

Asia-Pacific Forum on

Science Learning and Teaching

亞太科學教育論壇

亚太科学教育论坛, 第九期, 第一册

前言

科学课程学业成就调查的实施

杨宝山

中央教育科学研究所

中国 北京 100088

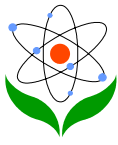
电邮: jks.yangbaoshan@cnier.ac.cn

内容

- [前言](#)
- [科学课程学业成就调查的基本框架](#)
- [科学课程学业成就调查的核心内容](#)
- [科学课程学业成就调查的主要方法](#)
- [参考文献](#)

前言

国际的学生学业成就调查作为 IEA（国际教育成就评价协会）和 OECD（经济合作与发展组织）两个国际组织主持的大规模的评价项目。从调查取向来看，IEA 较为偏重于考查学生基于学校教育的学业情况；OECD 较为侧重于考查学生离开学校教育的成年后的生存能力。从调查内容来看，两者均包括国际上公认的阅读、数学和科学等核心学科。所有参与的国家大都根据学生学业成就的世界排序，认真检讨各自的课程教学、教育制度、社会文化和家庭环境等众多方面的影响。此外，美国的（NAEP）调查，英国的（APU）调查，日本的国研调查等都曾产生不同程度的影响。基于此，我们从调查的基



本框架、核心内容和主要方法三个方面探讨科学课程学业成就调查的实施具有重要的意义。

科学课程学业成就调查的基本框架

从科学课程学业成就调查（评价）的实施来看，在宏观层面上，一是直接参加 IEA 和 OECD 等组织的国际学生学业成就调查；二是如美国的 NAEP 调查、英国的 APU 调查、日本的国研调查等，定期组织不同年龄段学生（9、13 或 15、17 岁）科学学科的学业成就调查，建立国家常模，开展纵向或横向比较研究。

在上述科学学业成就调查中，大都包括成就测验和背景因素调查两个部分。其中学业成就测验分学科领域的核心内容和表现水平两个维度。关键内容指课程内容标准中的核心内容；表现标准是要求学生达到的掌握水平。表现标准划分的基本框架是布卢姆认知领域的教育目标分类（知识、理解、应用、分析、综合、评价）。此外，加涅的学习结果分类和安德森等人提出的认知目标二维分类体系也在不断影响着学生学业成就的调查研究。

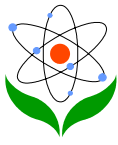
影响学生学业成就的因素，一般分为学习者、课程教学、学校及教育制度四个维度。学习者层面又包括性别、社会背景、自我认知、动机以及教育期望等；课程教学层面包括教学策略、课堂气氛、班级大小、教师支持、教材使用、学校组织和结构等；学校层面包括学校类型、学校结构、学校资源、学校管理以及课堂实践等。

TIMSS 的科学评估框架由内容维度和认知维度两个部分组成。在内容维度中，涉及的领域主要包括生命科学、物质科学、地球科学等几个部分；在认知维度中，主要包括领会、应用和推理等几个部分，每一部分又包括不同层次能力的要求。科学探究能力作为主要评价指标，始终贯穿于整个评价的内容与过程。

PISA 的科学评估框架基于终身学习的动态模型。其基本理念基于两个层面：第一，学生在校的学习是将来所需知识和技能的基础；第二，学校的功能在于使学生具备终身学习的能力。在评估中，重点测评学生所具备的阅读、数学、科学等基础性知识、技能、态度、情感等方面在真实情景中的运用情况。其中科学素养是其测量中的一部分。

NAEP 的科学评估框架包括科学领域与认知要素两个部分。科学领域涉及地球科学、物质科学和生命科学；认知要素包括概念理解、科学探究和实用推理三个要素。该评估框架规定了 4、8、12 三个年级学生学业的评定水平。

APU 的科学评估框架基于过程技能的评估。主要包括图表、器材的运用，观察、设计、分析、探究等几个主项目。在评估中，重点测评学生所具备的科学过程技能。



在中国内地，有关科学素养的测评框架问题也曾进行过诸多的探讨和实践。特别是新课程实施后，科学素养作为科学课程教学的首要目标。在科学学科的学业成就评价中，更加突出了基于科学的知识与技能、过程与方法、情感态度价值观三个维度的科学评估架构的科学素养的考察。

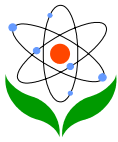
科学课程学业成就调查的核心内容

在 TIMSS 科学评估中，包括内容维度和认知维度两个部分。在内容维度中，TIMSS 科学评估框架的每一领域包括几个主题，每一主题涵盖了多数参与国家的科学课程标准中一系列目标，同时提供了每一主题相应的一组评估目标。在认知维度中，包括三个认知领域：一为领会，包括学生需要知道的事实、过程和概念；二为应用，集中在学生遇到问题时应用知识和概念理解的能力；三为推理，超出常规问题的解决方案以围绕新情景、复杂关系和多步骤的问题展开。TIMSS 科学评估框架中的三个认知领域由不同层次行为区分构成，描述在不同认知领域中学生的技能或能力。领会中包括识记、下定义、描述、举例说明、使用工具；应用中包括比较/分类、使用模型、关联、说明信息、拣出解决办法、解释；推理中包括分析/解决问题、整合/综合、假设/预测、设计/计划、得出结论、概括、评价、证明。另外，每个领域中列出了与其相对应的具体的行为。

有关学校、教师和学生的问卷调查，主要调查学生科学学习的情况以及对科学学习的信念；教师问卷主要调查教师对科学的信念和关于教学实践的信念；学校问卷主要调查校长对学校政策和实践问题的看法。在 PISA 的科学评估中，重点测评学生所具备的科学基础性知识、技能、态度、情感等方面在真实情景中的运用能力。有关科学能力的测量，PISA 将科学能力定义为运用科学知识的能力、识别问题的能力和在事实上做出决定的能力，帮助人们对自然世界的认识。科学能力被认为是衡量教育产出的一个关键部分。该定义并非意味大量知识的贮存，而是强调科学地思考面临的现实问题。

例如，关于科学的概念。学生需要掌握一定数量的关键概念，从而可以理解自然世界中确定的现象和人类活动所导致的变化。关于科学的过程。对科学知识运用和理解能力的评估。亦即获得、解释、作用于现象的能力。此外，内容还涉及关于科学的条件和应用的领域等。

在 NAEP 的科学评估中，包括科学领域与认知要素。其中科学领域涉及地球科学、物质科学和生命科学，其中物质科学包括物理和化学；认知要素则细分为概念理解、科学探究和实用推理三个要素。为获取更多有用的评估信息，NAEP 设计了学生、教师、学校领导等问卷，用于了解学生的种族、家庭状况、学校类型等相关信息。



在认知领域中，概念理解要素重点考察学生对科学知识和概念的理解。其中，科学知识包括从学校教育和自然界中学到的各种事实、事件，以及用于解释、预测自然现象的科学概念、定律和理论。科学探究主要考查学生使用科学工具的能力，包括制定计划，使用工具获得信息，交流探究的结果等。实际推理考查学生在新的、真实世界中运用科学的理解等能力。

在最近十年三次测评中，该项科学评价非常强调对科学概念的理解，它在各个年级所占百分比都几乎达到 50% 左右，2005 年 12 年级最高，达到 56%；同时还可看出，年级越低，越强调科学探究。随着年级增加，对学生实际应用能力的要求也在逐步的提高。

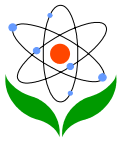
在 APU 的科学评估中，包括主要包括图表、器材的运用，观察、设计、分析、探究等几个主项目。在评估中，为每一个主项目提供了相应的一组评估目标。例如，有关实验器材的运用，主项包括：量度仪器的运用，物理数量的估算，根据提示进行实际操作等。又如，有关分析与应用的评估，主项包括：分析已经获取的数据，分辨不同程度的推论，以及评量学生应用科学概念包括物理、化学和生物概念的能力。再如，有关设计探究的评估，主要涉及三种不类型的活动，可以是设计整项探究活动，可以是提出可测试的假设，可以是设计探究活动中的某些步骤。

在中国内地，有关科学素养的测评内容问题也曾进行过诸多的探讨。新课程实施后，在科学学科学业成就评价中，更加突出了知识与技能、过程与方法、情感态度价值观等方面的综合考察。在科学知识的考查上，突出了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知识解释现象、分析和解决问题的能力。在科学探究的考查中，注意从实际问题出发，关注学生的探究能力。在情感态度与价值观的考察中，把情感教育渗透到具体的、生动的科学事例与活动中。

从科学学业成就调查（评价）的试题特点来看，在 TIMSS 科学评估的各种题型中，在领会方面，分别从识记、下定义、描述、举例说明、使用工具等几个方面入手展开问题；在应用方面，先后从比较/分类、使用模型、关联、说明信息、拣出解决办法、解释等几个方面入手展开问题；在推理方面，分别从分析/解决问题、整合/综合、假设/预测、设计/计划、得出结论、概括、评价、证明等几个方面入手展开问题。

例如，在领会方面的问题中，如说明期望学生能举出恰当的例子支持或阐明有关事实或概念的陈述，识别或提供详细实例来说明对一般概念的认识；在应用方面的问题中，如解释则期望学生能为观测或自然现象提供或识别一种解释，展示对一些基本科学概念、原理、定律或学说的理解等。

在 PISA 科学评估的各种题型中，重点考察学生所具备的科学知识、技能、态度、情感等方面在真实情景中的运用能力。



在科学的概念方面，要求学生需要掌握一定数量的关键概念，从而可以理解自然世界中确定的现象和人类活动所导致的变化。PISA 从物理、化学、生物学、地球和空间科学概念中集中提问。更具体一点，即从这些论题中定义概念，包括生物多样性、强度和运动以及物理变化。

在科学的过程方面，要求学生掌握对科学知识运用和理解能力的评估。也就是获得、解释、作用于现象的能力。PISA 测量五个过程：科学问题的知识、证据的鉴别、结论的提出、结论的交流和科学概念理解的演示。

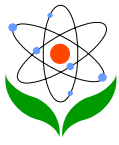
在科学的条件和应用的领域方面，命题时注意到科学能力的应用条件是指日常生活，而不是课堂或者实验室，这里包括对整体生活方向的论点，如个人直接关心的事件等。

可以看出，PISA 试题的一个显著特点就是十分强调知识和技能与日常的生活相关。它虽然考核的是学生科学学科的成绩，但却十分注重从日常生活取材，在设计问题时跳出课程具体的知识内容，大量地运用学生生活的真实情景，注重学生在相应学习领域的基础能力。学生必须理解关键的概念、掌握特定的过程，同时也要有在各种不同情景中运用这些知识与技能。此外，PISA 还特别注重收集学生态度以及学习方法等信息。

在 NAEP 科学评估的各种题型中，在科学知识方面，要求学生从科学教育以及自然界中学习到的各种事实、事件，用于解释、预测自然现象的科学概念、定律和理论；在科学探究方面，考查学生使用科学工具的能力，包括制定计划，使用多种科学工具获得信息，交流探究的结果等；在实际推理方面，考查学生在新的、真实世界中运用其科学理解能力。

在 APU 科学评估的各种题型中，有关实验器材的运用，在量度仪器的运用方面，利用量筒量度液体的体积或细小固体的体积，利用杠杆秤量度对象的质量，利用安培计量度电流等；在物理量的估算方面，估计箱子的体积、叶片的面积、铁线的长度、包裹的重力、皮球的质量等；在根据提示进行实际操作方面，转移指定份量的固体化合物和稀酸，将固体化合物和稀酸在试管中混合，然后利用本生灯将试管加热，最后过滤试管中的液体等。

在中国内地，有关科学素养的测评内容问题直接反映在科学学科的学业成就评价中，更加突出了知识与技能、过程与方法、情感态度价值观等方面的综合考察。例如，在科学概念知识方面的考察上，突出了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。如将食用油、酱油、食盐、酒精分别放入水中，请判断哪一种不能形成溶液。实际上，在考察学生对溶液的概念把握的程度。借助这样一个联系生活实际的情境，让学生利用所学知识回答现实生活中常见的问题。



科学课程学业成就调查的主要方法

在科学学业成就调查（评价）的方法上，宏观监测大都是以纸笔测验和问卷调查为主。目前，以自适应理论为依据的计算机因人施测技术和成长记录袋评价的运用逐渐增加。在问卷调查中，可选择学生、教师、校长或家长等作为对象；有时进行必要的访谈、实地观察及相关文本分析等，可进一步了解学生学业成就差异的成因。

在调查对象的选择上，常采用分层抽样技术。例如，在 PISA 调查中，设计为两阶段分层取样，第一阶段的取样单元为含有 15 岁学生的学校个体；取样之前，将在取样范围内的学校（学校的最低参与率为 85%）或内隐或外显地分配到层；然后，以随机抽样的方式在每层抽取学校，各层之抽取率与各层学校多少及学校内的学生数量成正比。第二阶段的取样单元为样本学校中的学生个体，从每一所样本学校中等概率抽取 35 名学生（底线为 20 名）。所抽取考生的作答率要求达到 80%（底线为 50%）。为了保证样本的总体覆盖率，学校和学生层面的总排除率必须低于 5%。

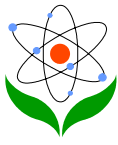
在 TIMSS 的科学知识的测试和调查中，除了测试和问卷调查之外，还包括课程分析、数学课堂的录像、观察和有关政策的研究。它的设计集中在学生的三个不同的学校阶段：小学后阶段、中学低年级阶段和中学的末尾阶段。由于每个国家学生的入学年龄不完全相同，学生被测试时必须考虑年龄和年级两个因素。

有关学校、教师和学生的问卷调查，主要调查学生的科学学习的情况以及对科学学习的信念；教师问卷主要调查教师对科学的信念和关注教学实践的信念；学校问卷主要调查校长对学校政策和实践问题的看法。

在课程分析中，主要对科学课程的指导书和教科书进行比较，研究科目的内容、内容的衔接和对成绩的期望等。他们通过对课堂教学进行观察，对教育管理部门、校长、教师、学生和父母进行访谈。研究的内容包括教育标准、处理个性差异的方法、教师的的生活和工作状况以及学生生活等对学校教育的影响等。

NAEP 的科学评估方法基于相同的评价框架指导，使用相同的评估程序，以及整个评估使用了有效的评估工具：大型题库和矩阵技术，并在不同年份的测试、不同年级的测试卷中特意安排了一些重迭的问题，与其它类型的测试共享部分相同的样本，使得测试结果既有纵向可比性，又有横向可比性。科学评估量表由学生问卷、教师问卷、学校领导问卷等组成。

在 NAEP 的科学评估中，在参加评估的州内根据人口统计学和地理组成进行抽样。为保证样本的均衡性，NCES 和 NAGB 规定州和地方学校的参与率不得低于 85%。通常在各州 4 年级和 8 年级各选取 100 所学校作为样本，再在作为样本的学校和年级选取 25 名学生参加每个科目的评估。



在测试时间方面，该组织制定了详细的评估计划表。全国评估通常与州评估和实验性城市地区评估是来年进行并且避免重迭，而全国长期趋势评估则是四年一次。今年（2008年）进行全国长期趋势评估。

在APU的科学评估中，由于学生和考题都是以随机方式抽样，因此两方面的样本都应具有代表性。在学生试卷的设计方面，由于评估所牵涉的内容项目众多，每次测试只能评估部份项目。每个子项目都设有一个试题库，部份试题适用于两个或三个年龄组别，以便于比较跨年龄的成就差异。在笔试前或后，会抽取部份考生进行实作测试。

在学校问卷的设计方面，包括科学教学的师资和资源，进行科学教学的目的，对科学教育的重视程度，分配予科学教学部门的财政资源等。在学生问卷的设计方面，主要考察学习环境与学生学业成就差异的关联。如学生的性别，种族，修读科目，职业取向，对有关科学议题的兴趣；问卷亦问及学生课余的兴趣及嗜好等。

在中国内地的科学评估中，由于多种原因，我国大陆虽然未整体直接参加IEA和OECD等国际评价组织的学生学业成就评价项目，但是从20世纪70年代后期开始，我国的理论界先后参照布卢姆等的教育目标分类框架，从认知、技能、情感态度和学生品德等多方面地进行了引进、吸收和本土化改造。20世纪90年代起，我国在学生学业成就现状调查和影响因素分析方面进行了多次的探索和实践。“十五”课题以来，特别是自新课程实施开始，一些项目组、课题组进行了广泛的理论研究和实践探索，并且已经取得了一些具有重要意义的成果。

参考文献

1. 国际教育成就评价协会: <http://www.iea.nl>
2. 经济合作与发展组织: <http://www.oecd.org>
3. 美国的NAEP调查: <http://nces.ed.gov/nationsreportcard/itmrls/>