

# 英国 APU 的科学探究评价案例简析\*

文 罗国忠

APU (the Assessment of Performance Unit) 是英国教育部下设的成就评估机构, 专门为监测学生科学成就而设立。在英国, 中小学开展科学探究活动有悠久历史, 但教育界对它的实际效果却认识不一。为此, APU 在 1981~1984 年组织了全英国范围科学成就评价, 其中就包含了科学探究评价 (Donnelly, 1987; Murphy, 1988)。①②它采用纸笔测验、工作单和现场观察三种方式, 评价中学生的科学探究能力, 该评价在全世界范围产生的影响深远, 至今仍被很多研究者关注。本文结合案例, 从评价目标、探究任务和评分细则三方面, 介绍我国教师相对陌生的工作单和现场观察的具体运用, 期望能对我国的科学探究评价有所启示。

## 一、评价目标

APU 在表 1 的评价框架中规定了科学探究的评价目标和评价方式。该框架把评价目标分为若干主类, 每个主类又分为若干子类, 并规定了每个主类的评价方式。

表 1 APU 的科学探究评价框架

主类	子类	评价方式
1. 使用图表和符号描绘	从图表中读取数据 用图表描绘数据	纸笔测验
2. 使用材料和测量工具	使用测量工具 估计数量 按照说明进行操作	群体实践评价 (工作单)
3. 观察	进行观察	群体实践评价 (工作单)
4. 解释和应用	解释给出的信息 应用: 生物概念、物理概念、化学概念	纸笔测验
5. 设计探究	设计部分探究 设计整个探究	纸笔测验
6. 做探究	做整个探究	个体实践评价 (现场观察)

从表 1 看到, APU 采用纸笔测验、工作单和现场观察来评价学生的科学探究能力。所谓工作单, 就是让学生动手探究, 同时把探究的过程和结果记录在一张单子上, 然后教师对单子的内容进行延时评分。由于一个教师可以监考一个群体的学生, 因此这种方式也叫群体实践评价。所谓现场观察, 就是让学生动手探究, 评分者进行观察, 辅以访谈, 即时评分。一般采用两个教师观察一个学生, 或一个教师观察一两个学生, 因此这种方式也叫个体实践评价。显然, 现场观察的人力和时间成本比工作单高得多。

从表 1 还看到, APU 的纸笔测验和工作单分工

明确, 各司其职, 两者的评价目标没有交叉。纸笔测验评价应用科学概念、从图表中读取数据、用图表描绘数据、设计方案等心智性能力; 工作单评价使用测量工具、估计数量、按照说明进行操作、进行观察等实践性能力; 现场观察则评价做完整探究的能力, 包括实践性能力和心智性能力。APU 这样做的潜在假设是: 纸笔测验能有效地评价某些心智性能力, 而工作单能有效评价实践性能力。因此, APU 的工作单侧重于评价纸笔测验不能评价的操作、观察等实践性能力, 它所用的任务往往不是完整的、复杂的探究任务, 而是完整探究任务的一些片段或比较简单的任务。

## 二、探究任务和评分细则

### 1. 工作单

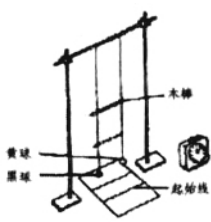
表 2 是一个基于工作单的“摆动”任务。

表 2 工作单的任务

**摆动**

一、探究问题: 不同质量的物体对摇摆有什么影响?

二、材料和器材: 两架铁架台、一根金属棒、一根两端有洞的木棒、两个小球 (一个黑、一个黄)、一张标着起始线的卡片、一个计时器、一根铅笔、粉笔、磁带。这些材料装配成如图所示。



三、探究任务:

(一) A 部分

1. 一只手让黑球处于竖直静止状态, 另一只手把黄球拉至卡片的起始线。然后, 两只手同时释放小球。
2. 观察两只小球至少一分钟。
3. 你发现黄球摆动后发生了什么现象?

(二) B 部分

4. 把木棒拉下来, 使之停在标有黑色的位置上。
5. 又重复原来的动作, 然后观察两只小球至少一分钟。
6. 这次你发现了什么不同的地方?

\* 本文系广西哲学社会科学“十一五”规划课题“科学探究能力的表现性评价”(编号 06JFYX005) 和广西教育科学“十一五”规划课题“初中生科学探究能力评价的研究”(编号 2006-B69) 的阶段成果。

这个探究任务相对简单,学生只需按照工作单上的程序进行操作、观察,进行简单的比较推断即可,探究性较差。这与 APU 工作单的评价目标是一致的,主要评价学生做简单操作、观察和推断结论的能力。

表 3 为“摆动”任务的评分标准,它给出正确操作所应该观察到的若干现象,每个现象为一个评分点,每个评分点分值 1 分。

表 3 “摆动”的评分标准

观察到的现象		分数
A 部分	黑球开始运动	1
	黑球逐渐摆得变远	1
	黄球摆动逐渐变慢	1
	黄球摆动接近停止	1
	木棒开始运动	1
	周期自己重复	1
	黑球摆动的距离与黄球摆动的距离差不多	1
A 部分总分: 5 分		
B 部分	运动的变化更快了	1
	侧移运动更小了	1
	周期自己重复更快了	1
B 部分总分: 2 分		
总分: 7 分		

在这个任务中,A 部分有 7 个评分点,满分 5 分。也就是说,只要有 5 个评分点得分即可得满分。一样的,B 部分有 3 个评分点,满分 2 分。这样,该任务满分 7 分。

2.现场观察

表 4 是一个基于现场观察的探究任务“弹簧上下振动的周期与弹簧的长度、粗细的关系”。

表 4 现场观察的探究任务

<p><b>弹簧的振动</b></p> <p>弹簧在生活中的应用非常广泛,小到圆珠笔大到汽车,都有弹簧在其中。如果把钩码挂在弹簧下面,然后轻拉一下钩码,松手后钩码会上下周期性地振动。相同的钩码挂在不同的弹簧下面,钩码的振动周期会不一样。</p> <p>你的任务就是要弄清哪些因素造成弹簧振动快慢不一样?是弹簧的长度还是弹簧的粗细?在你的桌面上有如图所示的器材(图略),你可以任意挑选你需要的器材来做这个探究。</p> <p>在探究时,把你的探究结果清楚地记录下来,以便别人也能理解你所发现的。</p>
--

这是一个开放的探究任务,学生进行自主探究,难度要大于上一任务。表 5 为教师所用的观察核查表,由三大主类构成。第一大主类“识别关键变量”,包括识别自变量、因变量。在该探究中,如果学生依次更换不同长度或不同粗细的弹簧,表明学生已经

识别了该探究的自变量;如果用计时器测量振动时间,也表明学生已经识别了因变量。但是,学生识别了关键变量,不一定能有效处理这些关键变量。因此,还需要观察是否满足第二大主类“有效处理变量”。例如,某学生依次用了几个长度不同的弹簧,但这些弹簧的粗细有些一样,有些不一样。显然,该学生已经知道要改变弹簧的长度,但他没有保持弹簧粗细不变,意味着该学生并没有有效处理长度这个自变量;类似地,如果某学生只测量弹簧振动的一个周期,或测量到弹簧振动静止,意味着该学生已经知道要测量振动时间,但没有有效地测出振动周期。即使学生能识别,并能有效处理变量,也并不意味着结果是有效的,因此还需要观察是否满足第三大主类“结果的有效性”。例如,某学生用了两个粗细相同、长度相差很小的弹簧,满足了第一、二项,但由于长度相差小,再加上误差,振动周期有可能没有差别,该学生有可能得出错误结论。

表 5 观察核查表

主类	子类	学生 A	学生 B	学生 C	学生 D
识别关键变量	改变弹簧的长度				
	改变弹簧的粗细				
	测量振动周期				
有效处理变量	使用不同长度的弹簧时,控制弹簧的粗细不变				
	使用不同粗细的弹簧时,控制弹簧的长度不变				
	测量几个振动周期的时间				
结果的有效性	长度和粗细的范围、改变的次数				
	是否静止释放弹簧				
分数		2	1	3	0

但是,纯粹的观察有可能会造成误判,因为观察者只能观察到学生的外显行为,而外显行为有可能与内隐思维不一致。例如,某个学生在实际操作中控制了相关变量,但他究竟是有意识地控制还是无意识地控制相关变量,纯粹的观察是难以作出可靠判断的。如果该学生无意中控制了相关变量,观察者也判对,这样的误判无疑对其他学生是不公平的。另外,某学生在第一、二次实验中没有控制变量,但在第一、三次中控制了变量,但他究竟是有意识控制还是无意识控制,纯粹的观察也不足以作出可靠的判断。因此,在观察过程中,如有不清楚之处,可询问学生,但要尽量少,而且也不能提示;探究结束后用表 6 的问题

进行访谈,通过这样的询问、访谈可确保观察的可靠性。

表 6 访谈问题

- |  |
|--|
| 1.你得出了什么结论?<br>2.你用了哪些弹簧来探究弹簧长度的影响?<br>3.你用了哪些弹簧来探究弹簧粗细的影响?<br>4.你是怎样测量或判断振动快慢的?<br>5.对于每根弹簧,你是怎样确保公平实验的?<br>6.如果再次探究,你会采取哪些不同的做法? |
|--|

对于现场观察的观察核查表,有几种评分方式。第一种,每个核查点赋予相同的分值(与上一案例类似);第二种,根据核查点的重要程度,赋予不同的分值;第三种,标准参照。以标准参照为例,说明其评分方式。首先,为了得到第1分,不管学生能否设计出适当的方法来改变这些自变量、测量这些因变量、控制某些相关变量,但他必须知道要改变哪些自变量、测量哪些因变量。然后,为了得到第2分,学生必须有效处理这些变量。最后,为了得到第3分,学生不仅要识别出关键变量、有效处理这些变量,还必须保证探究结果可以得出有效的、可靠的结论。表4是4个学生在“弹簧振动”中的表现及得分。这个基于标准参照的评分方式可以适用于所有的探究任务。

### 三、结语

对于整体性探究能力的评价,APU采用了公认最有效的现场观察,是世界上为数极少采用现场观察的大规模评价项目之一;尽管纸笔测验很难评价整体性探究能力,但能有效评价某些心智性探究能力,而且成本低,因此也被APU采用;对于实践性探究能力的评价,APU采用了工作单。然而,APU使用的工作单受到一些质疑,主要是因为它评价的往往是低水平的操作和观察技能,而不是高水平的探究能力(Buchanetal., 1992; Matthewsetal., 2005)。<sup>③④</sup>但

有研究表明,如果学生动手做了探究,那么工作单能可靠、有效地记录其完整的探究过程和结果,即工作单是学生动手探究的可靠指示器(Woolnoughtetal., 1990; 罗国忠, 2007)。<sup>⑤⑥</sup>

毋庸置疑,这对我国如何有效评价科学探究能力有积极的启示意义。高成本的现场观察对我国大规模评价没有现实意义,因而只好采用低成本的纸笔测验来评价某些孤立的心智性探究能力,但这样一来便窄化了探究能力。由于工作单的成本和有效性介于纸笔测验和现场观察之间,在我国现实条件下有较强的可操作性,因此我们可以尝试把工作单和纸笔测验结合起来,改变目前评价方式单一化的现状,并进一步研究降低它的成本、提高它的有效性。

#### 注释:

① Donnelly JF. Fifteen-year-old pupils' variable handling performance in the context of scientific investigations. *Research in Science and Technological Education*, 1987, 5(2): 135-147.

② Murphy P. Insights into pupils' responses to practical investigations from the APU. *Physics Education*, 1988, 23: 330-336.

③ Buchan AS, Jenkins E W. The internal assessment of practical skills in science in England and Wales, 1960-1991: some issues in historical perspective. *International Journal of Science Education*, 1992, 14(4): 367-380.

④ Matthews P S C, McKenna P J. Assessment of practical work in Ireland: A critique. *International Journal of Science Education*, 2005, 27(10): 1211-1224.

⑤ Woolnough B E, Toh Ka. Alternative approaches to assessment of practical work in science. *School Science Review*, 1990, 71(256): 127-130.

⑥ 罗国忠. 初中生科学探究能力评价方式的比较研究[M]. 南宁: 广西教育出版社, 2007. 201-218.

[罗国忠 广西教育学院教研部 530023]

(上接第26页)

是“设计者”、“督导者”和“参与者”。

(3) 重视资料积累:不只是课题中文献检索、调查访谈等形成的资料,也包括每次研究型心理课程的总体设计及最后的研究报告、课程评价、反思等材料,能妥善保存,系统整理,对以后的各届学生和新的课程,都起到信息资源共享和促进发展的作用。

(4) 关于评价方法:我们已完成的研究课题都是以小组成员共同操作完成的,研究成果是集体智慧的结晶。由于课题成员在研究过程中的贡献与成长是有差异的,所以要将评价个人与评价小组相结合;评价过程与评价结果相结合;评

价态度与评价能力相结合。

由于历史和地域的原因,七宝中学一直致力于在国内心理健康教育领域做一些开创性、示范性的工作。研究型心理课程作为以活动型课程、拓展型课程和研究型课程为一体的整体课程设计中的一个新层面,是对心理课程建设的进一步完善,体现了心理健康教育课程基础性、整体性和选择性的结构特点。我们希望探索一条真正适合我国国情的路子,同时延伸出适合城市、乡村、东部、西部的不同路子,以期通过研究型心理课程,普遍提高学生的研究能力和创新能力,也提高他们的道德和心理素质。

[杨敏毅 上海市七宝中学 201101]