

科学教育中的总结性评测浅析 ——以一次区级调研考为例

文_陈 彧/中国科协“做中学”科学教育实验项目教学中心(东南大学)

当我们想要了解学生对于一定的学习目标达到什么水平时,我们要设计一些方法来搜集相应的信息,并作出判断,这就是评测。从目的上看,评测可以分为形成性评测和总结性评测。本文将以某市一次区级调研考为例,分析在进行书面总结性评测(参见本期专题文章《科学活动能力现场作业检测的设计和实施》)时有哪些值得注意的问题。

◎ 总结性评测的目的

与“为了”学习的形成性评测(参见本刊上期专题文章《科学教育中形成性评测的原则与策略》)不同,总结性评测是“对于”学习的评测,它的主要目的是在特定的时间(如学期末、升学时)向学生、教师、学校或教育行政部门报告学生的成绩。可能有些报告的结果能为教学带来指导,但这并不是它的主要目的。比如学生十分熟悉的期末测试,它也许会对学生敲响警钟,但并不像形成性评测那样即时地作用于教学过程,影响教师的下一步教学计划。

国际科学院联合组织 Global Network of Science Academies (IAP) 的科学教育工作组 2013 年的研究报告《评测与探究式科学教育:政策与实践问题》(Assessment and Inquiry-Based Science Education: Issues in Policy and Practice) 将总结性评测的流程归纳为图 1 (Harlen 2013, p21)。与形成性评测环(参见本刊上期专题文章《科学教育中形成性评测的原则与策略》)不同,施测者得到

学生的成绩报告后,并不会用它来设计新的教学活动,而是用来向学生、家长和教师反馈(如期末考试),或是用来决定学生是否能够进入高一级的学校学习(如升学考试),或是进行学校之间、地区之间、人群之间的比对应。

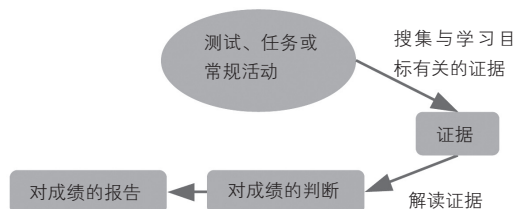


图 1 总结性评测的流程

以某市一个区 2013 年上半年对该区 5 年级学生的调研考为例,作为一次调研考,它的目的是检测全区学生对 5 年级下学期教材中的科学知识的掌握情况(教材内容见表 1),向教育行政主管部门、教研室、学校和科学教师反映全区和各个学校的教学质量,以便有针对性地指导以后的教学。

◎ 总结性评测的形式

搜集学生学习证据的形式是多样的,可以是书面的测试题、学生的手工作品、在学生完成某项任务时对学生的观察等。这些形式各有千秋。以书面测试题为例,它易于实施,体量可以比较大,也便于制订客观统一的评判标准,有利于保证评测的公平性和重复性,容易提高评测的信度。但是学生的探究能力、在实践中应用科学概念的能力、合作以解决问题的能力

表1 调研考对应的教材内容

第1单元 神奇的机械	杠杆、斜面、滑轮、轮轴、齿轮的基本作用及其在生活中的应用
第2单元 形状与结构	锥、台、柱、球等形状和它们所能承受的力的相对大小；三角形支架的稳定性
第3单元 它们是怎样延续后代的	植物除种子繁殖以外的繁殖方式，动物的卵生和胎生等繁殖方式
第4单元 岩石与矿物	岩石和岩石的种类；常见矿物及其用途
第5单元 人体的“司令部”	大脑的特征；大脑作为“司令部”的作用；脑神经和脊神经的工作流程

等在书面测试中常常无法得到充分的评价，这在一定程度上影响了评测的效度。再进一步，书面评测本身也包括有确定答案的问题和没有唯一答案的开放式问题，不同题型也是各有利弊的。因此，在设计评测时，难免要根据评测的目的在不同评测形式间进行权衡，作出取舍。下面结合本文所引调研考实例，分析两种较有价值的总结性评测题。

整理题挖掘学生的思维，凸显理解

这次考试想要考查的是学生对教材中的知识的掌握情况，那么，书面测试是一个合理的

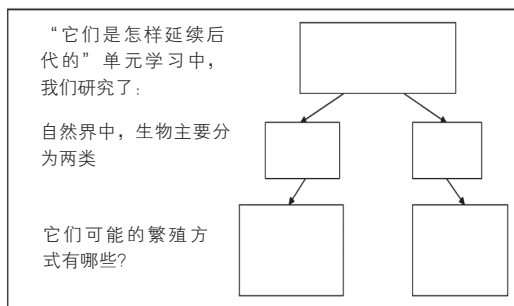


图2 调研考中的整理题

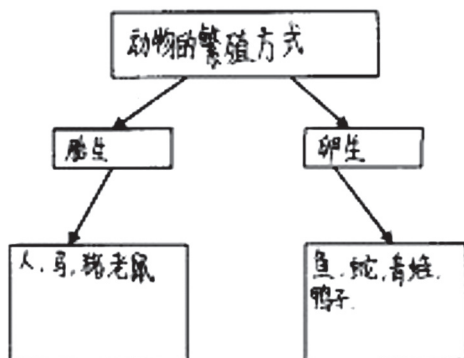


图3 学生答卷例1

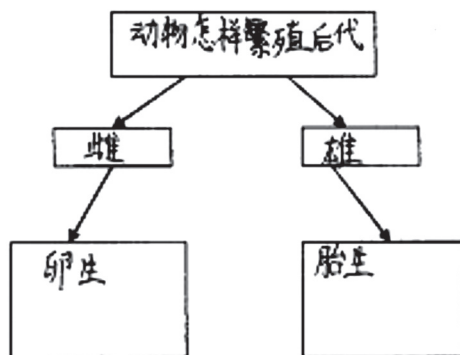


图4 学生答卷例2

选择。试卷以单选题、填空题和连线题这样的有标准答案的客观题为主，也包括一个需要学生系统地提取已学知识的整理题（图2）。这一题的满分为36.51%，比较低，反映出了其他题型所没有反映出的问题，值得借鉴。

根据教材的结构，这一题的本意是让学生填写他们在这个单元中研究了植物和动物的繁殖方式；生物主要分为植物和动物；植物可以用种子、根、茎、叶繁殖，动物有卵生、胎生、卵胎生等繁殖方式。相对于试卷中的其他题目而言，此题较为开放，学生的答案也是五花八门的。

以图3为例。这位学生忽视了他们还学过植物的繁殖方式。这样的答卷并不少见。可能的原因之一是学生对植物的不同繁殖方式印象不深。无论是种子繁殖，还是根、茎、叶繁殖，都是通过植物器官来繁殖，这两类器官在繁殖意义上有什么差异、两类繁殖方式之间究竟有着怎样的更深层次的区别，学生并没有理解。相反，动物的胎生与卵生是学生有密切接触的，在他们眼中区别更大。这说明学生对生物的繁殖，尤其是植物的繁殖，可能尚处于描述科学事实的水平；他们没有将植物的繁殖纳入整个生物繁殖的知识体系中并灵活取用。通过更进一步的访谈，能够获得学生忽略植物繁殖的确切原因。

再如图4。独立地看这份答卷的第2行和第3行，自然界中的生物可以按照性别分为雌、雄两类，生物（这份答卷也只考虑了动物）的繁殖方式可以分为卵生和胎生——对5年级学生来说，这样的概念都可以认为是正确的。但是学生却在这两个概念之间建立了错误的联

系。一种可能性是，他没有理解这两行之间为什么要用箭头连接起来，以为只是两个互不相关的填空题，因为他习惯了零散的填空而不是系统的整理。更糟糕的情况是，学生的脑中只有“卵生”、“胎生”这样的词汇，却没有把它们与相应的意义联系起来。不管是哪一种情形，都说明知识在学生的头脑中是散乱的、不成系统的。相应的课堂教学尚未达到让这个学生建构成体系的概念网的目的。

采用新的情境，考查学生对知识的应用

真正被学生理解的知识是这样的知识：它们能被用于解决问题、作出决策；因此，为了检验学生对知识的理解，给学生的评测任务“必须包含对知识的应用”，“必须是新的”（Harlen 2013, p49）。也就是说，好的评测题应该让学生在新的情境——而不是他们学习知识所用的情境——中应用知识。

在这次“做中学”高级研修项目中，Rowell 教授给教师们提供了一个评测量规的例子（见表 2）。在这个例子中，学生的发展水平 4 和水平 5 分别是将概念应用于与已知情境相似的情境和应用于新的情境。当学生的概念发展趋于成熟时，他们应能在新情境中应用概念。在本刊上期专题文章《嵌入形成性评测的教学实例》中，我们也可以看到教师采取了将概念应用于新情境的评测方法。

我们将课本上出现的情境视为学生已经学习过的情境，而将课本上没有出现过的情境视为对学生而言“新”的情境，那么在本文所引的调研考案例中，创设了新情境的问题如图 5 所示。

这 2 个题目引入了与学习情境不同的新情境，并且没有离学生的经验太远而使学生无法

表 2 学生科学概念发展水平量规

水平	在给出一个解释或作出一个预测时，学生
1	描述情况，但不解释
2	使用他 / 她自己的想法，而不是科学概念
3	引用相关的科学概念，但并不说明这些概念如何应用
4	只将相关概念应用于与已知情境相似的情境
5	将相关概念应用于新的情境
6	将几个相关概念综合起来，给出一个理由充分的解释或预测

如图，用同样的剪刀剪同样的纸，_____种剪法更省力？
A. 方法一 B. 方法二 C. 两种方法一样



方法一 方法二

请把下列情形下的信息传递过程填写完整。
当你看到这张试卷到目前 (3)

(1) 接受刺激

→

(2) 神经

→

(3)

→

(4)

→

传出神经

→

(5) 作出反应

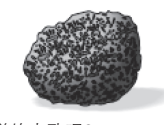
图 5 调研考中的新情境问题

理解；同时，它们是关乎真实世界、真实事物的这些特征符合 IAP 报告对新情境评测题的要求（Harlen 2013, p49）。

作为例子，让我们再来看一个创设了新情境的评测题（图 6）。

这是 TIMSS (Trends in International Mathematics and Science Study, 国际教育成就调查委员会 IEA 开展的多个国家和地区 4 年级与 8 年级学生数学与科学成绩的评估项目) 2011 年 8 年级科学评测中的一题，围绕岩石这个主题，涉及岩浆岩的生成过程与相应特征，与本次调研考所考查的“岩石与矿物”单元对应。在本次调研考中，相应的题目为图 7 所示。

一些火山岩中有许多小孔。



为什么会有这样的小孔呢？
A. 在岩石尚软的时候，小虫子钻进了岩石里。
B. 当岩石冷却的时候，气泡被留在了岩石里。
C. 在岩石尚软的时候，雨滴进入了岩石里。
D. 当岩石冷却的时候，有小石块掉了出来。

资料来源：TIMSS 2011 年评测题
Copyright© 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and IEA, IEA Secretariat, Amsterdam, the Netherlands.

图 6 新情境评测题案例

当地球表面的岩石被深埋于地下时，往往会在高温和高压下发生变化，由这种方式形成的岩石叫_____。
A. 变质岩 B. 岩浆岩 C. 沉积岩

图 7 调研考试题

这个题目重复了书中关于变质岩的定义，如果学生记得这个定义，就能正确回答。但是，“高温和高压下发生”的变化究竟是怎样的？它们所带来的特征与灼热的岩浆冷却后形成的岩石特征有什么区别？如果拿到一块岩石，怎样区分它是变质岩、岩浆岩还是沉积岩呢？3类岩石的定义能指导学生去解答这些问题，但仅仅记忆定义是不能确保解答的。从另一个角度看，变质岩、岩浆岩、沉积岩是从岩石的生成过程来划分的，那么学习这种岩石分类的意义在哪儿？对应着什么样的大概念或具有更高解释力的概念？一个可能的意义和概念目标在于理解地球内部和表面的过程。因此，为了检验学生是否真的理解了3种岩石的定义，是否建构了这种分类背后的科学概念，像上面的TIMSS例题一样，给出具体的岩石图片、给出对地质过程或地球历史的描述，是更好的评测途径。当然，这里展示的TIMSS评测题是面向8年级学生的，在不同年级采用这种方法的时候，要作出相应的调整。

◎ 评测结果的分析

总结性评测是为了向学生、家长、学校或决策者报告学习情况。带有调研性质的评测通常还需要对评测结果进行分析，以期呈现教学中反映出来的问题。

本文所引用的调研考以反映全区和每个学校的教学质量为基本目的。比较不同学校的情况是调研考所要收集的重要信息之一，以便支持教育管理部门和学校有针对性地开展接下来的教研和教学活动。因此这个区的教研员对评测中各个学校的情况作了比较。类似地，TIMSS和PISA (Programme for International Student Assessment, 经济合作与发展组织对多个国家和地区15岁学生技能和知识水平的评估)这类面向世界范围的学习水平测试，也会比较各个国家和地区的情况。

该区教研员统计了各个学校学生的平均分。平均分最高的学校获得了54.05分，平均分最低的学校获得了37.10分。结合试卷总分60分来看，这样的均分差异较大。由于本次调研考除了对科学知识的测试外，还以问卷的

表3 全区学生不同题型的正确率

试题类型	全区学生的正确率
识记	85.24%
理解	72.26%
应用	67.21%

形式调查了各个学校科学课的开展情况，包括课时数、学生动手实验的频率、学习科学的兴趣等，更直接地了解了一线教学的实情，因此教研员下一步可以结合学生在调查问卷中反馈出来的情况分析校际差异产生的原因。另一个有意思的结果是，在这个区的20余所小学中，有大约一半的学校为班额不超过28人的小班额学校，而在这次考试中均分排名前10的学校里，有8所为小班额学校。虽然严格评估班额对学生学习表现的影响仍需要更多数据，但这已经在一定程度上体现了小班额的优越之处。

教研员还将全部试题分成识记题、理解题和应用题，对不同题型的正确率进行了统计。全区学生的正确率见表3。从识记题到理解题再到应用题，正确率逐渐下降。由于识记题考查的是学生对某种科学现象或事件的记忆，而理解题和应用题需要学生对某个科学概念（这里的“概念”是指“对所观察到的相互关系或特性进行解释后的抽象”（哈伦2011，p8），是可以解释科学现象或事件的科学道理，而不是科学现象或事件本身）进行理解、推理和举一反三地在实际情境中使用，因而这三类问题对知识在头脑中的建构要求是越来越高的。这样的正确率梯度表明，整体看来学生更善于回答描述现象或事件的浅层次问题，而对于现象或事件背后的科学概念的建构水平要逊色一些。经过这样的分析，以后该区的教学会更加重视科学概念的应用。

致谢

本文得以完成，离不开相关教研员老师提供必需的试卷、调查问卷、学生答卷案例、各题型正确率统计和各学校的得分统计，在此表示诚挚的感谢！

参考文献

[1] Harlen, W. 2013. Assessment and Inquiry-Based Science Education: Issues in Policy and Practice. Global Network of Science Academies (IAP) Science Education Programme (SEP).
[2] 哈伦. 2011. 科学教育的原则和大概念. 韦钰译. 科学普及出版社. ■