

## 发展「真实性评量」来探讨台湾中学生的科学素养

张淑女

真理大学通识教育学院

收稿日期：二零零七年十月廿五日(于二零零八年六月三十日再修定)

---

### 内 容

- [研究背景](#)
  - [理论背景](#)
  - [测试方法](#)
  - [研究结果](#)
  - [结论与建议](#)
  - [致谢](#)
  - [参考文献](#)
- 

### 研究背景

自从科学素养一词在 1950 年代被 Paul Hurd 所提出后，科学素养成为近数十年来全球从事科学教育的研究人员与在职教师们的终极目标(Laugksch, 2000)。而伴随着科学素养的科学教育目标，许多国家的教育改革潮流也随之展开，台湾也不例外。

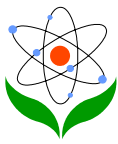
台湾最早在 1968 年开始实施九年国民义务教育改革，而最近的一次课程改革是在 2001 年开始实施所谓「九年一贯课程」(Chang & Chiu, 2005)。在这一次的教育改革理念指出，国民中小学课程应以人的生活为中心，配合学生身心能力发展历程；尊重个性发展，激发个人优良潜能；涵咏民主素养，尊重多元文化价值；培养科学知能，适应现代生活需要(教育部，1998)。根据上述的理念所发展出来的课程目标主要有三个面向(教育部，1998)：

「人与自己」—强调个体身心发展的关系，与促进个体的身心发展。

「人与社会」—着重于社会与文化的关系，与增进社会文化的参与。

「人与自然」—自然与环境的关系，正确认识自然并適切运用。

基于上述三大面向所产生的课程目标，可以进一步转化成个体在九年一贯课程所应积极培养的十大能力，如表一所示。而除了此十大能力的转化，课程改革人员也规划出七



大学习领域来加以落实，从国民教育的一年级至九年级，均提供语文、健康与体育、社会、艺术与人文、数学、自然与科技及综合活动等七大领域，用来取代过去升学考试的传统科目，希望统整学科知识与学生生活经验，避免知识的片断性，希望强化课程之间的连贯性与生活经验的结合；就七个学习领域之中的「自然与科技」领域而言，此领域所包括的学科包括自然科学（物理、化学、生物、地球科学）及科学应用（食品、织品、营建、运输、制造等方面的科技和信息传输、生物科技等）（教育部，1998）。

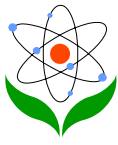
表一、九年一贯课程目标与十大能力（整理自教育部，1998）。

课程目标	十大能力
<b>（一）「人与自己」的面向</b>	
1. 增进自我了解，激发个人潜能。	1. 了解自我与发展潜能
2. 培养表达沟通、创作及审美能力。	⇒ 2. 欣赏、表现与创新
3. 获得运动常识与技能，陶冶体育情操。	3. 生涯规划与终身学习
<b>（二）「人与社会」的面向</b>	
4. 培养负责守法、积极参与、团体合作的民主法治知能。	4. 表达、沟通与分享
5. 发展互助合作的群己关系，重视伦理价值，提升道德判断能力。	⇒ 5. 尊重、关怀与团队合作
6. 启迪包容异己，尊重多元文化，恢弘国际视野。	6. 文化学习与国际了解
<b>（三）「人与自然」的面向</b>	
7. 启发探究兴趣、批判反思、创造发明等科学态度。	7. 主动探索与研究
8. 学习科学方法，理解科学概念，注重科学实验，应用科技新知。	⇒ 8. 规划、组织与执行
9. 了解自然环境，维护自然生态。	9. 独立思考与解决问题
10. 掌握科技信息，培养终生学习意愿。	10. 运用科技与信息

由上述的九年一贯课程改革的理念与目标，不难看出台湾对于科学发展与培养学生科学思考及科学态度的重视，以及促进学生能将科学知能与生活情境结合的决心，而这样的思潮十分符合全球所提倡的科学素养理念。基于此教育改革理念，以下内容将藉由笔者过去在真实性评量的研究经验来探讨其对于促进科学素养的可行性，以及介绍台湾在2001年的九年级学生科学素养中的认知指标表现情形。

## 理论背景

就 Miller 所提出的科学素养观点认为，在现今科学与科技快速发展的时代，科学素养应该要包含了解科学的方法、重要的科学概念与词汇，及科学与科技可能对社会所带来的冲击 (Miller, 1983)。其它的一些学者也提到科学素养的提升是为了改善个体在现今科技主导社会中的生活，而终极目标是希望能够提升国家的国际竞争力 (Aikenhead & Ryan, 1992; Laugksch, 2000; Thomas & Durant, 1987)。然而，在最近一次的台湾教育改革思潮，亦符合了上述科学素养的意涵。在台湾教改所规划出的七大学习领域中，在自然



与科技的学习领域, 教改委员们进一步规划出六个学习指标来促进科学素养的提升, 包括过程技能、科学认知、科学本质、科学态度、思考智能、科学与信息的应用 (教育部, 1998)。然而, 在这样的学习指标之下我们要如何知道学生是否有达到科学素养的目标呢? 在过去, 有学者发展出选择题来评量 11 及 12 年级学生对于科学—技术—社会 (Science, Technology and Society, STS) 的议题 (Aikenhead & Ryan, 1992), 或是利用是非题来评量学生对于科学本质, 科学认知, 与科技对社会冲击的观点 (Laugksch & Spargo, 1996)。然而, 从 Champagne 和 Newell 的研究发现, 美国学生在这些传统的评量方式表现很差, 认为应该要以真实性评量来评断及促进学生的学习动机与技能 (Champagne & Newell, 1992)。同时有学者也强调, 将科学概念融入真实情境是非常重要的 (Champagne & Newell, 1992; Yerrick, 2000)。而根据 Newmann 及 Archbald 的想法, 真实性成就 (authentic achievement) 的意涵指的是有受过训练的探究、高层次的思考及问题解决能力, 进而能够将这些能力及所学的知识从学校转移到真实的生活 (Newmann & Archbald, 1992)。这样的概念事实上符合于台湾教育改革理念以及科学素养的精神。因此, 在公元 2000 年及 2001 年的一个教育部及国科会共同补助的两年期研究计划 (计划总主持人为邱美虹教授), 即以真实性评量的理论为依据来设计发展评量系统, 探讨 1,469 位台湾九年级学生在此六项学习指标的表现 (Chang & Chiu, 2005)。

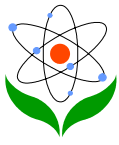
真实性评量 (authentic assessment) 是由 Wiggins 所提出之概念 (Wiggins, 1989)。在过去, 标准化的传统考试被视为是量化, 而真实性评量较属于另类质化的评量, 这两者在过去被视为是两种不同的文化, 但现在已经可以被整合来评量学生的高层次思考能力 (Dori, 2003; Wolf, Bixby, Glenn, & Gardner, 1991)。根据 Cumming 及 Maxwell 的真实性评量观点, 真实性评量具有评量四种能力的面向: (1) 探究评量—评量可以探究知识的能力; (2) 情境评量—特定情境的能力; (3) 问题导向评量—可以将学校所学迁移至生活的问题解决; (4) 价值评量—可以持有对职业价值观点的能力 (Cumming & Maxwell, 1999)。本两年期的研究计划即为发展评量自然与科技学习领域六项学习指标的真实性评量试题, 进一步根据上述四种真实性评量面向加以设计, 皆是以情境评量为基础, 再分别整合探究评量、问题导向评量及价值评量。评量的形式分为两种: 纸笔测验 (选择题与开放性试题) 与实作评量。详细分类与试题型式, 请参考后述之测试方法中的试题发展内容。

## 测试方法

在这个段落, 将呈现本研究的研究设计、取样方法、及试题的发展。最后, 亦将阐述本研究的研究资料分析。

### (一) 研究设计

本研究是一个经由国科会与教育部共同补助的两年期计划, 台湾师范大学邱美虹教授为计划总主持人。研究主要分为三个阶段: 研究与工具的发展阶段、实际施测阶段、与资料分析阶段。在研究与工具的发展阶段, 主要在发展具有核心情境的评量工具来评量科学素养的六个学习指标, 包括过程技能、科学认知、科学本质、科学态度、思考智能、科学与信息的应用 (教育部, 1998)。经由评量工具的信、效度检验之后, 进行实际施测阶段。在数据分析阶段, 进一步要了解真实性评量对于探讨科学素养的可行性, 及学生的表现为何。



## (二) 取样方法

本研究所采用的分层随机抽样方式是依据 TIMSS 的抽样原则, 从台湾的北部、中部、南部及东部, 共抽样了 44 所学校, 1,503 位九年级学生 (平均年龄 15 岁) 参加此真实性评量, 最后的数据分析有效样本数为 1,469 位学生, 而由于经费的限制, 只有邀请六所学校参加实作评量的部份, 总共 193 位学生参加。台湾的北部、中部、南部及东部各区的详细参与人数, 请参考表二。

表二、在各区域中依学校大小抽样的班级数目

学校大小 区域	北区 班级数/人数	中区 班级数/人数	南区 班级数/人数	东区 班级数/人数	总和 班级数/人数
大	17/517	8/264	8/271	—	33/1052
中	3/93	4/134	3/100	—	10/327
小	1/21	1/34	2/55	1/14	5/124
总和	21/631	13/432	13/426	1/14	48/1503

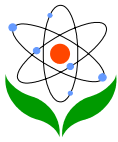
注: 学校大小之定义, 「大」为校内国三班级数超过 13 个班级; 「中」为校内国三班级数超过 6 个班级, 但少于 12 个班级; 「小」为校内国三班级数少于 5 个班级。

## (三) 试题的发展

本研究评量工具之发展是根据 Cumming 及 Maxwell (1999) 所提出的真实性评量理论为基础, 以调查台湾九年级学生的科学素养。为发展评量自然与科技学习领域六项学习指标的真实性评量试题, 进一步根据 Cumming 及 Maxwell (1999) 所提出的四种真实性评量面向加以设计, 皆是以情境评量为基础, 再分别整合探究评量、问题导向评量及价值评量。评量的形式分为两种: 纸笔测验 (包含选择题与开放性试题)、与实作评量。详细分类与试题型式, 请参考表三。

表三、符合自然与科技学习领域六项学习指标的真实性评量设计说明。

真实性评量的特征		六项学习指标	试题型式
情境评量	问题为本评量	科学认知	选择题 开放性试题 实作试题
		科学与信息的应用	选择题 开放性试题
		思考智能	开放性试题
	探究评量	过程技能	开放性试题 实作试题
	价值评量	科学本质	开放性试题
		科学态度	李克氏量表



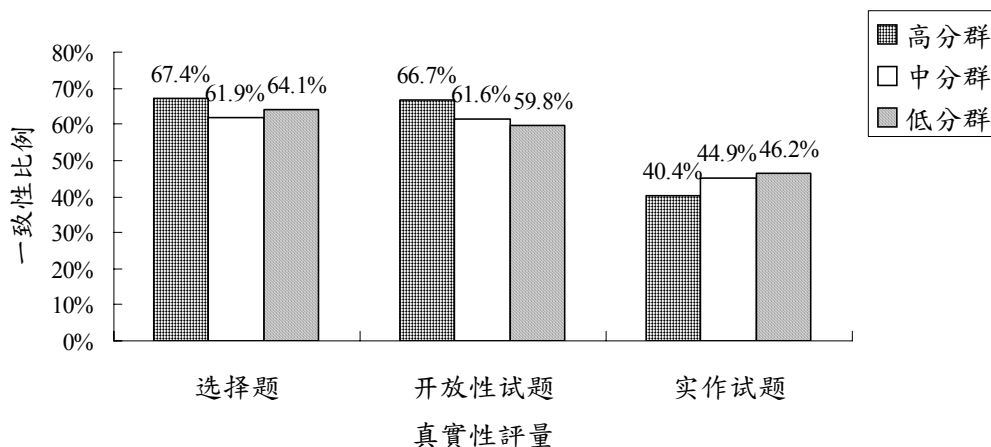
本文章主要着重在**科学认知**指标的设计与分析结果的陈述。在试题数目方面, 认知指标有 21 题选择题、4 题开放性问题 (60 分钟的施测时间), 而实作试题只有一题 (30 分钟的施测时间)。纸笔测验的选择题与开放性试题之题目的核心情境是以五位同学一起到阳明山露营为主, 试题包括: 学生们在组装驱动车时, 电学概念的应用; 煮水饺时温度与热的概念应用; 吃火锅时的现象; 咖哩鸡料理包的营养成分标示等等。而实作试题则是提供学生们每人四小瓶的眼药水滴瓶, 里面盛装四种不同的透明溶液, 同时提供学生酸碱指示剂与透明胶片, 让学生在透明胶片上自己进行实验, 辨别出四种透明溶液分别为  $\text{HCl}(\text{aq})$ 、 $\text{NaCl}(\text{aq})$ 、 $\text{CaCl}_2(\text{aq})$ 、还是  $\text{CaCO}_3(\text{aq})$ 。详细试题的内容, 请参考 Chang and Chiu (2005)。

经由专家效度的检验之后, 进行 Guttman 信度的测试结果, 21 题选择题为 0.63 ( $\text{Lambda } 2$ ), 而 4 题开放性问题为 0.80 ( $\text{Lambda } 2$ ), 实作试题为 0.85 ( $\text{Lambda } 2$ )。

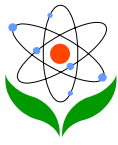
## 研究结果

研究结果, 以笔者所参与研发的**科学认知**评量为例, 试题包括三个部份: 21 题选择题、4 大题 32 小题的开放性试题、与一题实作测验。就纸笔测验 (21 题选择题、4 大题 32 小题的开放性试题) 之答对率而论, 对于生活中常见现象的问题形式, 学生成绩表现不理想, 只有一些记忆性问题 (如: 酒精的成分为何?), 或问题与课本内容较接近的题目 (如: 气体的原子表现形式) 回答的较好。而针对实作测验的结果, 在 193 位学生的施测后, 就每一试题之答对率而论, 学生对于操作式问题形式表现颇理想, 只有在回答「为什么」的这类开放性问题表现较差。

进一步, 结合台湾的高中入学考试成绩 (基本学力测验) 与本真实性评量的结果进行分析, 发现高分群 (前 27%) 的学生, 只有 66% 左右的学生在选择题与开放性试题的表现也属于高分群, 而甚至只有 40.4% 的学生在实作评量的部份也属于高分群, 此结果说明了真实性评量的传统标准化测验的差距 (图一)。此外, 若依据真实性评量的总成绩与实作评量的表现将学生分别分成高、中、低三群, 进行相关性分析, 发现不管选择题、开放性试题或实作评量皆与真实性评量具有高相关性, 其中实作评量与真实性评量的表现, 以低分群学生表现的一致性最高 (表四)。



图一、基本学力测验与真实性评量表现的一致性。



表四、真实性评量总成绩与实作评量部分表现的相关性分析。

实作试题	高分群	中分群	低分群
真实性评量			
总成绩	0.414*	0.273*	0.507*

\*.显著水平为 0.01 (双尾检定)。

## 结论与建议

由本研究的真实性评量结果，除了证实了发展真实性评量于评量学生科学素养的可行性之外，从本文章所呈现之台湾九年级学生在科学认知指标的表现，更告诉我们 真实性评量可以测出一般传统标准化测验所不能测量的认知能力，传统标准化测验表现优良的学生未必能在真实性评量有同样的表现，换句话说，在传统标准化测验 表现优秀的学生，是否踏出校园工作后，可以展现同样的能力，值得深思！

此外，针对在学校的智育成绩表现较差的学生，身为教师者应该利用各种多元的评量方式来了解学生的其它潜质。在本研究的施测过程发现，大部分学生对于实作评量的试题非常感兴趣，在评量的过程，也都显露出微笑，这些是传统纸笔测验无法观察到的表情；而本研究结果也告诉我们，实作评量与基本学力测验的一致性较高，因此，从简单的实作活动来教导与评量低成就的学生是从事科学教育的我们，可以使力的方向。

在现今科学与科技快速发展的时代，科学素养的意涵包括要让学生了解科学的方法、重要的科学概念与词汇，及科学与科技可能对社会所带来的冲击 (Miller, 1983)。综合上述，从真实性评量的研究结果，相信藉由真实性评量的发展，除了可以较确切地评量到学生是否能将课堂上学习的成果展现于真实生活的现象与问题解决之外，更可以激发学生将学校学习的知识转换至生活真实情境的动力；此外，笔者也认为，在知识爆炸的现代，获得知识的管道十分多元且方便，因此，学生所需具备的能力应该从过去的知识导向转变成知识应用与批判信息的能力养成，相信从发展真实性评量的方向着手，对于了解学生的科学素养与促进科学素养目标的达成，是可以有实质的帮助与贡献。

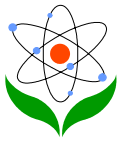
## 致谢

感谢国科会与教育部之研究补助 (NSC89-2515-S-003-012-X3 及 NSC90-2511-S-003-101-X3)，邱美虹教授在研究上的领导，以及研究团队所有教授们所提供的宝贵建议。

## 参考文献

教育部 (1998)。国民教育阶段九年一贯课程总纲纲要。

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "views on science-technology-society" (VOSTS). *Science Education*, 76(5), 477-491.



- Champagne, A. B., & Newell, S. T. (1992). Directions for research and development: alternative methods of assessing scientific literacy. *Journal of Research in Science Teaching*, 29(8), 841-860.
- Chang, S. N., & Chiu, M. H. (2005). The development of authentic assessments to investigate ninth graders' scientific literacy: in the case of scientific cognition concerning the concepts of chemistry and physics. *International Journal of Science and Mathematics Education*, 3(1), 117-140.
- Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education*, 6(2), 177-194.
- Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: students' performance in the matriculation 2000 project. *Journal of Research in Science Teaching*, 40(1), 34-52.
- Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84, 71-94.
- Laugksch, R. C., & Spargo, P. E. (1996). Development of a pool of scientific literacy test-items based on selected AAAS literacy goals. *Science Education*, 80(2), 121-143.
- Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, 112(2), 29-48.
- Newmann, F. M., & Archbald, D. A. (1992). The nature of authentic academic achievement. In H. Berlak, F. M. Newmann, E. Adams, D. A. Archbald, T. Burgess, J. Raven & T. A. Romberg (Eds.), *Toward a New Science of Educational Testing and Assessment*. Albany, NY: State University of New York Press.
- Thomas, G., & Durant, J. (1987). *Why should we promote the public understanding of science?* Oxford, UK: University of Oxford.
- Wiggins, G. P. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wolf, D., Bixby, J., Glenn, J. I., & Gardner, H. (1991). To use their mind well: Investigating new forms of student assessment. *Review of research in education*, 17, 31-73.
- Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry instruction. *Journal of Research in Science Teaching*, 37(8), 807-838.