



## 国内学生学业成就评价研究综述

钟媚

华南师范大学课程与教学系

广州 510631

电邮：[zhmei80@163.com](mailto:zhmei80@163.com)

收稿日期：二零零八年三月二日(于六月三十日再修定)

---

### 内容

- [摘要](#)
  - [《教育部初中毕业学业考试评价理科科目组》](#)
  - [《新课程背景下高中学生学业评价》课题组](#)
  - [《促进教师发展和学生成长评价研究》课题组](#)
  - [结语](#)
  - [参考文献](#)
- 

### 摘要

弱化评价的管理性功能，突出评价的教育性功能是新一轮国家课程改革中一个亟待解决的问题。虽然从理论层面上学者们对学生学业成就评价研究进行了大量的探讨，但从实践层面上对这一方面的研究还非常薄弱，直接关注科学学科学业评价研究的更加少。本文选取了三个比较有代表性的学业成就评价研究课题组：（1）《教育部初中毕业学业考试评价理科科目组》，主要关注初中理科试题的命制工作；（2）《新课程背景下高中学生学业评价》课题组，主要关注过程性评价和开放性试题的命制工作；（3）《促进教师发展和学生成长评价研究》课题组，主要关注发展性评价和档案袋评估的研究工作，对他们的具体工作进行简要综述。

### 《教育部初中毕业学业考试评价理科科目组》

《教育部初中毕业学业考试评价理科科目组》于2003年开始对初中理科会考试题的命制和开发进行了广泛的实践探索。虽然该课题组在总体理论框架上没有明确的导向，但在对科学概念知识、科学探究能力和情感态度价值观方面的考查上提出了很有建设性意见的命题原则和要求（全国初中毕业升学考试综合理科评价课题组，2004；教育部



初中毕业学业考试评价理科科目组，2005)。另外，在命题的总体形式上，无论是对科学概念知识、科学探究能力还是情感态度价值观的考查，该课题组的关注点都集中在对情境性试题的研究工作上，即考察学生在具体情境中运用所学知识与技能，分析和解决问题的能力，这是很值得理科考试命题工作者借鉴的地方。

### (一) 科学概念知识方面

在科学概念知识方面的考察上，该课题组认为命题的原则是：要密切联系学生的学习和生活实际，考查学生对科学知识的理解以及应用科学知识解释现象、分析和解决问题的能力。比如对溶液概念的考察，传统的做法是直接考定义，现在的做法是考概念的运用。该课题组认为，对于命题基础知识的命题要避免两种倾向：一是只测试一堆高度抽象、高度概括的知识，完全脱离生活实际；另一种则相反，测验只是简单地重复一下学生已有的生活经验，没有超越经验上升为科学（资料来源：教育部初中毕业学业考试评价理科科目组，2005，页 30-34）。

例题 1： 以下是溶液的一些说法： ① 无色 ② 混合物 ③ 化合物 ④ 均一 ⑤ 稳定 ⑥ 各部分性质相同，其中正确的是（ ）

A. ① ③ ④ ⑤      B. ② ④ ⑤ ⑥      C. ① ② ④ ⑤      D. ① ③ ⑤ ⑥

答案： B

例题 2（上海理化，2003 年）： 将白醋、食用油、食盐、酒精分别放入水中，不能形成溶液的是（ ）

A. 白醋      B. 食用油      C. 食盐      D. 酒精

答案： B

例题 3： 下列物质中，没有颜色、没有气味、没有味道的液体是（ ）

A. 啤酒      B. 食盐水      C. 纯净水      D. 食醋

答案： B

分析： 和例题 1 一样，例题 2 同样考察溶液的概念，但却不像例题 1 那样对溶液概念的定义直接设问，而是通过设置一个实际的情境，让学生回答将生活中常见常用的白醋、食盐、食用油、酒精分别放入水中能否形成溶液。例题 3 则停留在常识水平，不需要教学和考试也可以回答。

### (二) 科学探究能力方面

该课题组认为，对科学探究的考查应该从实际问题出发，重点关注学生的探究能力。应该避免将科学探究分解为生搬硬套的知识和按部就班的程序进行考查。在实际教学中，学生应有较完整、自然地体验科学探究过程的机会，但在大规模考试的情景中，要考查一个完整的探究过程往往是困难的，选择探究过程的部分内容，对学生探究能力的某一方面进行考查更现实可行。也就是说，为了保证考查的有效性和可操作性，一道试题不必贪多求全，可以只集中考察探究能力的一个或两个方面。如考查学生提出猜想的能力、设计和实施实验的能力、表达和交流的能力等。



## 1. 考查针对科学性问题的猜想的能力

例题 1：在煮饺子的过程中，可观察到如下现象：饺子刚放入时，沉在水底，煮了一段时间后，就浮起来。饺子为什么会浮起来？请提出你的猜想并简要说明。（资料来源：全国初中毕业升学考试综合理科评价课题组，2004，页 13）

我想原因可能是：\_\_\_\_\_

我的根据是：\_\_\_\_\_

还有可能的原因我认为是：\_\_\_\_\_

我的根据是：\_\_\_\_\_

.....

可能的答案是：

- (1) 沸水上下翻滚带起饺子。
- (2) 水沸腾后饺子表面吸附了一些气泡，托起饺子。
- (3) 煮熟后的饺子密度小，因此会浮起来。
- (4) 饺子由于温度升高而体积变大，所受浮力也增大，且大于重力，因此浮起来。
- (5) 饺子内的空气受热膨胀，使饺子体积变大，所受浮力也增大，并使浮力大于重力（尽管煮过的饺子因吸水会变重，但浮力增加显然更大），因此浮起来。
- (6) 随着饺子内的温度升高，水蒸气增加，导致饺子体积变大，所受浮力增大，并使浮力大于重力（尽管煮过的饺子因吸水会变重，但浮力增加显然更大），因此浮起来。

评分标准：满分为 6 分，答（1）或（2）得 1 分；答（3）得 3 分，其中若只答饺子煮熟了就会浮起来，而未提到密度变小的，得 1 分；答（4）、（5）和（6）中的任意一个得 5 分，同时考虑饺子中的空气和水蒸气受热膨胀的得 6 分，其中若没有考虑煮过的饺子吸水变重而认为重力不变则扣 1 分；能综合考虑饺子本身和饺子内气体受热膨胀并判断气体受热膨胀可能是主要因素的得 6 分；提出其它合理猜想的酌情给分。

分析：一般来说，学生对煮饺子的过程并不陌生。对于饺子煮后会浮起来的现象，学生可根据经验和所学的知识，提出许多有一定根据的猜想。评分主要依据的是学生通过猜想所表现出的正确运用科学知识、方法和推理的能力。

## 2. 考查设计和实施科学实验的能力

例题 1：（新疆物理，2002 年）几个同学爬杆比赛时，要比较谁的功率大，分别需要哪些测量仪器和测出哪些物理量？（资料来源：罗星凯等，2004）

答案及评分标准：本题有多种答案，只要合理均得分。几种可能的答案如下：

- (1) 用体重计测体重，用秒表测时间，用长杆和卷尺测高度。
- (2) 用体重计测体重，用秒表测时间。（要求比赛的同学都爬到同一高度）
- (3) 用体重计测体重，用长杆和卷尺测高度。（要求比赛的同学在几条杆上同时开始，同时停止）



分析：要设计一个实验来测量人爬杆时的功率，首先就要建立一个科学模型，高度抽象、概括地描述人爬杆时的物理状态。有了正确的模型，爬杆人的体重、爬杆的高度和时间是三个必须考虑的因素。常规的方法是像答案(1)那样分别去测量三个量，而用控制变量的思想（保持同样的高度或用同样的时间），就可以在时间和高度这两个量中少测一个（作为试题，这样的设计应该得分最高）。因此，这样的设计性实验试题，虽然并不复杂，但却考查了学生建立科学模型和正确应用控制变量的方法解决问题的能力。

### 3. 考查解释和表达科学探究结果的能力

例题 1：（南通物理，2004 年）小明同学在做完“观察水的沸腾”实验后，又进一步研究了沸水自然冷却过程中温度随时间的变化情况，将实验数据记录在下表中（数据来源：教育部初中毕业学业考试评价理科科目组，2005，页 81-82）：

时间 $t/\text{min}$	0	5	10	15	25	35	45	55	65
温度 $t/^\circ\text{C}$	100	71	55	45	35	28	24	22	22

请在坐标图中通过描点画出上述温度随时间变化的曲线。

如果有一天你要喝一杯奶茶，可以有两种方法供你选择：① 先将滚烫的热茶冷却 5min，然后加一匙冷牛奶；② 先将一匙冷牛奶加进滚烫的热茶中，然后冷却 5min。你认为方法\_\_\_\_\_的冷却效果较好。

分析：一杯热水放在室温下，温度不断下降，是一个司空见惯的现象。通过探究，从这个司空见惯的现象中发现规律，并把这个规律巧妙地应用到日常生活中，可以让学生感受到科学探究的独特魅力。

### （三）情感态度价值观方面

该课题组认为，情感态度与价值观是新课程三维目标体系中的重要一维，其重要性和知识与技能、过程与方法至少是等同的。但在科学学科的教育活动中，不能把这 方面的东西当成一些条文背诵，而在考试中要求简单地复述这些条文恰恰会导致这样的教学。该课题组提倡通过在试题中渗透情感态度与价值观教育的方式实现考察 意图，比如渗透到具体生动的科学事例与活动中，设计一些开放性的试题等等。该课题组认为，是否完全可以通过考试的方式来检测这方面的内容，是值得商榷和进一步研究的问题。

例题 1：在今年的气象日，《海口晚报》刊登了一则气象资料：“由于空气污染的影响，地球上阳光越来越少。与二十世纪 50 年代相比，如今到达地表的太阳光平均减少了 10% 左右……”。请你依据光的传播知识分析，影响地表日照的原因可能是什么（数据来源：罗星凯等，2005）。

答案：空气中的污染物（尘埃）将太阳光向各个方向反射。



分析：学生做完这道题后，收获可能不仅仅是这道试题的正确答案。很多学生都知道大量排放烟尘会造成污染，但没有想到连光照也会影响。无形中也许学生就意识到防止大气污染，保护环境是多么的重要。

## 《新课程背景下高中学生学业评价》课题组

《新课程背景下高中学生学业评价》课题组于2003年开始对学生学业成绩评价进行理论研究和实践探索。该项目研究分为两个阶段，第一阶段是2003-2005年，主要就高中学业评价的基本理念和框架、模块终结性测验的定位和方案、过程性评价的理念和方法、开放性试题的命制等进行了探讨；第二阶段是2005-2008年，主要就高中必修课程学业评价进行探讨，主要考察必修模块的水平测试，在命题与考试方面开展实践研究。这里主要对该课题组在过程性评价的理念和实施、开放性试题的命制和开发两个方面进行探讨。

### （一）过程性评价的理论和实践研究

#### 1. 过程性评价的理念

该课题组认为，过程性评价不是对微观意义上的学习过程的评价，也不是只注重过程而不注重结果的评价，而是对课程实施意义上的学习动机、过程和效果的三位一体的评价（高凌飙，2004）。为此，相对于终结性的或形成性的测验方式，过程性评价更关注动态的学习表现，评价的内容主要包括以下三个方面（高凌飙，2004；钟媚，2005）。

##### （1）学习方式

新课程要求改变被动的、接受式的学习，提倡自主、合作和探究的学习方式，使学生的学习发生实质性的变化。自主学习、合作学习和探究性学习是对学习方式的形式化描述，那么，学习方式的本质是什么？

这里指的学习方式来源于英语“approaches to learning”，具体含义包括：当学生面对学习任务时，他们对学习任务的价值取向和兴趣，以及他们对学习本身的认识，导致他们对学习效果产生特定的期望，这种期望驱使学生采用特定的策略和方法完成学习的任务，从而产生与期望相匹配的结果。可见，学生学习前的预备状态（presage）、学习过程中思考和活动的方式（process）、学习的结果（product），是一个密切相关的系统。简言之，学习动机 + 学习策略 = 学习方式，学习方式是决定学习质量的重要一环，对学生的学学习结果产生重大的影响。根据所涉及的动机和策略的不同，比格斯(Biggs 1987)把学习的方式分为表层式、深层式和成就式等。新课程所倡导的是深层式的学习，即学习者的学习由内在动机或好奇心产生，为满足兴趣、探究意义而自发、主动学习。

##### （2）反映学生智力发展的过程性成果

终结性测验可以很好地考察经过一个学习阶段后学生对知识技能的理解和掌握，而对反映学生智力发展的过程性成果则无法很好地反映。这就成为过程性评价关注的内容之一，具体来说，包括：①学习策略和方法。在学习过程中，以过程性评价作为交流的平台，通过学生之间的相互观察和提醒，可以促进他们不断自我反思，从而提高学生的学



习策略和方法水平。②与学科学习紧密相关而难以通过终结性评价衡量的智力因素，如解决现实科学问题的能力、创新能力等，这是考试测验不能完全考出来的。③通过学科学习而达成的高层次学习目标，如正确的科学观等。

### (3) 与学习密切相关的非智力因素

学生的学习过程受多种因素的影响，其中非智力因素对学生的学习有十分重要的作用。过程性评价通过考察与学习密切相关的非智力因素，比如对学习任务的意向，学习的习惯，以及透过学习过程反映出来的情感和态度等，给学生以信息回馈，促进学生进行反思和总结，使其对自身建立更加全面和深入地认识，从而提高学习的效果和质量。

## 2.过程性评价的实施

如何进行过程性评价？评价的实质在于价值判断。从操作层面上看，评价的过程包括设计评价方案和工具，利用评价工具收集数据数据，对数据数据进行分析得出结论，对评价结论进行解释。因此，过程性评价的实施应该包括以下四个工作环节（钟媚，2005）：

### (1) 明确评价的内涵和标准

从评价的使用目的及结果来看，过程性评价属于课堂学习评价，一般由学校自主进行，评价的主体是教师和学生。在评价的过程中，教师虽然也作为评价者参与评价过程，但更多的是作为评价的监控者和组织者，所以评价的主体主要是学生，评价的责任也就相应地落到了学生的身上。因此，评价工作的第一步是使学生理解评价的内涵和功能，明确评价的内容和标准。过去，由于学生长期处于被评价者的地位，对于什么是评价，评价的目的和作用是什么，以及如何正确地评价自己和他人等的认识是比较薄弱的，导致在评价过程中不能正确地把握和对待评价的标准。如何使学生提高对过程性评价的认识，使评价的水平由完全的下意识或潜意识水平以及情感水平逐步上升到理智水平，应该成为评价过程中的一项重要内容。当然，教师作为评价的组织者和主体之一更应该积极地提高对评价的认识，在教学的过程中逐步引导学生学会评价。

### (2) 设计评价方案和工具

从评价的内涵中可以看出，过程性评价是基于人们对学习质量的认识提出来的，是一种较具体方法更为上位的评价方式，因此，不能将过程性评价与某种特定的评价方法甚至评价工具等同起来，应该根据学校和学生的实际情况，针对具体问题，选择合适的评价工具。该课题组开发了评价量表作为过程性评价的工具（见表 2-1）。

表 2-1：高中语文过程性学习评价量表（资料来源：钟媚，2005）

模块\_\_\_\_\_高\_\_级\_\_班 学生\_\_\_\_\_学号\_\_\_\_\_200\_\_年\_\_月

评价项目	评价内容	评价等级			
		自评		组评	
1. 情感、态度、价	①热爱祖国的语言文字，积极提高运用欣赏能力				
	②维护祖国的语言文字，抵制不良语言现象				



价值观	③关注生活中的语言现象，并能分析判断、取舍					
	④按质按量完成学习任务					
	⑤学习中表现得自信、勤奋、坚毅，追求卓越					
2. 过程与方法	①独立思考，善于发现问题，有自己独到的见解					
	②合作意识强，勇于表达，专注倾听，乐于交流					
	③注重体验美，创造美，陶冶性情，涵养心灵					
	④善于计划、比较、反思、总结、调整					
	⑤勤于整理、辨析、积累，以各种方式充实自己					
3. 学习成果	①阅读广泛，读高文化品味著作，读完整著作					
	②表达交流语言规范、清晰、诚恳、有感染力					
	③积累丰富、完整、准确，灵活运用					
	④书写工整认真，整洁美观					
	⑤作品丰富、完整，有创意；笔记有条理，清晰					
特长及成果展示列举	背诵的篇章					
	阅读的书籍					
	发表及其它					
个人小结						
小组评议						
自评等级	本人签名	组评等级	组长签名	教师总评	教师签名	



### 表格填写说明：

1.过程性学习评价的方式为等级制，分四个等级：A等（优秀）、B等（良好）、C等（合格）、D等（仍需努力）。

2.各“评价栏目”的等级确认办法：先给5个具体分项评定等级；再将等级转换为分值：A等为3，B等为2，C等为1，D等为0分；最后把5个分数相加，得12—15分的为A等，8—11分为B等，4—7分为C等，0—3分为D等。

上述高中语文过程性学习评价用表的优点在于：评价所指向的内容比较全面，涉及情感态度与价值观、过程与方法、学习成果等领域；评价的主体多元，有学生自评、小组评、教师评；评价指标体系的分解比较详细、科学、可行；各指标的权重比较清晰；量化评价的等级的计算比较清晰；评价含有成果展示等评定的证明材料，使得评价的结果比较可信。

### （3）解释和利用反映学习质量的结果

在学年或一个学习阶段结束之后，应如何报告、解释、利用过程性评价的结果？有的学校建议把“过程性评价的结果赋予权重，折算成一个分数，然后与终结性评价的结果相加，作为最终学习结果”。那么，过程性评价与终结性评价要不要整合？过程性评价属于个体内差异评价，评价的目的并不是对学生的下一个终结性的结论，而是促进学生的学习与发展；评价的内容主要不是学生最后达到的水平，而是学生的进步情况。教师和学生应该通过对收集资料的分析，让学生了解自己的进步和不足，在此基础上提出建议，使学生明确将来继续努力的方向。所以，简单地把过程性评价的情况折算为一个分数加到考试成绩里面去是没有意义的。过程性评价的结果与终结性评价的结果应该分别给予呈现报告。

### （4）反思和改进评价方案

整个评价方案实施结束以后，还需要对于评价方案进行再评价，也称“元评价”，即“在评价的过程中，为检讨评价方案，实施过程与结果，藉以总结成功的经验和纠正评价工作之不足，而对正在进行或已完成的评价进行价值判断。”元评价的关键是确定元评价的标准问题，即一份好的方案的标准是什么。我们可以以下述问题作为准则考察评价的方案：

①有效。这个方案是为了解决什么问题，是否最合理地解决了问题，在多大程度上达到了我们的目标？

②可行。方案的要求是否超出了学校现有的条件和能力？评价的工具是否繁琐，是否给教师和学生带来很大的负担？

③可信。评价的标准是否恰切？评价工具收集的信息是否准确？

通过对这三方面的考察，学校和教师可以找出评价方案的不足，改进评价的方式和工具，更好地完成下一步的评价工作。过程性评价是一个动态多元的，而不是一个僵化单一的评价方式，因此，即使在实施的过程中，学校和教师仍然可以根据变化的情况及时修正评价方案。



## (二) 高中必修课学业水平测试

新课程强调学生的模块学分由各高中学校自主授予,这一措施大大增加了高中学校的办学自主权和灵活性,有利于各高中学校根据自己的实际进行教学,全面落实素质教育方针,形成自己的风格和特色。然而,由于学校条件和办学理念上的差异,各高中学校的教学质量也有一定的差别。需要通过适当的检测,对各高中学校的教学质量进行监控,保证每一间高中学校的教学质量以及每一位高中毕业学生的学业都达到国家规定的基本要求。为此,教育部华南师范大学基础教育课程研究中心,从2006年开始,系统组织了高中语文、数学、英语、物理、化学、生物、政治、历史、地理九学科的必修课学业水平测试。为了避免过去考试命题过于倚重封闭性试题的弊端,该课题组对开放性试题的命制方法进行了有益的探索。

### 1. 开放性试题的命制理念

过去的学业测评主要采用封闭性试题,该课题组认为,适当采用开放性试题在很大程度上弥补了封闭性试题的种种不足,特别在考查学生思维的灵活性和广泛性,考查学生的实践能力和创新意识,以及情感、态度、价值观等方面有着封闭性试题所无法取代的优点。开放性试题更符合新课程发展性评价的理念,在新课程的模块学业评价中应提倡适量使用开放性试题(高凌飙,吴维宁,2004)。

该课题组认为,开放性在编制上需要注意两方面的问题:①试题是否有明确的检测目标和检测效度;②试题是否违背学科的基本规律。为了避免开放性试题评分主观随意性的弊端,该课题组集中研究了SOLO分类法(Structure of the Observed Learning Outcome)在命题中的运用。SOLO分类评价法的基本理念源于皮亚杰的认知发展阶段论。皮亚杰的认知发展阶段论指出,儿童在成长的过程中认知的发展是有阶段性的,不同的阶段之间的认知水平有质的区别。比格斯(Biggs 1982)和他的同事通过研究发现,人的认知不仅在总体上具有阶段性的特点,在对具体知识的认知过程中,也具有阶段性的特征。人在学习新知识过程中表现出来的思维阶段是可以观察到的,因此称为“可观察的学习成果结构”。也就是说,学生在具体知识的学习过程中,都要经历一个从量变到质变的过程,每发生一次跃变,学生在对于这一种知识的认知就进入更高一级的阶段,可以根据学生在回答问题时的表现来判断他所处的思维发展阶段,进而给予合理的评分。SOLO分类法将学生学习的结果由低到高分五个不同的层次,即前结构(prestuctural)、单点结构(unistructural)、多点结构(multistructural)、关联结构(relational)、拓展抽象结构(extended abstract)等。这五种结构的基本含义如下。

- (1)前结构:没有形成对问题的理解,回答问题逻辑混乱,或同义反复;
- (2)单点结构:回答问题时,只能联系单一事件,找到一个线索就立即跳到结论上去;
- (3)多点结构:回答问题时,能联系多个孤立事件,但未形成相关问题的知识网络;
- (4)关联结构:回答问题时,能够联想多个事件,并能将多个事件联系起来;
- (5)拓展抽象结构:回答问题时,能够进行抽象概括,结论具有开放性,使得问题本身的意义得到拓展。

SOLO的五种层次代表了学生对于某项具体知识的掌握水平,从学生对某个问题的回答中,教师可以对照上述标准就学生对该项知识内容的掌握情况做出判断。因此,这种评



价方式可以帮助教师进行教学诊断，同时，也可以向学生提供有用的反馈，所以 SOLO 分类法可以用于形成性的学生学业评价；另一方面，如果将上述五个层次赋予不同的等级分数，那么学生对问题回答的质量就可以被量化，量化的分数可以作为终结性评价的依据。以下是依据 SOLO 的基本思想进行等级评分的历史开放性试题和物理开放性试题，从中我们可以了解基于 SOLO 的评分标准编制的一般方法。

## 2. 开放性试题的命制实例

### (1) 历史开放性试题

题目：请结合英国资产阶级革命的相关史实，谈谈你对英国议会制度的理解（资料来源：黄牧航，2004）。

本题为一道问答题，考查学生思维能力的种类包括再现历史知识、概括历史知识、分析历史事物的本质、阐述历史发展的规律等等。而根据 SOLO 分类法，可以把学生的回答分为以下五种情况：

①前结构的回答。这种回答只出现了一个简单的判断，我们不知道学生得出这种判断的根据何在；或者只提供了一个史实，我们不知道这个史实说明了什么问题。如：

答（1）：英国的议会制度对英国历史产生了巨大的作用。

答（2）：1649 年，议会把查理一世送上了断头台。

②单点结构的回答。这种回答只依据了一个方面的史实，说明学生还没有充分理解历史发展的复杂性。如：

答：英国资产阶级革命期间，议会组成了议会军，打败了王党军队，使立宪立宪制能够在英国确立。

③多点结构的回答。这种回答能够依据两个以上的史实，但是未能建立起这些史实之间的联系。如：

答：1640 年，查理一世为筹措军费而被迫重新召开议会，而议会却以此为契机要求限制王权。当两者的矛盾不可调和时就只能兵戎相见了。1645 年，议会军最终击败了王军，并于 1649 年把国王送上了断头台。

④关联结构的回答。这种回答不但能够提供多个史实，而且能够在这些史实之间建立起某种联系。如：

答：英国资产阶级革命其实是国王与议会之间的矛盾斗争引起的。1640 年，查理一世为筹措军费而被迫重新召开议会，而议会却以此为契机要求限制王权。当两者的矛盾不可调和时就只能兵戎相见了。1645 年，议会军最终击败了王军，并于 1649 年把国王送上了断头台。但是，议会与国王之间的矛盾并未根本解决。1660 年，因为议会内部矛盾不断，查理二世借机复辟。斯图亚特王朝复辟后，国王与议会的矛盾更加尖锐，最后，议



会不得不抛弃查理二世,把荷兰的执政威廉三世迎来英国当国王,使立宪君主制最终在英国确立,国王与议会的矛盾也最终得以解决。

⑤抽象拓展结构的回答。这种回答不但能够提供多个史实,而且能够把对问题的认识上升到抽象的理论认识层次,能够从理论上分析为什么会出现这种现象,以及这种现象的发展规律。如:

答:英国的议会起源于中世纪,议会与国王之间的矛盾是一直存在的。在中世纪,两者的矛盾主要体现在王权与贵族特权的纷争。到了近代,随着资产阶级势力的发展,两者的矛盾则体现在新兴的资产阶级对君主专制统治的不满,要求分享政治的权利。……(省略具体的史实)“光荣革命”后,资产阶级立宪君主制最终在英国确立起来,表面上这种制度的特点是国王与议会并存,而实质上是议会的权力日益超过国王,因此,资产阶级终于能够通过议会来实现对国家政权的控制。

## (2) 物理开放性试题

教师任务:将一气垫导轨倾斜放置,用标尺标明导轨上各点的位置,让滑块从导轨的高端滑下。不断重复这一过程(数据来源:高凌飙,2005)。

学生任务:观察上述过程并进行你认为必要的测量(可以利用自己的手表和导轨上的标尺),记录下结果,利用观察和测量的结果,对上述过程进行描述。应清楚地写出必要的文字说明、推导论证、公式计算、数据结论、图线表格等等。

评分细则:

本题无参考解答,教师可参照下面标准给学生评分:

- 0分:没有回答,或给出的回答言不及义、不对题。
- 1分:只有定性的描述,基本上能说清滑块的运动过程。
- 2分:有定性的描述和定量的测量。其中定量的测量只是测出几个时间及距离而没有做进一步的分析整理,也没有结论;或测得的数据与实际的差距很大。
- 3分:有定性的描述和定量的测量。其中定量的测量包括对滑块运动到不同位置时,滑块与原点的距离及运动所经过的时间,并能用这一数据计算出滑块运动的平均速度或实时速度这两个量中的一个。测得的数据基本上合理。
- 4分:情况同上,但能将滑块运动的平均速度和实时速度两个量都求出来。
- 5分:在前面的基础上,整个报告的叙述,包括目的、原理、过程、数据记录和整理、最后的结论等,条理清楚,结论明确。
- 7分:在前面的基础上,进一步求出滑块运动的加速度。
- 9分:在前面的基础上,能利用所掌握的多种数学工具,包括图表、图线等等,将实验观察的结果非常清楚地表达出来。
- 10分:在前面的基础上,能从实验事实出发对滑块的运动以至普遍情况下的匀加速运动,或其它任何问题进行讨论,且言之成理。

其它形式的答案,可参照上述等级给分。

## 《促进教师发展和学生成长评价研究》课题组



《促进教师发展和学生成长评价研究》课题组于 2002 年开始对发展性评价的理念和实施进行了广泛的探讨。课题组认为, 发展性评价区别于选拔性评价和水平性评价, 指的是在发展的整个过程中进行的, 旨在促进被评价者不断发展的评价(董奇, 赵德成, 2003)。发展性评价的目的是为了更好地促进学生的成长, 促进教师教育教学水平的提高, 促进学校的发展。发展性评价主张评价主体、评价内容的多元化和评价方法的多样化, 提倡形成性评价和终结性评价的有机结合。在评价方式和工具的开发方面, 该课题组着重探讨了质的评价方式—档案袋评估的实施。

### (一) 档案袋评估的定义

档案袋评估(portfolio assessment), 又称作“成长记录袋评估”。该课题组认为, 档案袋评估就是指“学生作品的有意收集, 以反映学生在特定领域的努力、进步或成就。它必须包括内容选择过程中的学生参与、选择的指南、评分的标准, 以及学生自我反思的证据”(赵德成, 2003a)。档案袋评估之所以受到重视是因其具有呈现学生学习进步情况和学习过程的潜能, 以及重视学生对学习的自我反省能力, 是一种能够适应学生个别差异的评估方式。综合而言, 档案袋评估以促进学生发展为根本目的, 强调个别差异的学习和成果的呈现, 而且着重教学与评估交互进行的过程。档案袋评估具有以下的特点: (1) 档案袋评估的基本成分是学生作品; (2) 学生作品的收集是有目的、有计划的, 而不是随机的; (3) 档案袋评估要给学生提供发表意见和对作品进行反思的机会; (4) 档案袋评估不是在整个教学结束后才进行, 其评估与教学是同步交互进行的。

### (二) 档案袋评估的类型

#### 1. 展示型档案袋

展示型档案袋是学生工作最佳成果的搜集, 是教室中最常实施的档案袋评估类型之一。展示型档案袋的主要使用者是学生自己, 其主要目的是促进学生对自己作品的反思。展示型档案袋中搜集的事物包罗万象, 包括手工作品、作业、照片、奖状、纪念品、论文、短评、录像等。对于每一项放入展示型档案袋中的作品, 学生应该讨论将其放入档案袋的原因, 下列问题可供他们参考: (1) 为什么我会选择这个作品放入档案; (2) 这个作品是如何显示我的学习和进步; (3) 我是怎样做使得这个作品成为“最优秀作品”等等。

#### 2. 文件型档案袋

这类型的档案是学生在学习历程的观察及记录, 主要是描述学生一段时间内的进步以及教师的期望。文件型档案袋的内容可以包括草稿、头脑风暴活动的结果, 也可以放入学习历程的任何作业、观察记录、评估清单、讨论记录等。要利用文件型档案袋有效评估学生的进步情况, 必须依据每一项学习内容选择数个作品作为其进步的证明, 因此, 选择的要包括三部分: (1) 原始的草稿; (2) 修改过的草稿; (3) 最后的草稿。与展示型档案袋相比, 文件型档案的限制在于耗费时间长, 数据太多, 不易评估。针对耗时的问题, 教师可以在平时有目的地记录。至于资料太多的问题, 教师应评估档案可以评估学生的进步的资料, 无需对所有数据加以评估。

#### 3. 过程型档案袋



过程型档案袋记录了学生在一学年当中对不同课题内容的学习过程, 主要包括学习历程与自我反思两个方面: 第一方面包括作品从萌芽到最终定型的过程, 第二方面包括学生对作品的反思过程。自我评价与反思是学生参与档案袋评估的最高水平。要想使学生的自我评价与自我反思起到预期的作用, 教师不能只是简单地提出要求, 要给学生以必要的指导, 让学生知道评价什么、反思什么, 以及怎么评价和反思。给学生设计一个用于自我评价或自我反思的表格, 是一种值得提倡的做法。表格的栏目设置应该鼓励学生反思他们为什么选择这些内容, 以及他们所选择内容的突出特征是什么。

#### 4. 评价型档案袋

与其它类型的档案不同, 评价型档案是在预设的要求下搜集数据, 并在预设的标准下评估学生, 其目的是收集事实以系统评价学生的学习, 并将结果报告给其它人。评价型档案袋试图通过简化评价任务、统一评价要求以及指导档案袋里面应该放什么材料来保证评分过程的信度, 评价型档案可以应用于区域性的大规模评价活动。评价型档案的优势在于有比较可靠的评分标准, 缺点在于: 缺乏真实性, 资料搜集方向的选择权在教师而非学生, 其主要目的是评估学生在特定工作上的表现, 减低了教学目的中所要求的真实性。

### (三) 档案袋评估优势与局限性

该课题组认为, 一方面档案袋评估具有许多传统的评估方式不具有的优势, 但同时也存在一定的局限性(徐芬, 赵德成, 2001)。档案袋记录了学生成长的“故事”, 是评价其最终发展水平、努力、反省和进步的理想方式。相对于纸笔测验和其它快照式评价而言, 它能向教师、家长和学生本人提供丰富的内容, 不仅能够反映学生知道些什么和能做些什么, 而且能够反映学生的问题解决能力和思维能力、运用策略和程序性技能的能力, 以及建构知识的能力, 为评价者提供很多关于学生的学习和发展的重要信息。由于档案袋评估并不是局限于教学之后进行, 因此可以有效地促进教学与评价的有机结合, 帮助教师及时、准确地获得有关学生的学习信息, 为学生提供适合其特点的教学。另一方面, 源于档案袋评估本身的特性以及我国的教育背景, 也带来很多问题, 有待于我们进一步研究: (1) 档案袋评估加重教师和学生负担的问题; (2) 档案袋评估的信度和效度问题。

## 结语

学业成就评价是教育评价、课程评价、教学评价、学生评价的重要组成部分, 是在系统收集学生课业学习情况的客观事实材料基础上的价值判断过程(吴维宁, 2007)。学业成就评价的方法, 既包括以纸笔测验为主要方式的量的方法, 也包括以观察、谈话、收集各种反映学生学习过程与结果的作品在内的质的方法。从国内三个主要课题组对学生学业成就评价的研究成果来看, 在量的方法和质的方法方面都取得了一定的进展, 但仍然存在很多问题。



首先,在量的评价研究方面,虽然现有的研究根据课程标准的教学要求,提出了与之相匹配的命题原则,但是如何细化和具体化这些原则使之更具有系统性,建立规范的评价理论依据和基本框架,仍然需要进一步的思考。例如,如何界定科学概念理解的层次水平,科学探究能力发展的层次水平。在纸笔测验中,如何衡量情感态度价值观的考察限度等等。

其次,在开放性试题的命制方面,现有的研究已经意识到评分标准的效度和信度问题,尝试引用教育评价领域的分类理论克服开放性试题的主观随意性弊端,但是如何结合各个学科的特性,选择与之相适切的评价理论和方法,还需要研究者在实践中不断摸索。例如,SOLO 分类评价理论适合哪些学科的试题命制?如何借助于新一代的测量理论和技术促进开放性试题的发展?

最后,在质的评价研究方面,实现评价与教学的有效结合,开发具有可行性和操作性的评价工具,避免质的评价方法异化成管理和束缚学生的手段,是未来研究的着力点。值得思考的问题是,现代教育评价理念提倡将各种非正式评价纳入评价的范围,如课堂提问,这样的做法是否会造成评价概念的泛化?如何结合我国的教育背景开展质的评价?

## 参考文献

Biggs, J. B., & Collis, K. F (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press.

Biggs, J. B. (1987). *Student Approaches to Learning and Studying*. Hawthorn, Vic.: Australian Council for Educational Research.

董奇, 赵德成 (2003)。发展性教育评价的理论与实践。中国教育学刊, (8):18-21。

高凌飙 (2004)。关于过程性评价的思考。课程·教材·教法, (10):15-19。

高凌飙 (2005)。探索和思考:“新课程背景下高中学生学业评价”项目阶段总结。

<http://xybj.cersp.com/XMYJ/XGZL/PJLL/200509/19.html>

高凌飙,吴维宁 (2004)。开放性试题如何评分?——介绍两种质性评分方法。学科教育,(8): 1-6。

高凌飙, 吴维宁, 黄牧航 (2006)。开放性试题的编制与评分。人民教育, (1): 36-38。

黄牧航 (2004)。SOLO 分类评价理论与高中历史试题的命制。历史教学, (12): 58-63。

教育部初中毕业学业考试评价理科科目组.中考命题指导(理科)。江苏教育出版社。

教育部华南师范大学基础教育课程研究中心 (2004)。高中学生学业评价研究资料汇编。

罗星凯, 刘小兵, 曾平飞, 赵光平, 薛跃规, 黄都 (2004)。新课程:“双基”和“探究”如何考——关于科学学科中考走向的研究与思考。人民教育,(9):33-38。

罗星凯, 曾平飞, 赵光平, 刘小兵, 薛跃规, 吴德汉, 唐力 (2005)。理科中考:从 2004 看 2005。

全国初中毕业升学考试综合理科评价课题组 (2004)。2004 年中考命题指导丛书(综合理科)。江苏教育出版社。

吴维宁(2007)。理科教师学业评价观研究。华南师范大学博士学位论文(未出版)。



徐芬，赵德成（2001）。档案袋评价在中小学教育中的应用。教育研究与实验，（4）：50-54。

赵德成（2003a）。成长记录袋的创建与使用。天津师范大学学报，（4）：38-42。

赵德成（2003b）。成长记录袋在大规模、高利害评价中的应用。教育理论与实践,(8):26-29。

赵德成，徐芬（2002）。成长记录袋应用的反思与改进。语文建设，（9）:43-44。

钟媚（2005）。过程性评价：概念、范围与实施。当代教育科学，（14）：44-47。