔的这些看法为我们反思仅仅向学生传授知识的现代教育制度提供了很好的借鉴。只关注学生认知方面的发展,忽视学生情感或品格的发展,这样的教育及评价系统是不完整的、残缺不全的,也是畸形、扭曲的教育。

3. 对策与建议

虽然我们承认,学生心理是否健康、人格是否健全,是否喜爱学习、敢于探索、具有社会责任感,像这样一些宝贵的品质,很难通过传统的考试办法加以评估,更难以用某个“分数”进行量化,但只要我们对它予以足够的重视,总能找到适当而有效的办法对其进行评估。比如,教师可以使用李克特自陈量表,让学生以匿名的方式对问卷中的提问做出反应;教师还可以对发生在自然情境中的真实的学生行为表现进行观察,把观察到的、认为特别重要的或有意义的、典型的、发生在学生身上的事例或事件片段用简

短的文字记录下来;教师也可以运用同伴提名法,给学生呈现一系列简短的有关学生的行为描述(如班上愿意同其他人分享物品的同学、愿意帮助他人完成作业的同学、总是鼓励其他人做得更好的同学等),要求学生写下最适合每项描述的同学的姓名;教师还可以运用访谈法,与学生进行面对面的、一对一的个别交谈,以获得有关学生行为的自述信息。在获得各种有关学生行为表现的评价信息后,教师就可以对这些信息进行综合判断,撰写有关学生行为表现的评语。这些评语年复一年地积累起来,放入专为每个学生建立的成长记录袋,就可以作为学校录取新生的一个重要的参考依据。

据说,美国的高中生能否获得毕业证书,除了学分、成绩等要求外,还要求学生能自觉自愿地参加社会公益活动,无偿地为社会服务一定的时间。比如,某高中规定学生提供社会服务的时间为:高一为5个小时,高二为10个小时,高三为10个小时,高四为35个小时。学生做义工的形式多种多样:有的到教会做事,有的做洗车工,有的到敬老院陪老人聊天,有的帮穷人修建房子,有的到建筑工地做搬运工,有的到幼儿园做“男阿姨”,有的为伤残人提供服务,有的甚至自掏腰包到贫穷国家搞扶贫活动,有的为灾区举办义捐、义演活动。除了做“义工”外,还要有一定的打工经历。在大学招办审核人员的眼里,穷人的孩子不打工是要“扣分”的(因缺乏责任感),而富人的孩子打工则可以“加分”(因自立精神)^[3]。由此可见,对学生品格方面的发展进行评估并不是做不到,关键是我们愿不愿意去做。

二、在知识获得的评估上,偏重陈述性知识和显性知识的考察

1. 表现及证据

在列举表现与证据之前,有必要先对相关概念作一个简要的解释。所谓“陈述性知识”(declarative knowledge),主要是相对于“程序性知识”(procedural knowledge)而言的,主要说明事物“是什么”、“为什么”和“怎么样”,是个人可以有意识地回忆出来的关于事物及其关系的知识,常以事实、概念、原理、法则、命题等形式出现。“程序性知识”则是关于“怎么办”或“如何做”的知识,或者说是关于完成某项任务的行为或操作步骤的知识,有时候,人们又把这种知识称为

“实践知识”或“过程知识”。例如,关于怎样下棋、怎样钓鱼、怎样修剪花木、怎样弹奏一种乐器、怎样开车的知识,即是一种程序性知识。程序性知识常常镶嵌在特定的实践情境之中,并通过个体外在的行为显露出来,因此,人们常常把“程序性知识”称为根据某人会做什么而推知某人所具有的知识。人们通常所讲的各种操作步骤、实践技能,包括各种学习策略、问题解决策略,本质上都属于程序性知识的范畴。

另一个相关概念是“显性知识”。显性知识是相对于“隐性知识”而言的。1958年,英国化学家、哲学家波兰尼在《人的研究》一书中首次提出“隐性知识”(tacit knowledge)的概念。在波兰尼看来,人类所有的知识基本上可分为两类,即“显性知识”与“隐性知识”。所谓“显性知识”,是指能够用语言、文字、数字和图表清楚地予以表达的知识;“隐性知识”则是一种“只可意会不可言传”的知识,是一种经常使用却又不能通过语言、文字或符号予以清晰表达的知识。隐性知识常常隐含在个体的行动之中,表现为个体的经验、眼光、趣味、技巧、诀窍、灵感、信念、习惯等,其主要特点是难以编码与度量、不能以正规的形式加以传递、不易大规模储存和传播,具有非逻辑性、非公开性、前言语性、模糊性、个体性等特点;虽然它不能用言语明确地加以表达,但却可以在行动中被展现、被觉察、被意会。

之所以说现行中小学的教育评价偏重对陈述性知识、显性知识的评估,忽视对程序性知识、隐性知识的评估,证据主要是我国中小学普遍倚重纸笔测验来对学生的进行学习评估与考核。从教育测量学的角度看,对陈述性知识或显性知识的检测,采用传统的纸笔测验(包括填空题、是非判断题、选择题、匹配题、简答题、辨析题、论述题等)是最有效的,但采用纸笔测验却无法有效地评估和检测学生对程序性知识或隐性知识的掌握。再者,从人们经常谈论的我国中小学生掌握“双基”比较牢固、扎实,但普遍缺乏运用知识解决实际问题的实践能力来看,也能从一个侧面证明我国中小学教育评价确实存在一种忽视对“程序性知识”或“隐性知识”进行评估的倾向。

2. 危害及后果

陈述性知识与程序性知识的区分,与英国分析哲学家赖尔关于“know-how”与“know-that”的区分十分相似。赖尔所讲的“know-that”往往表述为关于

事物的各种命题,而“know-how”则往往表现为各种做事的知识。赖尔指出,“愚蠢”不等于“无知”,前者是由 know-how 方面的欠缺所引起的,后者则仅仅意味着个体缺乏 know-that 方面的知识。在赖尔看来,个体“智力”的核心在于拥有“know-how”方面的知识,而仅仅拥有“know-that”方面的知识很可能是一个愚蠢的人。赖尔还认为, know-how 相对于 know-that 具有逻辑上的在先性,无论是发现还是拥有有一种 know-that 方面的知识,都要以拥有 know-how 的知识为前提;试图将“know-how”还原为、归结为 know-that 是错误的。由此看来,对学生智力或能力的发展而言,仅靠掌握和拥有“陈述性知识”(也即 know-that 方面的知识)是远远不够的,学生还必须掌握和拥有大量的“程序性知识”(也即 know-how 方面的知识)才行。

无独有偶,提出隐性知识理论的波兰尼也有类似的想法。波兰尼认为,隐性知识是显性知识的基础,一切显性知识都有其默会的根源;人类所有的知识不是隐性知识,就是植根于隐性知识;显性知识不过是大树上结出的果实,而给大树提供营养的树根则是隐性知识。从教育的角度看,教学不仅要致力于让学生掌握书本上那些显性的、明确的公共知识,还要致力于让学生掌握那些镶嵌于实践情境之中、难以言表的经验知识或行为知识。因为,一方面隐性知识是个体获得显性知识的基础,另一方面隐性知识还是个体增强实践能力的关键。人们常说的“高分低能”现象,很大程度上就是由学生掌握的知识主要局限于显性化的书本知识,缺乏从事有效的实践行动所需要的隐性知识(或称个人化的实践知识)造成的。

3. 对策与建议

要改变当前中小学教育评价中普遍存在的仅仅注重对学生习得陈述性知识或显性知识的评价,忽视对学生获取程序性知识或隐性知识的评估这一现状,除了要在教育评价体系中引入或采用适宜于程序性知识或隐性知识的评价方式(如诉诸实际操作的表现性评价)外,还必须对学校现行的偏重学术性知识的课程体系以及偏重教师讲解与直接传递的教学方式进行改革。比如,完善课程结构,建立学科(知识)中心课程、学生中心课程和社会中心课程三类课程相互支持、相互补充的课程结构体系,大力强化活动性教学、实践性教学、探究性教学在整个教学生态

中的比重,只有这样,注重对学生程序性知识、隐性知识习得状况进行考察的教育评估体系才有依托,否则,光有评价理念、方式的改变也是无济于事的。

三、在智力发展的评估上,偏重于低层次认知能力或单一类型智力的评估

1.表现及证据

前面讲到我国中小学教育比较注重书本知识的获得、认知能力的发展,这只是相对于学生健全人格的形成、正常情感的培养而言得出的结论。其实,单就知识的获得或认知能力的发展而言,我国中小学教育及其评价系统也存在相当的局限。这个局限概括起来主要有两点:一是偏重对低层次认知类教学目标(如布卢姆教育目标分类学中所讲的前几类认知目标)进行评估;二是偏重对单一类型智力(如美国学者伽德纳所讲的言语智力、数理逻辑智力或斯腾伯格所讲的分析智力)进行评估。

先看第一种局限的表现及证据。众所周知,布卢姆关于教育认知目标的分类早在上世纪60年代就已提出,不过早期的分类在实践过程中被证明存在各种缺陷。后来,安德森(Anderson, L.W.)等人对布卢姆原来的分类框架(所有认知类教育目标被分为“知识”、“领会”、“应用”、“分析”、“综合”和“评价”)作了较大的修订,新的分类框架不同于原来的地方是:采用“知识”和“认知过程”两个维度作为分类框架,“知识”涉及学习的内容或结果,依据“从具体到抽象”的顺序分为四个类别,即事实、概念、程序和元认知(学习策略);“认知过程”涉及学习的过程及心理能力,依据认知的复杂程度不同,“从低到高”分为六个类别,即记忆、理解、应用、分析、评价和创造。依据布卢姆提出的认知目标分类体系,再来观察和对照现实中的课堂教学,不难找到我们所需要的证据。大量的课堂观察和课例研究表明,在我国中小学,传统的“满堂灌”(教师一讲到底)正在被“满堂问”(边讲边问)所替代,高密度的提问已成为课堂教学的重要方式,但这种高密度、高频率的课堂提问并没有真正激发和调动学生高水平的思维技能。上海知名学者顾泠沅教授把教师在课堂上的提问分为五类,即常规管理性问题、记忆性问题、推理性问题、创造性问题和批判性问题。据其观察,在绝大多数中小学课堂上,教师的课堂提问以记忆性问题居多(74.3%),

推理性问题(21%)次之,极少有创造性、批判性问题;提问后让学生齐答或举手回答的比例很高,学生主动表达自己观点或向教师提出问题的极少出现,课堂几乎完全为教师所控制。^[4]同样的现象在教师课后布置给学生的作业中也有明显体现:绝大多数教师布置给学生的课后作业往往侧重于对课堂所讲知识的巩固与运用,学生只要提前预习一下,上课时认真听讲、记好笔记,课后及时复习,很少需要或几乎不需要投入真正高水平的思考(如对材料的分析、评价与创造),就能轻而易举地完成课后作业。类似的情况在大大小小的各种教育测验与考试中甚至在教材编写中也可以看到。

再看第二种局限的表现及证据。应该说,发展智力的重要性人们很早就已认识到。早在上世纪80年代,教育界就有学者指出,尽管智力的发展离不开知识的掌握,但掌握知识与发展智力并不是一回事,掌握知识并不会自动地导致学生智力的发展。两个学生掌握的知识可能完全相同,但认知发展的潜能很可能相差较大。这一认识相对于以前单纯强调“双基”来讲当然是一个进步。但人们对智力本身是什么的认识仍然存在一定的局限。比如,在很长一段时间内,人们谈到“智力”时,总是倾向于说某个人智力水平(或智商)高,某个人智力水平(或智商)低,但人们并没有意识到这种比较仅仅适用于单一类型的智力,如果存在多种类型的智力,如果人与人之间的智力差异不仅表现为水平的差异,还表现为类型与结构的差异,则上述比较就没有意义。

最近这几十年来,虽然国际上心理学家们提出了各种各样的智力理论,他们对智力本质的看法各不相同,但他们都对传统的智力概念提出了批评,认为传统的智力概念只涉及内涵宽广、结构复杂的人类智力的极小一部分,而且也是相对来讲不太重要的一部分,这种智力概念不能反映人类心理能力的多样性与复杂性。相反,他们都认为,人类的智力并不是某个单一的实体,而是多元化的,存在不同类型、不同性质的智力。例如,伽德纳认为,在每个个体身上都存在七到八种不同类型的智力,如言语智力、数理逻辑智力、视觉空间智力、身体动觉智力、音乐智力、人际智力、自省智力、自然观察智力;不同的个体擅长不同类型的智力,每个个体都有自己的智力强项或弱项;不能简单地用某个单一的、统一的标准

去衡量学生智力的高低。“问题不在于一个人有多聪明，而是他怎样聪明、在哪个方面聪明。”另一位心理学家斯腾伯格也持有类似的看法。他认为，智力是个体成功地适应和塑造环境的一种能力，它包含三个基本侧面，即分析智力、实践智力和创造智力。其中，分析智力主要用来分析、评论与评判，实践智力涉及到知识的使用与运用，主要用于方案或计划的执行与实施，创造智力主要用于规划、设计、发明与创造。

如果我们用上述这种智力观去看待和衡量学生智力发展的话，不难发现，传统学校教育所奉行、所采纳的智力观，基本上是单一类型的智力观，这不仅抹杀了学生智力存在“水平”差异的事实，而且还抹杀了“类型”与“结构”差异的事实。正如伽德纳和斯腾伯格所看到的那样，当前的学校教育实践，不论是教学系统还是评价系统，均倾向于不断强化和鼓励学生的言语智力和数理逻辑智力（或学生的分析智力），其它类型的智力往往受到忽视。最明显、最有力的证据就是，在传统的课堂上，教师总是站在教室前面演讲，在黑板上板书，就一些事先已经安排好的阅读材料向学生提问，要求学生完成指定的书面练习或作业题；相应地，学生的学习方式就是坐在教室里看书、听讲、做笔记，以便消化和巩固教师和书本告诉给他们的信息内容，并在测验中反馈出来。在这样的背景下，教师既不关心学生的智力在真实生活中的运用，也不关心学生对新知识的探索。这一点不仅反映在教学上，也反映在各种类型的教育测验与考试中。通常我们的教育测验只能反映出学生对静态的、死的书本知识的掌握程度，不能反映学生灵活运用知识解决实际问题的能力，也不能反映学生发现新问题、提出新看法、解决新疑难的创造能力。

2. 危害及后果

教育评价偏重低层次认知能力或单一类型智力的评估，将导致三个方面的不良后果。

其一，不利于高层次创新人才的培养。根据布卢姆的认知目标分类学思想，对不同层次、不同类型的认知目标需要采用不同的教学方法。例如，若要帮助学生达到和实现高层次的认知目标，课堂上教师就不能仅仅采用像“讲授”这样单向交流的教学方法，而必须采用师生双向互动交流或者鼓励学生自主探究这样的教学方法。根据教学对学生思维挑战性程度的不同，人们通常把教学分为三种水平，即记忆水平、说明

理解性水平和自主探究水平。仅仅诉诸“讲授”的教学，充其量只能使教学在记忆水平或说明理解水平上发生，很少或不太可能使教学在自主探究水平上发生。而如果我们的教育评价仅仅注重对低层次认知目标的达成进行评价，则通常最有可能发生的教学方式就是单纯的“讲授”，单纯的“讲授”对培养学生的自主学习能力、探究能力、创造能力都极为不利，而这些品质恰恰是高层次创新人才所必需的。

其二，不利于学生的成材与成功。(1) 按伽德纳、斯腾伯格等人的多元智力理论，天才有许多种，不同的学生智力各不相同。如果我们的学校只推崇某个单一类型的智力，用一把尺子、一个标准去衡量智力各不相同的学生，势必会低估相当一部分学生的真实智力，并给他们烙上与其实际情况不符的“愚笨”标记，导致这些学生内在的优势智力得不到承认、开发与利用，他们的才能很可能被狭隘的、单一的智力概念所“掩盖”或“埋没”，这无论是对学生个体还是对整个社会都是一种损失与浪费。(2) 当我们试图用一把尺子或一个标准去衡量智力各不相同的学生时，我们的教学往往也是单一规格的教学。正如条条大路通罗马，当我们承认不同的学生有着不同类型的聪明，不同的学生有着不同的通往成功的道路时，为了使每个学生都能成材，或者让每个学生都能按自己的方式取得各自的成功，就必须为学生提供多元化的、有选择的学习任务或学习机会。而要做到这一点，教育评价的对象与方式都必须是多元化的，而不能是单一的。(3) 伽德纳明确指出，在实际生活中，任何一项活动的有效完成，都有赖于多种智力的联合运用。例如，一个棋艺高明的人下棋时，眼睛盯着棋盘，脑中却在盘算下一步该如何走，他必须同时运用视觉空间智力与数理逻辑智力。即使是像做一顿饭这样简单的事也得用上好几种智力：看食谱需要语言智力，计算食品和调料的用量需要数理逻辑智力，切菜切肉、煎炒炸煮等得靠肢体动觉智力，而设计一道适合各人口味的菜单又得具备人际关系的智力。斯腾伯格也说，现实生活中，真正的成功往往取决于分析的、实践的与创造的智力这三种智力成份的联合、平衡使用：只有创造智力的人，也许可以产生革新思想，却意识不到自己创新的好坏或不能推广这些创新思想；只具有分析智力的人也许能透彻地批判他人的思想，自己却不能产生创新思想；只

有实践智力的人或许是个优秀的推销员,却无法将那些没有价值或价值很小的思想或产品同那些真正有价值的思想或产品区分开来。因此,为了让学生能在社会上、在生活中取得真正的成功,学校教育就必须同时关注和强调上述三种类型智力全面、均衡发展,而不能只钟情于其中某一种。

其三,不利于对教师的教学进行客观公正的评价。有些教师在课堂上很注重通过课堂提问、作业布置等教学环节,培养学生高水平的思维技能,但如果我们的教育测验和考试只偏重对学生达成低认知水平教学目标的情况进行考察,那么,这种教育评价很可能误导我们对教师的评价,即真正的好教师、优秀教师反而受到较低的评价,而那些平庸的、中规中矩的教师反而受到较高的评价。据说在课新改中,那些按新课改的要求去上课的教师,其学生反而考不过那些按老方法(教师讲学生听、反复不断地做各种题目、频繁地测试)教出来的学生,这很可能是一种假象:如果我们的教育评价系统过于偏重低认知水平教学目标的考察,那么出现上述这种现象就没有什么好奇怪的了。

3. 对策与建议

要改变上述倾向,就必须对教师的课堂提问、作业布置、教育测验题目的编制乃至教材中课后练习的设计等进行联动性的改革。

(1) 在课堂提问和布置作业的教学环节上,教师可以依据布卢姆的教育认知目标的分类框架,注意把涉及不同认知层次的提问或作业结合起来。比如,教师可以有意识地将教材内容与现实生活联系起来,问一些涉及知识使用的应用性问题,或者问一些没有固定答案、较为复杂的开放性问题;或者给学生布置一些开放性的主题作业,要求学生进一步去搜集资料、进行调查研究和深入思考才能回答的开放性作业(又称长期性作业)。

本着“为多元智力而教”的理念,在作业布置的教学环节,教师可以尝试在自己的学科教学中,给学生布置涉及各种类型智力或思维运用的作业任务。例如,语文课教师可以要求学生运用想象力,给《呼啸山庄》安排一个现代结局;或者要求学生读完《斧头》这本书后,考虑如何运用从作品中学到的技巧帮助自己在野外求生;或者让学生分析一下《夏洛特的网》这本书的作者,为什么要把书名取为“夏洛特的

网”。上述三个作业任务要求的智力类型各不相同,它们分别要求学生运用创造智力、实践智力和分析智力,这样的作业布置既能让学生展示其智力或思维的长处,又能让学生智力或思维的“短板”得到弥补,两者的结合即是《学记》中所讲的“长善救失”。

(2) 在编制有关课程的教育测验时,试题的编制者必须注意和做到,在一张试卷中用来考察高、中、低三种不同认知层次教学目标的试题都要占有一定的比例。尤其是要在各门学科的测验考试中,有意识地加入少量用以考察学生批判性思维、创造性思维的开放性试题,不要总是拿那些封闭的、有着固定答案的书本知识来考学生。

长期以来,在教育测验题目的编制上,一直有这样一种思维定势,即凡是作为教育测验与考试内容的东西都必须具有客观性与确定性,只有那些具有唯一正确答案的问题,才能纳入中小学的教育测验与考试中。我们认为,这种思维定势必须打破:要真正检测学生在学校里学到了什么、发展得怎样,既要用那些稳定的、客观的、具有确定性的知识去考他们,同时也要留出适当的空间,用那些在书本上找不到现成答案、学生必须综合运用所学知识、根据实际情况作出灵活应对的开放性试题去考他们。从某种意义上讲,采用后一种考试方式更能考出学生的实际发展水平。我们没有必要去担心:用开放的、答案不确定的问题去考学生,会导致考试效度与信度的丧失。尽管采用开放性的考题,学生对问题的解答存在某种程度上的不确定性,但我们依然可以制订出相对合理的“评价标准”,对学生反应的高下、优劣作出评判。比如,从学生的反应中我们可以看出,一个学生思考的视野比另外一个更广阔;论证更清晰、更严密、更有说服力;提出的观点更独特、更新颖、更有价值。其次,面对答案不确定的开放性问题,学生可以站在不同的立场、从多个不同的视角展开思考,得出不同的结论。承认这一点并不意味着学生得出的任何结论、做出的任何反应都一样“正确”、一样“合理”。答案的“多元化”与“随意化”是两个完全不同的概念,承认一个问题的答案是不固定的、多元化的,并不意味着对这个问题的回答可以是任意的。比如,在语文课中,教师可以鼓励学生对文学作品的意义做多元化的解读,但这种多元解读不是随意的、胡乱的解读,任何一种解读都必须有它的根据,都必须联系“文本”产生的特定社会历史条件

以及作者个人特定的生活经历来进行解读。也就是说,即便是在人文学科的学习中,学生对“文本”意义的自由解读也是有边界的。这意味着,在人文学科的教学及评价中,我们根本不用担心以开放性问题作为考试内容会影响到评分的客观性,我们也可以采取许多技术手段来避免这一点。比如,要求出题者事先拟好合理的评价标准,列出各种可能的答案,采用多人评分制以避免评分的主观性,等等。

(3) 改造传统的教材编写方式,发挥教材对教学和评价的导向作用。教材编写得好的话,也可以起到某种教育评价的作用。例如,台湾的语文教材编写体制一般分“课前导学系统”(含学习重点、课前预习、题解、作者简介等)、“课文及注释”和“课后练习设计”(含问题讨论、应用联系、课外学习指引等)。其中,“课前导学系统”中的“课前预习”和“课后练习设计”中的“问题讨论”、“应用联系”均以问题的形式呈现,对学生而言是一种练习、一种作业,对教师而言则是一种评价学生理解状况的工具。特别值得称道的是,“问题与讨论”栏目中设计的这些问题几乎每一个都是开放性的、没有标准答案的问题,至多只是提供了一个思考的方向或思路,它们对培养学生的复杂性思维、批判性思维均有很好的促进作用。^[5]

四、有的东西不值得考、没必要考或者超出了学科评价范围,却在现行的教育考试与评价中反复出现,甚至受到重视

在各门学科的教学,弄清哪些东西该考、哪些东西不该考,一直是个老大难的问题。由于这个问题没搞清楚,因此,现实中一些不该考的、意义或价值不大的东西常见于各级各类考试命题中,尤其是在语文学科中,这种现象体现得最为明显。比如,要求学生对本文内容作琐碎的微言大义式的分析;要求学生分析一篇短文可以分为几层、几段,每一层、每一段的意思是什么;要求学生分析某个字的偏旁部首、声母韵母是什么、有多少个义项,分析某个词的词性、某个句子包含的修辞手法或语法结构是什么。还有,拿一些琐碎无用的、根本无需记忆的、需要时查查工具书就知道的死知识(如“作者的籍贯”、“作品发表的年代”等)去考学生;更有甚者,出一些莫名其妙的、刁钻古怪的题目去考学生。这样做的后果是,既消磨了学生学习语文的内在兴趣,同时又浪费

了学生宝贵的学习时间。

其次,各门学科究竟应当考什么还必须考虑各门学科自身的性质或学科特点,尽管新课改强调要加强学科的交叉、渗透与综合,但这种综合是有限度的,毕竟各门学科有它自身的学科特点,学科与学科之间是有逻辑界限的,忽视这种界限,一门学科就不能称之为该学科。但在现实的教育评估与测验中,有些人为了追求学科的交叉与综合,往往置学科特点、学科界限于不顾。

要改变上述评价偏差状况,我觉得可以从两个方面去努力:首先,教育测验的编制者、教育评价者包括学校各门学科的任课教师都必须通过学习,深入了解和领会各门学科本身的性质与特点、逻辑界限及相互关联,在此基础上,深刻理解和把握各门学科内在的、独特的存在价值以及它所蕴含的教育价值。只有搞清楚这个问题,我们才知道各门学科应当教什么、考什么。不少教师常常会把“教材内容”混同于“课程内容”,以为教材中的课文讲什么,教师就应当教什么、考什么。以语文课的教学及考试为例。有学者指出,把“对美满爱情的追求”作为“牛郎织女”这篇神话故事的主题是可以的,但如果把它作为一堂语文课的主题就很离谱,因为“追求美满的爱情”显然不是语文课的主要教学任务,这个任务品德课、公民课、人生辅导课来承担更合适一些。导致这种现象的主要原因是,许多语文教师误把“教材内容”(或课文内容)当成“课程内容”;只有“课程内容”才是语文课教师要“教”、要“考”的对象。所谓语文课的“课程内容”,即是语文课标中规定的要教给学生的语文知识、要培养的语文智能(如听、说、读、写能力,尤其是对语言的理解力、使用与表达力);使学生学到一种平实的、合逻辑的、清晰的“有用语言”,才是语文课最重要、最基本的“课程内容”,至于通过语文课对学生进行情感和审美方面的教育,都是次要的、处于第二位的事情,而且,它们最好以潜移默化的、附带的形式来进行。^[6]

其次,在弄清楚各门学科本身的性质、内在特点及其教育价值的基础上,还要深入到各门学科独有的知识体系中去,找出那些不会过时的、对学生今后可持续发展真正有价值的东西,这便是美国学者布鲁纳所讲的各门学科的“基本结构”。按布鲁纳在《教育过程》中的观点,教育过程的核心在于教师向学生

教授各门学科的基本结构(即基本概念、基本原理和基本方法),因为只有掌握了各门学科的基本结构,才能促进知识的普遍性迁移。从事课程设计与开发的人都知道,衡量课程内容是否重要的一个基本标准就是看它是不是各门学科中最基本的成分;是不是应用性和迁移力最大的成分;是不是属于探究方法和探究精神的成分。“最基本的成分,也就是最核心的成分,缺乏了此一成分,知识不成其为知识,文化不成其为文化,它们都丧失了本质,学习也缺乏意义。应用性和迁移力最大的成分,可以促使学习趋于省力,同时内容间的关系也易于阐明,学习效果较

佳。探究方法和探究精神是知识和文化进步的基石,有了它,事实、概念、原理原则的发现和建立,才有可能。”^[4]课程内容的选择是如此,教什么、考什么的取向亦是如此。

本文系张民选教授主持的上海市教委重点学科“比较教育学”(J50404)的研究成果,亦属于上海市教委科研创新项目“课程设计与开发中基本理论问题的哲学思考”(CW0936)的部分成果。

(未完待续)

(责任编辑 林岚)

参考文献

- [1]张礼永,郭军.共和国教育60年:筚路蓝缕(1949-1966)[M].广州:广东教育出版社,2009.222-223.
- [2][英]A.S.尼尔.夏山学校[M].周德译.北京:京华出版社,2002.
- [3]黄全愈.“高考”在美国[M].北京:北京大学出版社,2003.54-58,92-93.
- [4]沈兰,郑润洲.变革的见证:顾泠沅与青浦教学实验30年[M].上海:上海教育出版社,2008.121-124.
- [5]王丽.语文丢失了什么[N].中国青年报,2006-06-21.
- [6]王晓春.语文课如何是好[M].北京:中国轻工业出版社,2010.
- [7]黄政杰.课程设计[M].台北:东华书局,2008.267.

Can Our Present Education Evaluation Promote Students' Development?

Xia Zhengjiang

(School of Education, Shanghai Normal University, Shanghai 200234)

Abstract: The education evaluation has a strong function of orientation, that is, we have what kind of education evaluation, and we will have what kind of education practice and what kind of students' development. Based on this standpoint, this paper conducts a systematic reflection on three main defects existing in our primary and middle school education evaluation system that hinder students' development and puts forward the corresponding suggestions for improvement. The first problem is that the dislocation between what should be evaluated and what shouldn't be evaluated has caused the distortion of education process. The second problem is that the impropriety of evaluation standards and methods leads to the distortion and invalidation of the result of evaluation. The third problem is that the unreasonableness of interpreting and using the result of evaluation leads to the narrowness and alienation of evaluation function.

Keywords: education evaluation, students' development