

分析科学学习作评估示例

郑美红

中国 香港

香港教育学院科学系

电邮：maycheng@ied.edu.hk

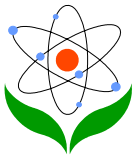
收稿日期：二零零二年十一月八日(十二月廿三日再修定)

内容

- [摘要](#)
 - [科学习作评估性质的转变](#)
 - [另类科学评估模式](#)
 - [可信有效的评估方法之条件](#)
 - [评估课业的开展模式](#)
 - [评分准则的运用及发展](#)
 - [方法](#)
 - [结果及讨论](#)
 - [综合分析习作示例](#)
 - [参考文献](#)
-

摘要

香港在近期的教育改革把评估重点从总结评估转移到进展评估(Curriculum Development Council, 2001)。学习评估除用以检视学生的学习表现外，亦反映了教师的教学情况，并能培养学生的自学能力及沟通技巧。面对这种评估模式的转移及新需求，教师须考虑使用同侪互评的形式、选用另类的学生习作、设计及运用合适的评分指引。为加强教师采用新评估方法的信心，本文提供了不同国家的科学习作示例，并作一分析。所选例子是从一项名为「寰宇学校计划」的国际性研究收集得来。计划旨在让教师主动讨论及交流习作样本，从而促进教师的专业发展。

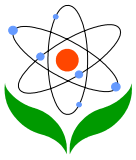


科学学习作评估性质的转变

本文以香港最近推行的教育改革为背景。根据政府的教育改革指引, 学生的学习经验可分为八个主要学习领域, 科学教育属其中一个; 九种共通能力亦列为教育重点, 包括协作能力、沟通能力、创造力、批判性思考能力、运用信息科技能力、运算能力、解决问题能力、自我管理能力和研习技能(Curriculum Development Council, 2002)。学与教的文化亦转向强调学会学习、全人发展、综合学习、多元化的学与教资源、走出课室的学习、以及为学生提供机会, 发掘自己的学习性向和潜能。因此, 教学与评估必须重新配合这些教育重点。科学教育学习领域的主要关注项目强调评估的重要性, 藉以提升学生的科学推理能力, 加强科学过程技能, 鼓励学生主动设计和参与实验, 探讨科学概念及将之应用于日常生活(Curriculum Development Council, 2002)。故此, 教师应留意标准化测验的种种限制, 例如未能量度学生的特质、把知识应用于现实生活的技能(Torff, 1997)、学习科学科的兴趣(Baron, 1991)、也未能鼓励学生深入及全面理解知识(Kulm, 1991, Haetrel, 1991; Raizen and Kaser, 1989)。没有一种评估形式能评核学生每一个学习层面的表现, 就如单一的教学策略并不能涵盖所有学习情况 (Doran, Chan and Tamir, 1998)。要包罗更多不同评估形式, 量度更复杂的技巧, 则须利用各式各样的评估课业。不过, 这并不代表背诵数据这种较低层次的学习形式对学习过程毫无作用, 而是学生熟读内容的程度, 不应视为单一的评估范畴。

香港的教育制度较偏重常模参照的评估模式。学生要在总结评估(测验及考试)取得理想成绩, 透过公开考试成绩作为编班及批准入学的依据。常模参照评估模式, 是指以其它接受评估学生所取得的等级作为评估比较(Marsh and Morris, 1991)。如教师采用常模参照评估而非标准参照评估, 学生之间则较常出现互相比对的情形。久而久之, 成绩较好的学生会感到学习缺乏挑战, 而成绩较逊色的学生则会感到灰心(Black, 1998)。面对教育改革, 教师应重新判断评估学生习作的根本价值。因此, Baird (1997)提出的评估要点实在值得教师详细考虑。首先, 评估应以改善学生的学习为目标。学生可根据学习的回馈改善学习。第二, 评估与教学应紧密连系起来。教师可透过评估学生习作, 了解教学步骤、活动、课本及其它以学习成果为本的教材是否有效。教师应多根据这些资料来更新教学计划。当教师了解学生的已有知识和尚未掌握的知识后, 便可在教学策略上作出相应改动, 支持学生的学习。Haetrel (1991) 亦强调了教学与评估的密切关系, 指出评估课业应融入日常的课堂活动中。另外, 评估亦可用以了解学生是否已全面掌握知识内容。最后, 评估亦可让教师为班内学生订定实际的评核准则或标准。

总括而言, 香港教育改革提出改变评估方法, 其实与最近的教育研究趋向一致。评估、教学和学习三者互有关连。要评估不同技能的发展及知识理解的程度, 必须利用多元化的评估模式。教师亦应注意使用评分准则或表现标准的重要性。因此, 评估课业的示例亦应能反映上述重点。



另类科学评估模式

为了评估学生是否达到不同学习目标, 教师应采用多元化的评估方法。在决定使用哪些评估形式的问题上, Jones (1998) 提出三个主要考虑因素: 首先, 教师须细心留意所用形式是否配合学生的能力及年龄, 并确保有充足的资源进行有关课业。第二, 所选课业应切合所要量度的学习成果。第三, 评估计划及形式应互相配合, 让学生有机会展示其技能及学习成果。因此, 评估计划不应使用单一的评估形式(如多项选择题)。就算采用多种评估形式, 亦不应只利用一种评估技巧(如书面响应)。评估计划应揉合各种形式, 同时牵涉学生不同的技能。Meng (1990) 进一步探讨有效的评估技巧与学习成果之间的关系。实务技能测试、观察、讨论、晤谈及专题研习, 都可更有效地量度科学能力及解决问题能力的发展进度。Doran, Chan and Tamir (1998) 强调运用科学学习作的「另类」评估模式。另类评估是指有别于传统所采用的形式, 通常要求学生制作、示范或表演, 倾向把焦点放在学生身上, 以学生为本, 而且着重真实性。这些课业形式通常为学生提供机会, 就某些问题找出多种解决方法, 而非从既定的表格中选择「对」或「错」的答案。此外, 另类评估亦可让教师了解学生的智能发展进度(Enger and Yager, 2001)。

Doran, Chan and Tamir (1998) 提出的另类评估形式包括:

口头报告及辩论: 口头报告及辩论提供另类的评估形式, 对于说话能力较书写技巧强的学生尤为适合。口头报告可以个人、二人或多人小组进行。利用这两种形式, 学生有机会就某一特定的科学课题作研究, 并向教师及同学展示研究结果。

组织图是以图画展示认知结构及思考过程的图表, 例如: 概念图及温氏图(Venn diagrams)。

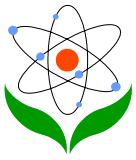
习作选辑集合了不同形式的学生习作, 可包括一般家课、专题报告、实验报告、测验或其它习作样本, 以显示学生的技能发展, 及特定范畴内的知识及技能概念转变。

晤谈及会议: 在晤谈形式的评估过程中, 教师会晤学生时, 应以学生的概念转变、所要纠正的错误观念、及提供数据为目的。在会议期间, 可邀请学生参与自我评估或同侪互评。

技能清单是加强实验室技巧及揉合评估与教学的理想方法。教师、学生及同侪之间都可于进行教学活动时利用该清单。

自我评估、二人评估及同侪互评: 在小组学习的过程中, 教师应请组内每位学生对解决难题作出贡献, 并乐于接受组员的意见, 向组员学习。如学生有机会共同工作, 上述行为亦可推动学生完成评估课业。

由此可见, 有效的评估示例应能如实反映学生的表现, 并配合既定的学习目标。



可信有效的评估方法之条件

无论是哪种评估模式, 所有评估方法都应可信而有效。所谓可信, 是指所取得的评估结果必须准确、稳定, 即以统一的方法评核所有参与评估的学生。Chicago Board of Education (2000)则指出有效的评估, 是就某一特定目的而进行, 所量度的范畴与教师原定要量度的范畴一致。评估效度的几个重要层面, 包括涵盖内容、公平及是否能扩大其应用范围。评估应与学习成果及教师的教学连成一线, 覆盖所有学习成果的重要层面。开展评估方法(Dougherty, 1997; Doran, Chan and Tamir, 1998)及使用评分准则都能有效达到这些评估目的。

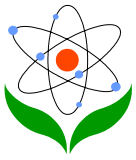
评估课业的开展模式

不断开展评估课业可确保其信度及效度。Doran, Chan and Tamir (1998) 为科学教育工作者简述了开展评估课业的八个非线性步骤, 包括: 列明目的、选取合适的课业形式、写下或改动课业、阐明课业进行的步骤、订立评分准则、测试课业、分析结果、修订课业。上述步骤注重持续的发展过程, 包括学生在测试课业提供的意见, 以及教学与评估之间的关系。Dougherty (1997)在订立发展科学教学的进展评估时, 提出包含参与、发掘、解释、进一步阐释、评估这五个阶段的模式, 强调教师在协助学生建立科学概念时, 可利用这五个阶段进行进展评估。虽然单从学生习作上未能清楚看到这个发展过程, 但这些模式都着重教学与评估的关系, 并应在有效的评估示例中反映出来。

评分准则的运用及发展

一套统一的准则, 应清楚界定及列明各项评估习作表现的准则定义(Enger and Yager, 2001)。Doran et. al. (1998)认为, 评分指引应列明对学生在评估课业表现的评分或评级标准, 而学生在这类课业的参与比重, 远胜选择题形式的习作。有效的评分准则不但能清楚界定习作进行过程的标准, 亦应列明对学科知识的要求。在开展进展评估模式时, Sadler (1998) 要求教师先深入了解所用评估课业的应有准则及标准。教师对学生表现的期望, 应包括对全班学生的一般期望, 以及对活动表现的特定期望, 然后利用评估准则及标准, 使评估课业与预期结果相配合(Erickson, 2001)。这种做法验证了教学与评估的密切关系, 以及评估模式的非线性发展过程。

一套评分准则可以是个有效的沟通工具, 用以清楚订明评核的动机, 并就学生的表现作出回馈(Enger and Yager, 2001)。当教师、学生及家长共同使用时, 评分指引就能把学校所注重的层面具体地表达出来(Chicago Board of Education, 2000)。它不但让教师及其同事弄清优异表现的定义, 将此定义传达给学生, 同时亦能就主观评分的课业, 提供理性的评级标准参考。评分指引应清楚阐明各项评核准则, 好让两位教师按之评核同一份学生习作时, 都会得到同一分数, 以体现评估课业的可信性。这种统一的数据, 对跨班评估及发展学校的整体改善计划相当有帮助。



教师在设计及订立评分准则时, 须考虑几方面的问题。首先, 教师应决定要订立一般或是特定的评分准则。教师可先设计一个整体评分架构, 然后就特定的习作加以修订。例如, 用以评估一份实验报告的评分准则, 也可以用来评核其它报告写作的习作。相反, 个别习作的特定评分指引亦可加以调整, 用来评核其它同一性质的习作。另一个要考虑的因素, 是如何厘订评估等级的分野。Enger & Yager (2001)认为三个等级已足以用来评估学生的表现, 而通常使用的等级记号, 则是 5、3 和 1。如学生习作被评为「不可接受」的类型, 教师便应给予有建设性的意见, 尽可能让学生重交习作。表现优异的学生亦应得到教师的响应或意见, 以确保学生在日后的习作有同样理想的表现。Erickson (2001)指出教师往往把评估的重点放在学生未能达到的标准上, 而不是为达到目标的学生给予鼓励。其实, 评分准则应以推动学生的学习动机、协助他们主动学习为目标。因此, 每一个表现等级所订立的标准均应以正面的字眼来描述, 让学生更积极面对自己的表现, 把自己看成是不断进步, 而非总是达不到要求的学习者。教师评估习作时, 应考虑学生迈进每一个标准期间, 在每个发展阶段所展示的学习能力。此外, 教师如能在学生进行习作前与他们共同制定评估准则, 更可促使学生主动学习。教师可尝试与学生一同想出评估的细节, 就每个表现水平制定具体内容及解释。当学生有机会参与这个过程时, 他们便会更深入了解怎样才能达到预期目标, 以取得较理想的成果。

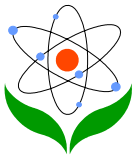
一般来说, 一套有效的评分准则能清楚界定各项表现标准, 作为沟通的工具, 向学生给予回馈, 从而推动学生学习。

总结上述各项分析, 本文所分析的科学评估习作示例均具备三大特点。第一, 以香港的情况来说, 评估课业应配合教育改革的方向, 鼓励发展学生的科学知识及学习技巧。第二, 多采用另类的评估模式。第三, 评估课业应附以评分准则, 达到可信、有效的评估目的。

方法

本文认为教师其实亦曾使用有效的评估课业, 而不是以「亏欠模式」假设教师的教学方法基本上有问题。研究结果由参与「寰宇学校计划」(香港网页: www.ied.edu.hk/cric/saw/; 国际网页: www.edc.org/CCT/saw2000/)的教师提供。这项计划由美国基础教育议会发起, 共有其它九个国家 / 地区参与, 当中包括香港特别行政区。参与计划的教师须提交小四、中二及中四级*的科学学习作样本, 并就每套习作样本提供评分准则及习作背景资料。每套习作包含三份高、中、低水平的样本。每个习作个案均会上载于计划专属网页, 供参与教师讨论。另外, 本文亦旨在提供实证, 探讨不同国家的教师曾采用的有效评估示例, 以引发教师就科学科评估课业的问题作出专业讨论。

*由于研究小组基于数个既定课题收集学生科学学习作, 而各国学生对有关课题的学习程度亦有所不同; 故此, 教师所提交的习作样本亦可能属于小五、中一或中三程度。

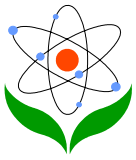


结果及讨论

作者于「寰宇学校计划」选取来自不同国家的六份习作（包括澳洲、美国、葡萄牙、法国），并根据上文提及的四个评估课业层面作出分析。以下将总结习作的评估形式、学习技巧或科学知识、评分准则的运用、以及订立评估课业的过程。表一总括了六份学生习作样本的数据分析。个别习作的详细资料将于下文逐一探讨。

表一 六份习作示例的资料总结

习作名称	级别	应用的科学知识	学习技巧及态度	选取的评估形式	评分准则的运用
动物研究——生态与环境的适应	小四	<ul style="list-style-type: none">动物对环境的适应栖息地	<ul style="list-style-type: none">从书本及互联网搜集资料观察记录绘图分类作出假设爱护动物的态度	<ul style="list-style-type: none">习作资料夹绘图	<ul style="list-style-type: none">事先派发予学生由教师进行评估
消费者研究	中三	进行科学实验的方法	<ul style="list-style-type: none">分析数据撰写报告撰写杂志稿运用信息科技控制变因处理数据	<ul style="list-style-type: none">杂志稿信函以图表组织概念	<ul style="list-style-type: none">只评核分析、表达及运用信息科技的技巧以评语的方式给予回馈由教师进行评估
太阳系	中四	太阳系	<ul style="list-style-type: none">模型制作创作力组织能力	<ul style="list-style-type: none">模型描述报告	<ul style="list-style-type: none">不同比重的评分准则由教师进行评估
接驳电路：串联与并联电路	小四	串联与并联电路	<ul style="list-style-type: none">制作及测试模型展示结果绘图描述能量的转换	<ul style="list-style-type: none">模型绘图	<ul style="list-style-type: none">评分准则以文字描述向学生解释习作要求由教师进行评估
电子的流动	中二	<ul style="list-style-type: none">电子电运用符号	<ul style="list-style-type: none">创意写作绘图空间概念想象力认识能量的转换	简图	<ul style="list-style-type: none">分为四个表现水平绘图及写作分别以两套准则评核预先派发给学生由教师进行评估
绘画配子	中二	配子的结构	<ul style="list-style-type: none">观察绘图反思	绘图	<ul style="list-style-type: none">学生可利用教师的评分准则作自我评估由学生自评及教师进行评估



示例一： 动物研究——生态与环境的适应

级别： 小学四年级

评估形式： 此习作展示了习作数据夹及图像运用的方法。教师要求学生为非洲矮种青蛙、招潮蟹及蜗牛建立及维持自然的栖息环境。学生在绘画温氏图及比较动物的适应能力这类简图时，均考验了他们的认知思考能力。

学习技巧：

透过维持动物的栖息地，可培育学生一些重要的技巧及态度，例如设法保护动物、与人交流经验、并在饲养动物及记录观察日志时，养成努力不懈、自给自足的态度。

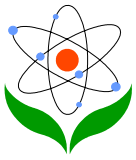
学生可透过写作练习学会记录观察的重要技巧（于问题一及二描述及绘画动物栖息地）、分类（于问题三绘画温氏图、问题四的比较图表）、推论（于问题五把动物栖息地与自然环境联系起来）、以及培养爱护动物的态度（问题六）。

评分准则的运用： 学生在进行习作前获发评分准则，每项习作表现均配合教师对学生的期望。学生有机会重交习作。教师使用评分准则评核习作时，每个等级都包含五个要点，与写作练习第一至五题相对应。教师没有为问题六订立任何评分标准，因为此问题要求学生给予个人意见。

示例一： 动物研究——生态与环境的适应

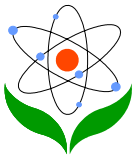
教师给予学生的指示：

学生进行了五星期的动物研究后，以此习作为最终评估。学生须建立及维持非洲矮种青蛙、招潮蟹、蜗牛的栖息地。教师鼓励他们从书本、互联网及百科全书搜集数据，并运用之前所学的笔记记录技巧，定期观察及详细记录有关数据。教师亦以透明胶片向学生解释写作部分的评分准则。



图一：教师提供的评分准则（中文译本）

<u>动物研究评分标准</u>	
A	B
<ul style="list-style-type: none">◆ 确认出生态中需要的四种基本元素。◆ 提出一个动物生态的详细图解并加上标签。◆ 列举出两种被观察的动物的四种相同及四种不同的地方。◆ 确认及描述两种被观察的动物在行为上及结构上的两种适应方法。◆ 提出人类对生态环境影响的一个详细解释。	<ul style="list-style-type: none">◆ 确认出生态中需要的四种基本元素。◆ 提出一个动物生态的图解并加上标签。◆ 列举出两种被观察的动物的三种相同及三种不同的地方。◆ 确认及描述两种被观察的动物在行为上的两种适应及结构上的一种适应方法。◆ 提出人类对生态环境影响的一个解释。
C	
<ul style="list-style-type: none">◆ 确认出生态中需要的二至三种基本元素。◆ 提出一个动物生态的图解，没有或有简单的标签。◆ 列举出两种被观察的动物的两种相同及不同的地方。◆ 确认及描述两种被观察的动物在行为上或(及)结构上一至两种适应方法。◆ 提出人类对生态环境影响的一个简单解释。	



习作样本:

014
USA SC 2601 050 04 S 01
Name/# _____ Date 1/30/01

ANIMAL STUDIES

1. List the 4 basic needs of a habitat.
The 4 basic needs of a habitat are: food, water, space, and shelter.

2. Choose one of the three habitats (frog, crab or snail) that you and your team have been observing and illustrate it below. Be sure to label your diagram. Crab habitat

第一页

014
USA SC 2601 050 04 S 01

3. Using the Venn diagram below, compare and contrast two (2) of your animals. (Frog/Crab/Snail)

第二页

014
USA SC 2601 050 04 S 01

4. List two behavioral adaptations and two structural adaptations for two of your animals (Frog/Crab/Snail). Describe how these adaptations help the animals survive in their environments.

	List	Describe
Animal: Crab Behavioral adaptation	Eats Plankton Burrowes	Eats Plankton to survive burrowes for Shelter
Structural adaptation	Poppy Eyes Male has large claw	Poppy Eyes for seeing Male has large claw for protection
Animal: Frog Behavioral adaptation	Swims Fast Eats	Swims fast for Protection Eats for survival
Structural adaptation	Webbed Feet Long tongue	Webbed feet for swimming Long tongue to catch food

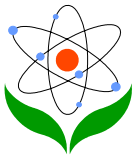
5. What organism has the greatest impact on the environment? Give at least three examples that show the important and unique way that this organism interacts with the environment.
The humans have the greatest impact on the environment. For one thing the pollute. Another thing is they start forest fires. And another reason is because we tear down trees.

第三页

014
USA SC 2601 050 04 S 01

6. What was your favorite animal to observe? Why?
My favorite animal to observe was the crab. It was my favorite because the did neat things. I didn't know that our biggest crab could fit in the tiniest space. That's why my favorite animal was the crab.

第四页



示例二： 消费者研究

级别： 中学三年级

评估形式：此习作是真实评估的例子。一份真实的评估习作应包含真实的任务、难题或专题，当中牵涉学生的判断及创新意念，使学生自己动手完成有关习作，模拟现实世界的情况，包含一定的复杂性，并提供机会改进表现(Wiggins, 1998)。习作让学生体验实地考察的工作。学生须根据既定的科学家研究结果，以记者身份撰写时事通讯形式的分析报告。

学习技巧：学生可学到在科学探究过程中控制变因的重要技巧，培养在分析所提供的证据时抱审慎、怀疑的态度。此习作亦加强了学生处理数据及数据的技巧、写作能力、运用计算机及复杂的分析技巧。

评分准则的运用：学生于低年级时已学习有关科学探究的方法，亦尝试设计及进行两份相似性质的科学探究活动。虽然教师不会与学生一同进行习作，但亦会事前派发列明评分准则的习作指引及背景资料予学生。评分准则并无列明细节，教师在派发习作时会写上评语。

示例二： 消费者研究

图二： 教师提供的习作指引（中文译本）

中三级消费者研究报告习作

背景资料

假设你是「声音」杂志的记者。该杂志定期向消费者提供澳洲市场新产品的报告或数据。

你收到美国"Conway"公司寄来的科学研究资料。该公司刚在澳洲市场上推出一系列的家居化学品及清洁剂。

你的工作是为「声音」杂志分析附上的研究资料，并撰写一份评论报告。

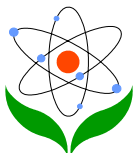
学生必须完成以下各项工作：

1. 为「声音」杂志撰写文章——用简单易明的方式（如图表）去表达所提供的数据。（注：学生最好运用如"Excel"的计算机程序。）
2. 以读者的角度，批判性地评估实验方法。 列举出方法内的存在问题。
3. 就所提供的产品数据，为读者提供意见。（例如产品是否实惠；公司声称的产品效用是否正确。）
4. 写一封详细的信给提供数据的科学家，建议两项改良实验的方法。

学生是否准时递交？ （是 / 否）

评分项目	最高可得分数
实验结果的表达方式（如图 / 表）	3
实验方法的分析（例如找出问题的数量）	5
为读者提供的意见	2
改良实验方法的质素	5
口头报告、表达时的清晰度和原创性	3
计算机的运用	2

总分=_____



背景资料：科学家的实验方法（中文译本）

“Shine”是最新研究出的衣物清洁剂。科学家发现，液体状的“Shine”清洁剂比其它同类产品有高出百分之二十的清洁效能！此清洁剂可以应用在热水或冷水中。它适合予手洗，对衣物和纤手都十分温顺。一点二公升的清洁剂只售价六元五角，极合乎经济原则。

“Shine”与市面上其它一些清洁剂的比较。测试结果如下：

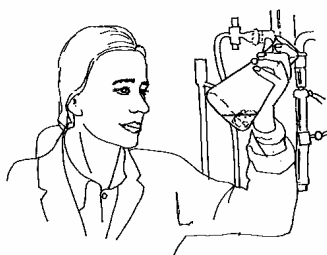
测试一：「清洁效能」



三样名为“Omo”、“Dynamo”、“Surf”的清洁剂与“Shine”作比较。实验中，把一杯清洁剂加入盛载「少量」衣物的洗衣机。经过干水的清洗过程之后，消费评审员比较结果。(本公司向每评审员支付每小时十五元的酬劳。)

产品名称	“Omo”	\$8	750毫升	「干净」
	“Dynamo”	\$9	2公升	「完全地干净」
	“Surf”	\$7.50	1.2公升	「十分干净」
	“Shine”	\$6.50	1.2公升	「惊人地干净」

测试二：「温顺度」



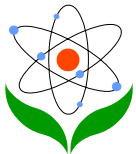
首先测试各清洁剂对衣物的温顺度，然后测试对手部皮肤的温顺度。

Scum太太、Sludge先生、Slime小姐及Sublime女士分别用不同的清洁剂清洗白衬衫。他们每天清洗，直到衬衫有明显的磨损迹象为止。

Scum太太是一位九十岁，很健壮的家庭主妇。她报告说，用“Omo”清洗过的衣服能经历至少一百零五次的循环清洗，衣物的钮扣才会脱落。她载满衣物的洗衣机从来没有坏过，不过她的手出了一点皱纹。

Sludge先生已经独居两年。他光顾市内名叫“Sticky Fingers”的自助洗衣店，用“Surf”每天两次清洗他最喜欢的、夜总会交际用的衬衫。很不幸地，在清洗过只有三十次之后，清洁剂已损坏了「泡泡袖」的针线口。他每天戴着胶手套洗衣服。

Sublime女士使用“Shine”清洁剂。她旧式的洗衣机仍然操作如常，而衬衫到目前为止还没有损坏的迹象。换句话说，衬衫能通过起码一百五十次的清洗考验。她的皮肤甚至逃离了以往牛皮癣的痛苦。



习作样本:

Voice

Volume 7, Issue 6 October 2000

SHINE WITH YOUR CLOTHES

Amazing results guaranteed.
The latest development in clothes washing products.

Written by

I have just received some information from the American company called Conway. They have just recently made their products available on the Australian market.

The new product they have released is called *Shine*. It is the latest development in clothes washing products. Scientific tests have shown that *Shine* cleans 20% better.

I was given two tests that were done on this product. *Shine* and three others products were used in the testing but *Voice* is unable to publish what the other products were called. The first test looks at the washing effectiveness. And the second, at the gentleness shown on clothes and hands.

Four people were asked to each test a detergent. After they tested their product they told me what they

The other side to the story.
Is it all what it really seems?

Perhaps all is not what it seems. If you have already read the main article you would have seen that *Shine* showed some fantastic results.

But if you take a closer look at the method used and the way things were done it doesn't seem so rosy.

In the first test done, there was a lot of varying factors. A capful of each detergent was placed in the washing machine. The caps would have all been a different size. And if the back of the

第一页

Continued from page 1

Cost of Products

Product	Price (\$ per 100 mL)
Shine	0.54
Brand 1	1.06
Brand 2	0.45
Brand 3	0.63

2 was also a good product. It showed that the clothes came up very clean. However the superior standard made by *Shine* can in no way compare.

On the left I have included a graph which shows you a price of each of these products. Brand 1 is by far the most expensive. Brand 2 is slightly cheaper than *Shine*. Nine cents cheaper per hundred mL. This can surely not win over in the factor of which one is better. *Shine's* results are wondrous, and you can not put a price on bright clean clothes.

The next test tested how gentle the products were on clothes and also on hands. *Shine*, Brand 1 and Brand 3 were used in this test.

The tests were done on white shirts and were washed daily until there were obvious signs of damage.

Tester	Product tested	Gentleness on clothes	Gentleness on hands
Mrs. Scum*	Brand 1	Lasted 105 washes until buttons came off.	Hands looked slightly wrinkled.
Mr. Sludge*	Brand 3	Stitching in sleeves wore out after only 30 washes.	Gloves used.
Ms Sublime*	<i>Shine</i>	No signs of damage. Tested right through to end of test. 150 washes.	Psoriasis on hands gone.

Brand 1 was good in gentleness on clothes but seemed to fail on the gentleness on hands. Brand 3 did not give very satisfying results. After only thirty washes it seemed to wear out. Brand 3 was not tested for gentleness on hands. Once again *Shine* showed tremendously great results, lasting up until the end of the testing. It may even hold the cure for Psoriasis!♦

第二页

Shine *Amazing* added to a small load of continued on page 2

As you can see above *Shine* showed some fantastic

Rachael's Recommendations

Shine is said to clean 20% better. And is said to be amazing. It is a great price and is in a good-sized 1.2L bottle. It is not too small that you run out quickly but not too big either.

The shirts that were tested with *Shine* did not show any signs of damage. And *Shine* was gentle on hands. If it really will get rid of Psoriasis, I can't tell you. But I guess it is worth a try.

There did seem to be a lot of errors in the way the method was done though, and I myself wrote away to the scientist who provided me with the information and told him about how I thought it could be improved.

But can we believe what we see? This time I'll let you make up your own mind on that one. ♦

Next Week: A review on the new computer game out. Called *Deddasgart* but known on the American market as *"Dedda"*.

Continued from page 1

washing. Each load would undoubtedly be a different size. No two people would agree on the size of a "small" load.

How long did each cycle go for? How were they dried? Was it hot or cold water used? What sort of machine was used to wash them? Each of these things would contribute to the result given in the end.

And how can four different people describe what the shirts look like. Each person would say something different from the same shirt let alone different ones.

In the second test there was also a lot of errors. White shirts were washed daily until obvious signs of damage were shown. But each detergent tested a different type of white shirt. They all need to be the same. They need to be the same size, age, thickness, brand and have the same sort of stitching. The same type of washing machine needs to be used and the same length of time in the cycle, and when drying.

How can one person wash their shirts twice a day, or wear rubber gloves on a test on your hands? If buttons came off or stitching came undone it would be the shirt not the detergent used which made that happen. And I certainly don't see how someone's favorite night clubbing shirts could all be white.

And with Mrs. Scum*. Of course her hands are going to look "slightly" wrinkled. She's ninety years old!

And lastly how can anyone really tell when signs of wear or damage are showing. Perhaps one of them forgot to look at the stitching or the collar. You can not have four different people looking for signs of damage each person will not see what another sees. ♦

Volume 3

第三页

VOICE
832 Collins St
Melbourne 3000
Australia

October 7, 2000

Dr Richard Harvey
CONWAY
763 Hudson Rd
Phoenix Arizona
USA

Dear Dr Richard Harvey,

I am a journalist from the Australian magazine "Voice". Your product *Shine* has recently been released onto the Australian market. I received some scientific information from your company Conway on the product. I appreciate the information that your company sent to me.

Our magazine reviews products that are on the Australian market. After the review I did on *Shine*, I thought I would write back to you and give you some feedback as to what I thought of the method used. Enclosed with this letter you will find a copy of the article I wrote about the product.

As you can see in the article, I have come across numerous errors. I rewrote what I thought would be an adequate method. You will find it on page two.

I hope you will find it to be helpful, and that you truly do consider using my new and improved method. Thankyou.

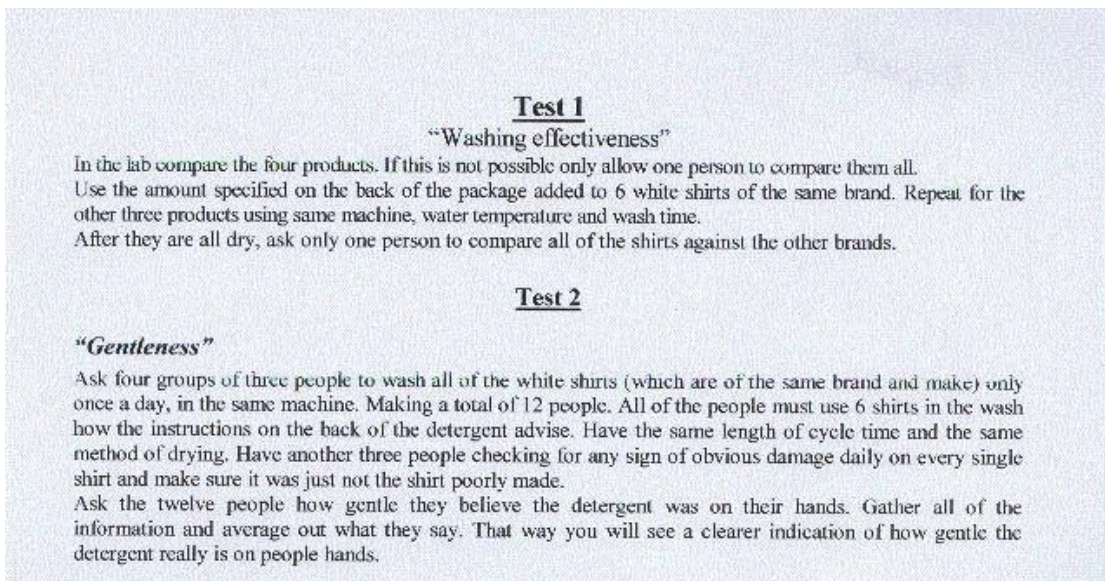
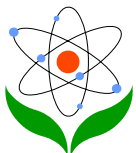
Sincerely yours,

Journalist
Voice
RAR

Please note: In magazine article *Product 1 - "Orno"*
Product 2 - "Dynamo"
Product 3 - "Surf"

Volume 4

第四页



第五页

教师的评估结果：

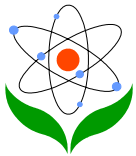
评分项目	教师评语	得分
实验结果的表达方式（如图 / 表）	各方面都表现出色	3/3
实验方法的分析（例如找出问题的数量）	文章内提及实验出现的问题数目在全班前几名	5/5
为读者提供的意见	所提供的改良意见能针对报告内的弊端	2/2
改良实验方法的质素	惟一缺点，是没有描述一部分实验方法	4/5
口头报告、表达时的清晰度和原创性	文字流畅，杂志设计见心思	3/3
计算机的运用	懂得利用 Excel 程序处理文字和图表	2/2

示例三： 太阳系

级别： 中学四年级

评估形式： 此习作是探讨太阳系及地球内部结构的专题研习，主要包括模型制作及报告两种评估形式。

学习技巧： 习作以小组形式进行。教师鼓励学生多表达个人意见，同时亦要尊重别人的意见。教师只提供有限度的指引及监察，以培养学生的责任感。习作鼓励学生发挥创意，训练组织技巧，这两项亦是教师评分的项目。



评分准则的运用：

评分指引所列的各项评分准则，都再细分为不同类别，每项准则所占分数比重不同。虽然教师可为各项准则定下不同评分比重，但如能在学生进行习作前，把习作的重点通知学生，则会更为理想。此习作较着重学生在「运用理论的准确性」表现，所占分数较「报告」一项高出一倍，而习作的「整体表现」及「创造力」，则只占总分 20%。学生知道有关的分数比例，则可适当分配时间，寻找合适的的数据，而不会把时间花在美化模型的外观。不过，使用评分指引来评核学生表现并非一定「万无一失」，因为学生及家长有时并不认同教师的评分。例如他们会质疑「为甚么学生在某项准则上只拿到 12 分，而不是 13 分？」因此，分数的比重只是用来显示不同层面的重要性，而实际的分数其实并不重要。

示例三： 太阳系

教师给予学生的指示：

学生可利用五个月时间来制作模型。习作共分三个阶段：

第一阶段：搜集有关太阳系和地球内部结构的数据，完成一份书面报告，介绍太阳系的模型、描述模型结构及相关现象、列出所用的材料、模型的草图、以及提交下一阶段习作的日期。

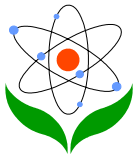
第二阶段：教师与学生商议特定时间，共同讨论制作模型及进行报告时所遇的问题。

第三阶段：学生公开汇报已完成的工作，由教师根据评分纸评估学生的表现。

图三：

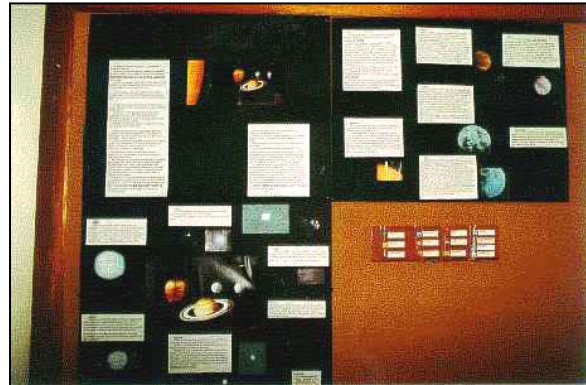
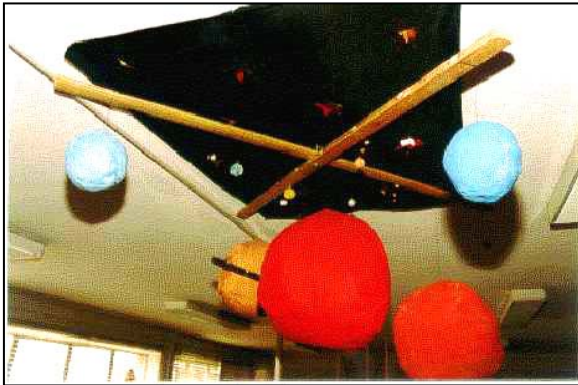
习作样本：





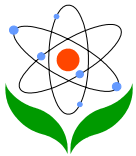
模型放大图：

学生描述太阳系的各个星球：



教师的评估结果（中文译本）：

<u>立体模型——太阳系</u>			评分纸		
准则		最高得分	获得的分数		
缴交日期		3	3		
创造力		15	10		
报告	确认	作者	3	3	
		工作	4	4	
	副标题	是/ 否	5	5	
		科学 准确性	10	10	
	整体表现		5	5	
运用理 论的准确性	模型制作		15	10	
	科学准确性		15	10	
	懂得运用其它颜 色作强调		10	10	
	填写的数据		15	10	
总分		100	80		



示例四： 接驳电路：串联与并联电路

级别： 小学四年级

评估形式： 此习作包含绘图及电路模型制作两部分。学生可利用所提供的材料，自行选择制作串联或并联电路，然后绘画电路图。最后，学生须向教师解释模型所涉及的能量转换。

学习技巧： 透过练习及进行实验，可训练学生以下各种学习技巧：制作及测试模型、展示探究结果、按照实物绘图、描述能量的转变。

评分准则的运用：

学生可多次重交习作，直至达到超出要求的表现。教师事先已派发评分指引，并向全体学生清楚解释对他们的学习期望。

教师订立了一套完整的描述性评分指引，指出及解释习作内三项评估范畴中四个不同层面的要求。总分达到 8-9 分的习作视为超出要求；5-7 分为符合水平；0-4 则未能达到要求。

Enger & Yager (2001) 指出，每项评分标准应只描述一项要求，否则对一些只达到部分要求的表现，则较难作出评分。教师可考虑把三项评估项目细分为不同类别，例如把评核电路图练习的卷标、着色及准确程度分为三个评分细项。

示例四： 接驳电路：串联与并联电路

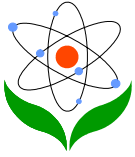
教师给予学生的指示（中文译本）：

学生二人一组，利用所提供的材料合力制作一个串联或并联电路模型，然后个别按模型绘画电路图，并向教师解释模型所涉及的能量转换。教师利用评分指引（参阅习作样本后页）解释对每项工作的要求。

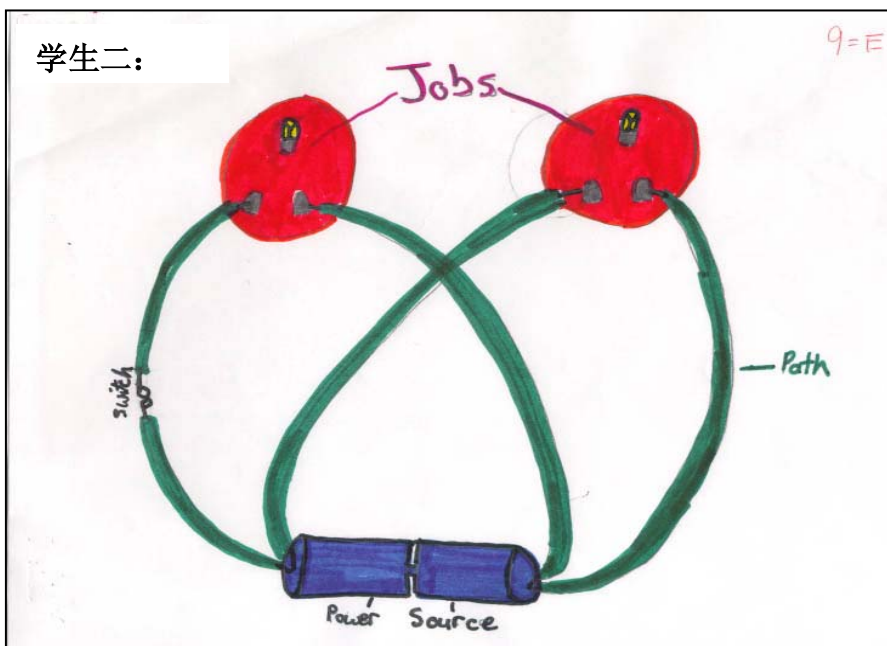
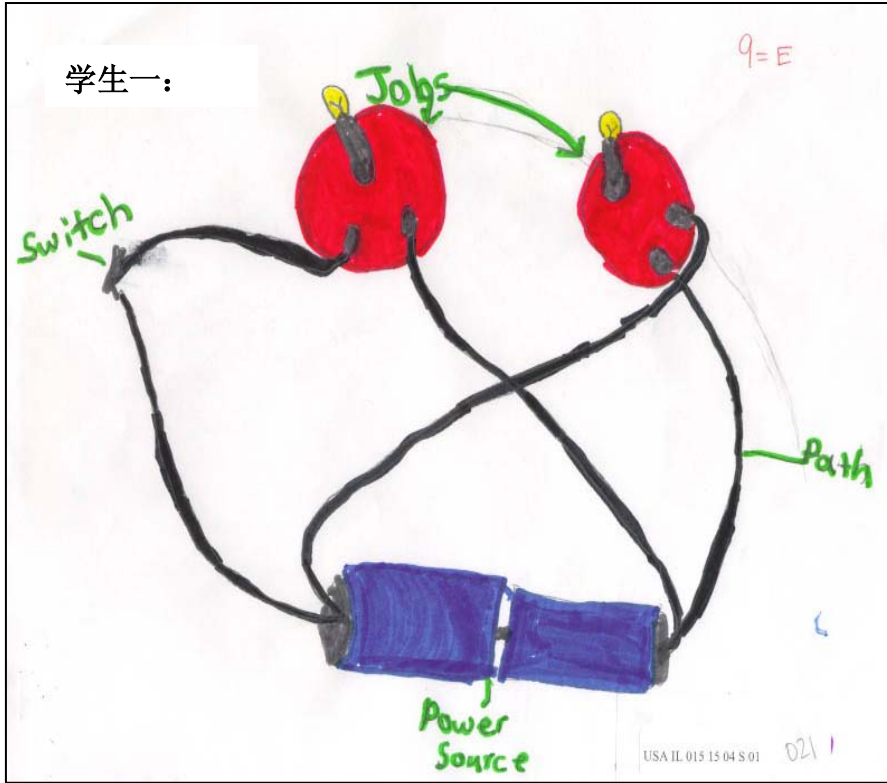
图四：

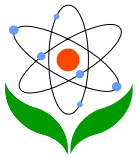
模型样本：





学生根据上述模型绘画的电路图：





教师的评估结果:

电路设计

请圈出你所示范的电路种类

串联电路

并联电路

	0	1	2	3
绘画电路图	没有付出努力或不求甚解。	只有某部份的电路图绘画正确。	电路图包含了电路应有的部份但不能完全正确地绘图。	电路图包含了电路应有的部份及完全正确地绘图。
创作电路模型	没有付出努力或不求甚解。	只有某部份的线路连接正确。	电路模型包含了电路应有的部份及学生能够解释模型的运作程序。	电路模型包含了电路应有的部份及学生能够解释模型的运作原理。
解释能量种类的转变	没有理解。	学生只解释了一种能量的种类。	学生解释了在模型中所有的能量种类。	学生解释了在模型中所有能量种类的转变。

超出要求: 8-9

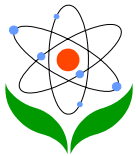
符合要求: 5-7

不符合要求: 0-4

* 学生的串联及并联电路模型一定要达到超出要求的水平, 才能获得甲等成绩。

组员一得分: 9 (超出要求)

组员二得分: 9 (超出要求)



示例五： 电子的流动

级别： 中学二年级

评估形式： 此习作包括绘画电路图及创意写作两部分，前者要求学生绘画评分指引所列的所有元素，后者要求学生运用想象力，假设自己是一粒名叫「哈利」的电子，解释有关电压、电阻器和能量这几个概念之间的关系。

学习技巧： 此习作设计顾及个别学生的多元智慧发展，如创意写作练习能提升逻辑数学思维及语言技巧，而绘画电路图则可训练学生以图像显示空间感（绘画观察图），凭空想象，然后把想象转化为物理符号，找出事物的模式（展示串联及并联电路都包含相同的基本元素），了解元素之间的转化情形（两种电路的电子流动路径）。

评分准则的运用：

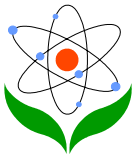
教师分别针对绘画电路图及创意写作两部分订立了两套不同的评分指引。学生须在没有教师协助下完成习作，但教师会于事前派发评分指引，并讲解习作指示。评分指引所列的各项要求，便成为学生习作表现的指标。

其实，教师亦可于每个评分项目加入描述性的准则。创意写作部分的评分是以四种表现水平来评估，包括「优异」、「良好」、「一般」及「没有提交」。虽然教师在评分时就每项准则写下意见，但其实学生最好于事前已得悉每个等级的特定要求，例如教师可考虑在评分指引「讨论所用的能量种类」一栏，列明「优异」及「良好」等级之不同要求。教师可指明拿取「优异」等级的学生须列明最少三种所用的能量，而「良好」等级的学生，则只列出其中两种。

示例五： 电子的流动

教师给予学生的指示（中文译本）：

学生须根据电路模型准确绘画电路图，当中必须包括评分指引所列的各个主要配件。当学生以「电子的流动」创作故事时，可根据评分准则作为指引。教师表明习作重点，是展示对科学概念的理解及创意的发挥。

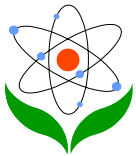


图五:

学生习作样本:

HARRY THE ELECTRON

Harry and his mates are all electrons. At 11:00 Harry and his mates had a rude awakening when Tom connected a series circuit with 2 light globes, a switch, a Voltmeter, a Ammeter and the 12V Power Pack were Harry is sleeping. When Tom turns the Power Pack on Harry and his mates get up and go to the Power Pack Check in desk. This is were Harry gets his energy from. After the Check in desk he enters a torpedo chamber were a chemical reaction takes place and sends him flying around the circuit. "Why go around this circuit" asked one of Harry's friend. "I want to go around this dangerous circuit because I am in love with the Positive side of the power pack and the only way I can get to it is around the circuit," said Harry. As soon as Harry is teleported out of the Power Pack his movement is restricted, but he knows alot worse is to come. These are called by the electron, "those damn charge that get in the way" but are known by every one else as resistors. The first structure he passes through is the copper wire. Harry enjoys this freeway because there is little resistance. Then Harry runs into a funny resistor thing called an ammeter. This counts all the electrons that pass through. When Harry went through the number was 0.2. Harry doesn't mind this because it has little resistance. Both the Wire, Freeway and the Resistor, Ammeter's always went arranged so that Harry and his mates could pass through without many collisions with other atoms. A few things in the hard work through the Mountains of resistors. These Mountains are commonly known as light bulbs. The atoms in these light bulbs hate Harry and his mates so they try their best to annoy him and get in their way. AS Harry comes up to this he can see another light bulb in his way so he thinks to himself he only way I'm going to get past this is to use less energy on each. So Harry charges, on the first globe he has many collisions but makes it through using only 6V to get through. But on the second light globe he uses only 6V again but he finds Voltmeter Gyn. In this gym he was wish to see how much energy he had lost in the light globe. All the energy Harry lost in the light globes come out as heat and light. With the little energy he had left he had a struggle run home. But All of a sudden the switch was opened right in front of Harry. Harry looked at the situation and decided not to jump the gap. If Harry had alot more energy he could jump the gap in the form of a spark. This is where Harry wished he was a bolt of lightning. If he was he could jump any gap. After he had gone around the circuit Harry had a chat with some of his friends and found out that they went through the same experiances he went through. Later in the day the circuit was changed to a rearranged circuit and Harry went around it again. He did the same things except for this time he come across a junction. He decided to go straight ahead but his mate went left so Harry felt lonely. This time Harry had only 1 light globe to go through so he used 12V instead of 6V releasing more light and heat on the light globe.



教师的评估结果 (中文译本):

47/48 A+

中二年级科学科 – 创意写作的评分工作纸

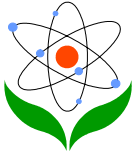
姓名:

电路绘图

准则	在两个电路中的位置及符号都正确	只有一个电路正确	在两个电路中的位置及符号不正确	没有提交
电路的设计	✓			
12 伏特的电池	✓			
开关制	✓			
电灯泡	✓			
安培计	✓			
伏特计	✓			
电压的读数	✓			
电流的读数	✓			
	8			
	× 3	× 2	× 1	× 0
总成绩	24			

创意写作

准则	优异	良好	一般	没有提交
能量来源及电势差的讨论	✓			
讨论所用的能量种类	✓			
电阻及有关结构的讨论	✓			
电流的讨论	✓			
讨论开关掣的运作	✓			
讨论开关掣的限制	✓			
创作意念	✓			
表达方式		✓		
	7	1		
	× 3	× 2	× 1	× 0
总成绩	21	2		



示例六： 观察记录图

级别： 中学二年级

评估形式： 此习作包括两部分：绘画配子的特征及填写自我评估表格。

学习技巧： 学生已学习绘画生物图的技巧，所用的自我评估方法能推动学生反思自己的思考方法，是主动改善学习的重要。

评分准则的运用：

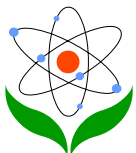
此习作的自我评估表分教师及学生评估两栏。自我评估表是最快最方便的方法，让学生根据既定的准则自行分析习作表现。不过，大部分教师都会担心这种评估方法的可信性，但研究显示，当学生有机会参与评估，而学习目标又在评估前及评估期明清楚界定时，学生普遍都会在自我评估及同侪互评如实反映(Black and Dylan, 1998)。

如教师与学生使用相同的评估标准，学生就能比较自己与教师的评分。为使评估课业更公平、更有效，可在评估表上加上空白栏，以便填写意见。更理想的做法，是让达不到所有准则的学生于限期内重交习作。另一项值得注意的地方，是此评估表以「成功」及「失败」为评分等级，对某些学生来说或会较难接受。教师可考虑以较正面的用字，如「达到要求」及「仍须改进」取而代之。

示例六： 观察记录图

教师给予学生的指示（中文译本）：

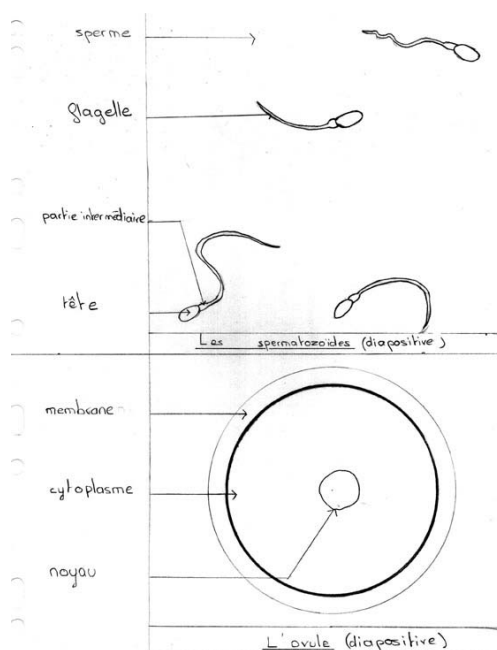
学生须根据以显微镜观察精子和卵子的结果，如实绘画观察记录图。习作由学生个别进行，并可根据以下评分指引绘图。学生可于限期前重交或分段提交已作修改的习作。

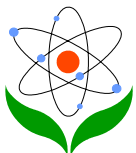


图六： 教师提供的评分指引（中文译本）：

	姓名： 班级：	自我评估		老师评估	
		成功	失败	成功	失败
工作纸的报告	用尺沿标签栏的垂直边界绘画一条直线				
	整齐，用铅笔横行书写				
	小心				
绘图技巧	平均地使用绘图框				
	置中绘图				
	整齐，细致及一笔过绘画				
	整体的外形正确，适当的绘图比例				
	绘画正确的细胞内部结构				
标签栏	用尺绘画一条指示箭咀来加上标签，不可与其它指示箭咀相交				
	指示箭咀的尾部须与适当的标签相连				
	卷标填写在卷标栏内，一个接一个垂直地填上				
	卷标内容须以细楷字母填写				
	正确及完整的标签				
	在图的下方写上准确的标题，并加上底线				
	列出观察时的情况				

习作样本（原文为法文）：





综合分析习作示例

评估准则的运用

透过分析本文的习作示例，在运用评分准则方面，建议教师可考虑以下要点：

应事先派发有关评分准则予学生参考，这可作为沟通工具（例一），亦可作自评之用（例六）；

把学生的表现水平划分为三或四个等级（例五）；

列出不同评分准则的内容，例如：科学概念、表达能力等，并标示各项准则所占的比重（例二及三）；

利用文字详细描述各项准则的要求 / 标准（例四）；

除提供准则外，亦加强对学生的回馈，教师评语对帮助学生自我完善及积极学习方面是十分重要的（例二）。

多元化的评估形式

从表一可见，用作评估学生的形式是可以相当多样化的，这些形式包括：习作数据夹、绘图、杂志稿信函、以图表组织概念、模型、描述、报告及简图等。

除使学生更容易掌握科学概念外，利用不同类型的评估形式，更可进一步发展学生学习科学的技巧及态度，例如：

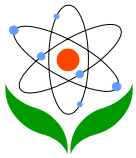
科学过程技巧：观察、记录、绘图、分类、作出假设、分析数据、控制变因、处理数据、制作及测试模型、认识及描述能量的转换等。

一般学习技巧：从书本及互联网搜集数据、撰写报告、撰写杂志稿、运用信息科技、模型制作、创作力、组织能力、展示结果、绘图、创作写作、空间概念、想象力及反思等。

态度：爱护动物的态度。

总结

上述六个示例符合了最近所鼓吹的评估方向，亦展示了各种另类评估形式，包括：习作数据夹、杂志稿、撰写信函、绘图、创意写作、模型制作等，各示例亦提供评分指引，用以交待教师对学生的习作要求和给予学生的习作指示。虽然订立评估工作的过程未能在习作示例中展示出来，但其中一位教师（示例六）



亦尝试加入学生自评部分, 让师生使用同一套评分标准。这种评估方法亦可成为学生学习过程的一部分, 让他们深入了解教师对自己的期望, 同时亦能培养自我评估的技巧。

如考虑香港的实际情况, 这些外国示例其实亦能让本地教师作为借镜。每个示例都评估了一连串的学习技巧, 评分指引的内容与培养这些学习技巧的目标亦属一致。然而, 建立科学知识是学习的一部分, 学习技巧的培养也应建基于此。当香港的教师正寻找测验考试以外的评估模式之际, 上述示例相信有助他们了解如何使用平日较少采用的评估方法。

文中讨论的示例是教师专业发展的实用参考材料。教师可从中了解其它教师的科学教学方法。由于所引例子均为实例, 教师尝试将之引入自己的教学过程时, 定能有更大的信心。纵使这些习作并不能切合特定学校或课堂需要, 然而教师必可举一反三, 按学生的学习需要作出改动。这种鼓励教师作专业发展的取向较为正面, 相信较容易为教师接受。「寰宇学校计划」希望能透过网上讨论习作示例, 让教师互相交流科学评估及教学经验。

参考文献:

Baird, H. (1997). Performance Assessment for Science Teachers. Utah: The Utah State Office of Education.

[Online]. Available: <http://www.usoe.k12.ut.us/curr/science/Perform/past1.htm>

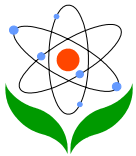
Baron, J. B. (1991) Performance Assessment: Blurring the Edges of Assessment, Curriculum, and Instruction. In G. Kulm, and S. M. Malcom. (Eds.). Science Assessment in the Service of Reform. USA: American Association for the Advancement of Science.

Black P. (1998, December). Formative Assessment: Raising Standards Inside the Classroom, *School Science Review*, **80**(291), 39-46.

Black P. and Dylan W. (1998, October). Inside the Black Box: Raising Standards Through Classroom Assessment, *Phi Delta Kappan*, **80**(2), 139-148.

Bloom, B. S., Englehart, N. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I. Cognitive domain*. New York: McCay.

Chicago Board of Education. (2000). Chicago Public Schools Instructional Intranet Assessment Page. [Online]. Available: <http://intranet.cps.k12.il.us/Assessments>



Curriculum Development Council. (2002). *Basic Education Curriculum Guide: Building on strengths*. Hong Kong: Government Printing Department.

Doran, R., Chan, F. & Tamir, P. (1998). *Science Educator's Guide to Assessment*. Arlington: National Science Teachers Association.

Dougherty, M.J. (1997). Formative Assessment. *Science Teacher*, **64**(6), 29-33.

Enger, S. K. and Yager, R. E. (2001). *Assessing Student Understanding in Science: A Standards-Based K-12 Handbook*. California: Corwin Press, Inc.

Erickson, H. L. (2001). *Stirring the Head, Heart, and Soul: Redefining Curriculum and Instruction*. California: Corwin Press, Inc.

Giuseppe, M.D., Bence, L., Pedrett, E., Goodnough, K., Hodson, D. and Mylchreest, L. (2001). Developing Assessment Tools for Improving the Quality of Student Performance. A presentation at the annual meeting of the National Association for Research in Science Teaching, St. Louis, Missouri, March 2001.

Haetrel, E. H. (1991) Form and Function in Assessing Science Education. In G. Kulm and S. M. Malcom. (Eds.). *Science Assessment in the Service of Reform*. USA: American Association for the Advancement of Science.

Jackie D., Jenny M., Keith V., Margaret H., Doug J., and Cherry A. (1998). *CSF: Making consistent judgments in Science*. Australia: Department of Education, Victoria.

Jones, R. W. (1998). Overcoming an all-too-common problem: Solving the mismatch between science curricula and assessment. *Australian Science Teachers' Journal*, **44**(3).

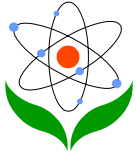
Kulm, G. (1991) The Control of Assessment. In G. Kulm, and S. M. Malcom (Eds.). *Science Assessment in the Service of Reform*. USA: American Association for the Advancement of Science.

Marsh, C. and Morris, P. (1991). *Curriculum Development in East Asia*. London: Falmer.

Meng, E. (1990) What Research Says...About Appropriate Methods of Assessment. *Science and Children*, **28**(1), 42-45.

Raizen, S. A. and Kaser, J. S. (1989) Assessing Science Learning in Elementary School: Why, What, and How? *Phi Delta Kappan*, **70**(9), 718-722.

Sadler, D.R. (1998). Formative Assessment: revisiting the territory. *Assessment in*



Education, 5(1), 77-84.

Torff, B. (1997) *Multiple Intelligence and Assessment*. USA: SkyLight Training and Publishing, Inc.