

多元化课堂教学评量的原理

Principles of Diversified Teaching Evaluation

王淑慧*

(WONG Shwu Huey)

摘要

有教学就必定有课堂教学评量，当前教育改革课题中不可忽视的其中一个议题就是如何在无法脱离传统框架的同时，能更进一步地有效进行课堂教学评量。本文以课堂教学评量概述为起点，进而说明多元化课堂教学评量是以多元智能及建构主义学习理论为理论支撑，再谈多元化课堂教学评量的种类与特点，最后针对当前对多元化课堂教学评量的一些误解做出厘清，为马来西亚华文教育当前课堂教学评量改革，提供可操作的参考原理。

关键词：多元化课堂教学评量、多元评量、课堂教学评量原理、教学评量

Abstract

Classroom teaching necessitates evaluation. One of the issues in the current educational reform is how, within the existing traditional setting of the classroom, to upgrade the evaluation of classroom teaching. This study on teaching evaluation will focus on the following aspects: to state that this method of evaluation is based on the theory of multiple intelligences and the theory of constructivism in learning; to elaborate on the types and characteristics of evaluation; and to clarify some of the misunderstandings and misconceptions surrounding this method of evaluation. The purpose of this discussion is to make available some points of reference on the principles of evaluation in the context of the current attempt to improve the effectiveness of this exercise in the Chinese secondary schools of Malaysia.

Keywords: diversified teaching evaluation, diversified evaluation, Principles of Teaching Evaluation, teaching evaluation

* 王淑慧博士 新纪元学院教育系主任。电邮地址：shwuhuey.wong@newera.edu.my

一、前言

教育改革是当前社会改革不可或缺的热门议题，在不尽相同的改革方案中，或因各自教育哲学的不同而有所差异，然而努力的方向和目标是一致的，就是期盼现行教育能回归教育的本质。其中，了解和掌握教学系统里课程、教学和评量三者之间的关系，对课堂教学评量改革是非常必要的。三者之间，相互影响、唇齿相依、密不可分，其中的课堂教学评量因为直接呈现出教与学的成果成效，所以在不同的改革方案中，课堂教学评量方式的改革已然是必要探讨的议题之一。

在课堂教学评量改革的浪潮中，传统的课堂教学评量方式已经成为众矢之的，批评者将学生在学习上的偏差及教学成效的失败归咎于它，认为学生在学校的学习成效不应只由传统的课堂教学评量来论断，而应考量学生多方面的表现，促进学生的全面发展。

二、课堂教学评量概述

课堂教学评量最基本的功能有两项，即选拔和甄别。然而，随着时代的变迁和社会的发展，教学评量的功能已不能只局限于上述两个方面，而应符合时代的需求发挥其自我教育和导向性的功能。为此，课堂教学评量的多元化在教学评量改革中是非常重要的一个面向。

（一）评量与课堂教学评量

要了解课堂教学评量，首先要明白什么是评量（evaluation），不同国家和地区，对“evaluation”这个名词有不同翻译，中国大陆译为“评价”、台湾地区译为“评量”、马来西亚则以“评估”称之，实则要表达的意涵是大同小异。评量指的是有系统、有步骤地从数量上测量或从性质上描述学生学习的过程与结果，据此判定是否达到了所期望的教育目标。（日本筑波大学教育学研究会 2003：409）检视过去和现在的文献，评量的发展与演进可分为三个阶段，在不同的阶段，使用的名词也有所不同。美国在“八年研究”（The Eight-Year Study）时代之前，强调的是“measurement”以量化的方法取得正确可靠的数据；到后来则认为应该从教育目标、人格的发展等各方面来进行评量，亦即除了客观的数据之外，尚须有一些价值标准来加以衡量，而将“measurement”提升至“evaluation”；近年来，学者们又将“evaluation”再提升至“assessment”，强调评量时应考虑相关的整体情境，从各种可行的途径，搜集全面性、多元化的资料，再从各个角度和不同的观点加以比较分析和综合分析，进行整合性的诠释，以获得全面充分的了解。（李坤崇 2006：1）

Miller和Linn认为评量最终要回答的问题是：“这个学生的表现如何？”，在描述过程中要兼顾量化和质化，其中量化描述指的是测量（measurement），而质化描述则是非测量。测量是获得量化的过程，用以呈现学生的学习程度，测量可以解答“这个学生的表现得多少分？”，其重点在于量化。测验（test）则意指某种特定的评量工具或系统程序，正如一套试题，学生在指定范围时间内完成测试，重点在于个人与群体

之间的比较，因此，测验可以解答的问题是“这个学生的个人或与其他人比较的表现有多好？”。（Miller & Linn, 2009）简而言之，评量、测量与测验是在研讨课堂教学评量中常使用的三个基本概念，厘清三者之间的关系，将有助于理解课堂教学评量的意义。

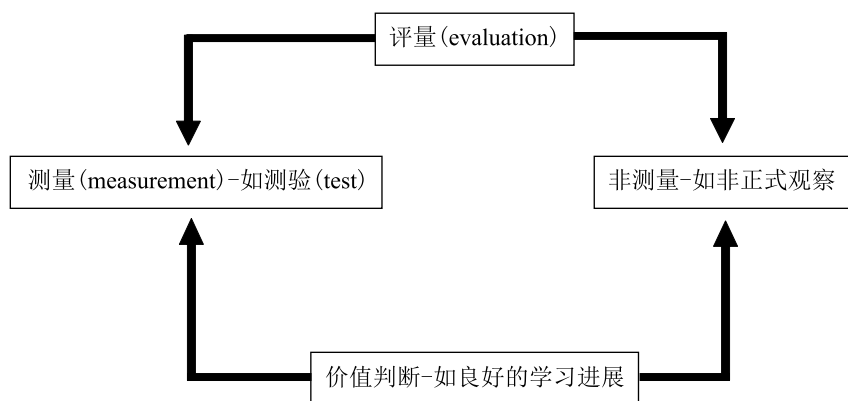


图1：评量过程（Miller & Linn 2009：9）

美国教学评量学者柯布勒（R. J. Kibler）在其1974年出版的《教学目标与评价》中提出，教学的过程应分为教学目标、学前评估、教学活动、学后评量等四个部分。（蔡斌祥 1997：68）

中国教育学者王道俊和郭文安在《教育学》一书中指出教学评量是对教学工作所作的测量、分析和评定。它以参与教学活动的教师、学生、教学目标、教学内容、教学方法、教学设备、教学场地和时间等因素的有机组合的过程和结果为评量对象，是对教学活动的整体功能所作的评量。

依据评量在教学过程中的不同作用，有以下三种类别：

- 1、诊断性评量：在学期教学开始或单元教学开始前，对学生现有的知识水平、能力发展的评量。
- 2、形成性评量：在教学历程中对学生的知识掌握和能力发展的比较经常而及时的测评与反馈。
- 3、总结性评量：在一个大的学习阶段，如一个学期或一门学科终结时，对学生学习的成果进行的较正规的、制度化的考查、考试及其成绩的全面评定，也称为终结性评量。其目的是给学生评定成绩。（王道俊、郭文安 2009：269）

Airasian认为，课堂教学评量（assessment in teaching）是指教师将所得的信息数据加以选择、组织和分析后对学生做决定或价值判断的过程，信息乃教师在课堂上收集到的种种量或质的信息。（李坤崇 2010：1）

中国学者覃兵则认为，课堂教学评量的定义可分为狭义和广义两个方面。狭义的课堂教学评量是指发生在课堂中的语言点评活动，就是教师在课堂中对学生的各种表现所做出的评量性语言。广义上的评量是对学生的学习情况、参与教学活动情况的了解、总结和反馈。（覃兵 2010：5）

由此观之，课堂教学评量绝不是教学的终点或结束，反之，课堂教学评量的主要目的在于分析教学过程的利弊得失，并藉以诊断教与学的难处，以作为后续个别辅助或实施补救教学之用。

（二）课堂教学评量的功能

课堂教学评量的主要目的在于了解教学过程中教师教学与学生学习的效果，在教学过程中有着重要的功能，主要的功能有以下几个方面：

1、检视教学目标。测量并诊断教学目标，是课堂教学评量其中一项重要的功能，教学目标是否达成？教师的教学是否依循教学目标而展开？学生是否掌握了预定的认知、情意与技能？这些问题都得通过课堂教学评量加以检视。教师可运用不同的评量工具，分析学生的学习成效，以作为后续教学活动设计的依据，从而改进教学状况，以达学前所预定的教学目标。

2、了解教学问题。从课堂教学评量的结果可探勘教学成效，教师可从中了解自己教学目标的设定是否正确、教学设计与方法是否妥当，同时了解和发现学生学习的难题，以调整教学的策略和措施，改进学生的学习情况。

3、掌握反馈信息。课堂教学评量的结果可同时反映学生学习、教师教学的优缺点，就教师而言，可了解学生学习的成就或潜能，亦可掌握学生学习的问题与难点；就学生而言，可通过课堂教学评量更清楚地了解自己学习情况的优劣好坏，激发学习的动机。

4、调整教学过程。课堂教学评量的其中一项重要功能就是对教学进程的调整，依据评量的结果，在教学效能、教学问题和反馈信息等基础上，针对教学目标、教学内容、教学方法、教学进度、教学情境等进行调整。借由教学过程的调整，使教学朝向预定的教学目标迈进，亦是课堂教学评量的重要功能。

三、多元化课堂教学评量的理论依据

自上世纪90年代以来，多元智能理论和建构主义学习理论在东西方各国被广泛地运用于教育领域，成为多个教育研究领域的重要理论依据。无论是多元智能或建构主义学习理论都强调评量学生应对学生的整体性、全面性的发展空间投入更多的关注。除了多元智能理论和建构主义学习理论，布鲁姆教学目标分类理论亦是探讨多元化课堂教学评量不可忽视的理论，若不清楚教学目标或教学目标含糊，则课堂评量难以进行有效地测量。因此，课堂教学评量必须对上述理论有较详尽的了解，并以此作为多元化课堂教学评量的理论基础。

（一）多元智能理论

1983年出版的《智能的架构：多元智能理论》（Frames of Mind: The Theory of Multiple Intelligences）一书中，加德纳将智能定义为：“解决问题的能力或是在各个文化背景中创作该文化所重视的作品的的能力。”在此定义中，加德纳强化了对人类认知的跨文化观点。因为这些智能就像是所有人都在使用的语言，并且有一部分是受

到个人所处文化背景的影响。它们是人类用来学习、解决问题以及创造的工具，是每个人都会使用的工具。2008年出版的《再构多元智能》（*Interintelligence Reframed*）一书中，加德纳又对智能的概念作了更进一步的诠释：“智能是一种处理信息的生理心理潜能，这种潜在某种文化环境之下，会被引发去解决问题或是创作该文化所重视的作品。”（Linda Campbell等 1998:4）现将加德纳提出的八种智能简介如下：

1、语文智能（*linguistic intelligence*）：包含用语言思考、用语言表达和欣赏语言深奥意义的能力。作家、诗人、记者、演说家、新闻播报员都展现高度的语文智能。

2、逻辑—数学智能（*logical-mathematical intelligence*）：使人能够计算、量化及考虑命题和假设，而能够进行复杂的数学运算。科学家、数学家、会计师、工程师和计算机程序设计师都展现很强的逻辑—数学智能。

3、空间智能（*spatial intelligence*）：让人有能力以三度空间的方式来思考，就如航海家、飞行员、雕塑家、画家和建筑师所表现的一样。这种智能使人能知觉到外在和内在的影像，也能重视、转变或修饰心像，而且不但自己可以在空间中从容的游走，还可以随心所欲地操弄对象的位置更能产生或解读图形的信息。

4、肢体—动觉智能（*bodily-kinesthetic intelligence*）：使人能巧妙处理物体和调整身体的技能。运动员、舞者、外科医生和手工艺术者都是例证。在西方社会，虽然动作技能不像认知技能那般受重视，但是对于许多令人尊崇的角色而言，善于支配自己身体的能力是他们赖以成功不可或缺的条件。

5、音乐智能（*musical intelligence*）：例如那些对音准、旋律、节奏和音质等很敏锐的人。具有这种智能的人包括作曲家、指挥、乐师、乐评人、制造乐器者，还有善于感知的听众。

6、人际智能（*interpersonal intelligence*）：就是能够善解人意，与人有效交往的才能。成功的教师、社会工作者、推动者或政治家就是最佳的例证。正因为近来西方文化已经开始认识心智与身体间的连结，所以也必然开始重视精通人际的行为的价值。

7、内省智能（*intrapersonal intelligence*）：是有关如何建构正确自我知觉的能力，并能善用这些知识来计划和导引自己的人生。确实有些人拥有高度的内省智慧，特别是神学家、心理学家和哲学家就是最佳的例证。

8、自然观察智能（*naturalist intelligence*）：包括观察自然界中的各种型态，辨认并分类物体，且能洞悉自然的或人造的系统。学有专精的自然观察者包括农夫、植物学家、猎人、生态学家和庭园设计师。

依据加德纳的解释，尽管他相信这八种智能比以往的一元论（智力，IQ）对人类的才能提供了一个更正确的面貌，但智能不应只局限于这八种。他同时强调每种智能还有次智能，例如，在语文领域中的次智能就包括演讲、朗诵、写作、编辑等。其他七种智能也都各自包含多种成份。

加德纳相信，由于每种智能都可以用来行善或作恶，这八种智能本质上都是价值中立的。例如德国的戈培尔（Joseph Goebbels）和印度的“圣雄”甘地（Mahatma Gandhi），他们都有很高的人际智能，但他们的行为方式却戏剧性地南辕北辙。戈培尔是希特勒统治下德意志的宣传部长，是使德国人盲从纳粹政权的始作俑者；甘地则是印度民族主义领袖，20世纪非暴力主义的倡导者。

由此可见，通过不同的智能都能表现出不同的创意。然而，正如加德纳指出，大部分人只在某个特定领域显现创意。例如，爱因斯坦（Albert Einstein）虽然在数学和科学方面天赋禀异，但在语文、肢体一动觉和人际方面则未显示与他人同等的过人天分。也就是说，一般人其实只能在一、两种智能上超凡出众。（Linda Campbell 等 1998：4）

后来加德纳又陆续提出神灵智能（spiritual intelligence）、存在智能（existential intelligence）和道德智能（moral intelligence）等其他智能的可能性。然而，多元智能的理论不在于研究人类究竟可能有哪几种智能，而在于提醒我们用宏观的视野看待所有的智能，尤其是教育工作者更应该提供多元的渠道，让具有不同智能的学习者各展所长，不以传统的一元论来论断学习者的成与败。如此，才能符合多元智能教学的精神——超越或者透过多元智能（through Multiple Intelligences）来教学，而不只是教授多元智能的观念（for Multiple Intelligences）。

教师在进行多元化课堂教学评量时，可参照David Lazear基于多元智能理论的多元评量方式，因地制宜，取精扬弘：

表1：多元智能评量清单（David Lazear 2000：117）

语文智能 （以语文技巧为基础的评价工具） 书面论文 字汇问答比赛 文讯息的回忆 录制音记录 诗词创作 语文式的幽默 正式演说 认知性辩论 倾听与报告 学习日志与日记	逻辑—数学智能 （以认知型态为基础的评价工具） 认知组体 高层次的推理 型态的游戏 摘述要点 逻辑与推理游戏 心灵清单与公式 演绎推理 归纳推理 计算过程 逻辑分析与批判	空间智能 （以意象为基础的评价工具） 壁画和混合画 图象表征与视觉图解可视化与想象力 阅读、理解与制作地图 流程图与图表 雕塑与建筑 想象的对话 网络化 录像与照相 操作示范	肢体一动觉智能 （以实作为基础的评价工具） 研究室的实验 戏剧化 原创与古典的舞蹈 比手画脚和模仿 扮演 活人画 发明的计划 体操与游戏 技巧示范 透过肢体语言与姿势来解说
音乐智能 （以声音为基础的评价工具） 创作概念歌曲与饶舌歌 用声音解说 辨认旋律的型态 谱曲 用概念把音乐和节奏链接起来 编管弦乐曲 创作打击乐型态 认识音调的型态与音质 分析音乐的架构 复制音乐与节奏的型态	人际智能 （以关系为基础的评价工具） 小组分工合作 向别人解说或教导别人 思考—配对—分享 接力赛 提供和接受回馈 访谈、问卷和调查 同理心的历程 随机的小组问答 评价你的队友 测验、训练、再测验	内省智能 （以内心历程为基础的评价工具） 自传式的报告 个人应用的脚本 后设认知的调查与问卷 较高层次的问答 专注力测验 心情日记与日志 个人的投射 自我认同的报告 个人历史的关联 个人的偏爱与目标	自然观察智能 （以环境为基础的评价工具） 动手做的实验或示范种族或自然型态的分类 与大自然邂逅及田野旅行 环境的回馈 大自然的观察 照顾植物和动物 感官刺激运动 力行自然保育 典型的型态认知 自然世界的模仿

多元智能理论强调智能的多元，教学面向的正是有着个别差异的学习者，因此，多元智能理论为课堂评量的多元化提供了强而有力的理论依据与支架。

（二）建构主义学习理论

建构主义（constructivism）学习理论建立于认知学派有意义学习的基础上，强调学生在理解信息方面的积极角色；是学生试图了解周围环境时构建知识的一种学习方法；是学生积极参与集体阐释问题、解释现象、研究复杂问题或解决问题的一种学习模式。建构主义的观点源于众多教育家，包括约翰·杜威（John Dewey）、琼·皮亚杰（Jean Piaget）、马利亚·蒙台梭利（Maria Montessori）、列夫·维果茨基（Lev Vygotsky）等。（Donald R. Cruickshank等 2007：280）

建构主义主张，知识的学习不是通过教师传授所得，而是学习者在一定的情境下借助学习有意义地建构获得的。因此，知识的建构是基于现实生活中的运用，学生应该是主动的建构者，而非被动的信息记录者。（崔永华 2008:17）

为了使学生在学习活动中能尽可能地理解知识而后加以应用，在建构主义者看来，学习活动应该具有一些特定的要求，学者们融合不同学派的观点并将其加以统整，归纳如下：

- 1、主动学习（学生直接参与独立发现答案的活动）比被动学习（学生被动接受教师所给的信息）更好。
- 2、学生应当从事“真实情境”活动，即他们面临的任务应当是具体的，而不是抽象的，是真实的，而不是象征性的。
- 3、学习活动应当有趣且富有挑战。
- 4、学生应当将新知识和他们已有的知识通过一些沟通方法联系起来。
- 5、学生应当反思正在学习的知识。
- 6、在学生团体中，即在小组或社会情景中学习，学习效果最好。
- 7、教师不是把信息提供给学生，而是侧面引导，促使学生获得信息。
- 8、教师必须向学生提供取得进步所需的帮助。（Donald R. Cruickshank等 2007：280）

由此观之，在学生学习的过程中，建构主义强调以学生为中心，承认学生的个别差异是必然存在的，并强调合作学习的重要性。

（三）布鲁姆教学目标分类理论

美国著名教育学者布鲁姆（B.S. Bloom）等人认为，完整的教学分类应该包括认知、情感和技能三个基本领域。每个领域的目标又可分为不同的层次结构，每个层次有包含数个类别。

1、认知领域（cognitive domain）

认知教学目标是指学生学习的结果，布鲁姆等人认为，学生学习的结果可以由学生行为的改变与否来表示，因而于1956年设计了教学目标分类学（Taxonomy of Education Objectives），将整个教学过程中所有预期的学习结果，以类似动物的科学分类方式予以系统分类。这有助于课程的发展、教学活动的进行及课堂教学评量的实施。（王财

印等 2012: 93) 布鲁姆等人将认知学习由低向高发展分成六个层级: 知识 (knowledge)、理解 (comprehension)、应用 (application)、分析 (analysis)、综合 (synthesis)、评量 (evaluation)。

2、情意领域 (affective domain)

情意是对外界刺激肯定或否定的心理反应, 如喜欢、厌恶、态度、价值和信念都归属这个领域, 学生的情意会影响他在行为上的选择与形成一定的态度, 提高鉴赏能力, 更新价值观念等方面的表现。D. R. Krathwohl等人于1969年提出情意目标, 并将之分为五个简单到复杂的层次: 接受或注意 (receiving or attending)、反应 (responding)、价值评定 (valuing)、组织 (organization) 和形成品德 (characterization)。(王财印等 2012: 95)

3、技能领域 (psychomotor domain)

布鲁姆主持的委员会没有建立技能领域的分类, 直到1972年才由Elizabeth Jane Simpson及Anita Joyce Harrow建立技能领域的目标分类。目前, Simpson的技能领域分类较被广泛的使用。Simpson以认知心理学为理论基础, 建构技能领域的目标分类模式和理论, 将技能领域区分为七个主要阶层: 知觉 (perception)、趋向 (set)、在引导下的反应 (guided response)、机械化 (mechanism)、复杂的外显反应 (complex overt response)、适应 (adaptation)、创作 (origination)。(王财印等 2012: 97)

布鲁姆等人认为, 教学目标的主要功能是为了协助学习或教学内容和行为的选择。

综上所述, 基于多元智能理论、建构主义学习理论和布鲁姆教学目标分类理论的多元化课堂教学评量, 无论是在评量的专业、内涵、过程、情境、方式、结果等方面都应以“多元化”为出发点, 开展有利学生全面发展的课堂教学评量。在评量过程, 学生不再是消极的被动者, 而应转化为积极主动的参与者; 教师亦不再只扮演主导者, 更应成为指引学生参与评量的引导者或中介者。简而言之, 多元化课堂教学评量为现有教学评量体系之革新提供了新的视角。

四、多元化课堂教学评量的种类与特点

现今的课堂教学评量无可避免的仍以传统的纸笔测验为主, 纸笔测验虽然能在认知领域呈现量化的评量结果, 然而在情意和技能领域评量结果的呈现则非常有限, 例如图形概念、实验口头报告、演讲、想法的组织、小组合作学习、校园活动设计等方面的能力就难以以纸笔测验的方式进行评量。所以, 课堂教学评量应兼顾认知、情意和技能三个层面的结果, 教师宜朝着多元化课堂教学评量发展, 方能更適切地评量学生在各项学习上的表现。

多元化教学评量, 顾名思义就是在评量学生学习表现时, 使用两种或两种以上方式, 而不局限于单一的评量方式, 它包括实作评量、档案评量、口头评量等多种课堂评量的方式, 较能顾及认知、情意和技能之学习结果。

(一) 多元化课堂教学评量的种类

多元化课堂教学评量能使课堂教学评量更趋于多元化、人性化和弹性化等方向发展。以下就多元化课堂教学评量的类型: 实作评量 (Performance Assessment)、档案

评量（Portfolios Assessment）、变通性评量（Alternative Assessment）、高阶思维认知评量、口语评量、轶事记录等，说明多元化课堂教学评量的真实内涵。

1、实作评量（Performance Assessment）：善于教学的教师会经常采用实际操作、观察、实验、写报告、口头发问等方式来判断学生的学习情形。由此观之，实作评量其实不是课堂教学评量改革的一种新策略，它的主要用意在于以多种评量方式来评量学生不同的能力，从中探视学生对所学知识的应用，而非仅仅展示记忆中的知识本身。

Airasian主张实作评量是要求学生完成一项活动，或制作一个作品以验证他们所学习的知识和技能，所以实作评量主要是让学生在与现实相似的情境去表现他们的所知和所能。Airasian对于实作评量与纸笔测验做出的比较，有助于更清晰地了解实作评量的功能：

表2：实作评量与纸笔测验的比较（陈湘媛 1997：10）

实作评量	纸笔测验
1. 涉及学生把知识转化成为可观察的表现行为或成品的能力。	1. 主要涉及学生的知识及信息的获得。
2. 设计及实施测验费时，但评量表可针对同一或新的学生重复施测。	2. 设计费时，但可同时施测许多学生，同组学生仅能使用一次。
3. 学生表现不佳，可予诊断及补救，可监控学生学习实况。	3. 除论文方式或开放式数学题外，甚少提供方向或指示如何改进表现。
4. 教学首重表现及过程。	4. 教学首重内容知识。

Stiggins则强调实作评量的目的在于评量将知识和理解化为真实行动的能力，他认为学生应学会善用知识和技能，所以主张让学生经由计划、建构及表达原始反应来评量学生的学习结果，他还认为凡以观察和专业判断来评量学生学习成就的教学评量方式均可称作实作评量，具体的体现在于教学评量形式的多元化。（李坤崇 2006：216）Stiggins亦对以下几个类型的课堂教学评量作比较，探视不同类型评量之间的差异：

表3：各种课堂教学评量类型的比较（李坤崇 2006：217）

项目	客观式测验	论文式测验	口头发问	实作评量
目的	评量具最有效率及信度的知识样本	评量思考技能或知识结构的精熟度	评量教学中的知识	评量知识、理解化为行动的能力
学生的反应	阅读、评量、选择	组织、写作	口头回答	计划、建构、表达原始反应
主要优点	有效率：能在同一测验时间内测量许多测验题目	可评量较复杂的认知结果	将评量与教学结合	提供实作技能的充分证据
对学习的影响	过度强调记忆，若妥善编制，可测量思考技能	鼓励思考及写作技能的发展	刺激学生参与学习，教师提供教学成效的立即回馈	强调在相关背景下，善用有用的技能与知识

彭森明为实作评量作了以下几项注释：（1）要求学生自做答案，而非选择答案，对某一问题的正确答案可能不只一个；（2）实作评量不单看成果（答案或作品），也应考虑学习的历程；（3）实作评量的内容，不局限于单科知识，而是同时考虑多项知识和技能，以及报告、实作等能力；（4）实作评量问题的取材应与实际生活问题有关，以促进运用所学知识去解决与真实生活有关的问题，以达到“学以致用”之效果；（5）评量时间有弹性，有些评量可要求学生于一定的时间内完成，如写作文。有些评量则让学生自定义和支配完成的时间，如在一学期之内完成报告；（6）实作评量是教学的一部分，它可促进教师的教学与学生的学习。（陈湘媛 1997：9）从以上的论述可得知，实作评量可弥补传统评量的不足，提升教与学和实际生活的联系，增强学生高阶思维与能力，使教师更深层思考课堂教学评量的真正用意，使学生更主动的学习。

陈英豪、吴裕益、余民宁主张实作评量是介于评量认知能力的纸笔测验和将学习结果应用于真实情境这两种呈现方式之间的评量方法，既是模拟不同的真实情境，提供教师有系统地评量学生在实作方面的表现的一种方式。

李坤崇在综合中外学者的观点后，将实作评量定义为：具相当评量专业素养的教师，编拟与学习结果应用情境颇类似的模拟测验情境，让学生表现所知、所能的学习结果。

2、档案评量（Portfolio Assessment）：是对“portfolio assessment”一词的译名，有者译为“卷宗评量”，“案卷评量”、“档案评量”、“历程档案评量”、“学生学习成果档案评量”、“作品集项评量”等之译法亦大有人在。本文选择以“档案评量”作为论述之用。

档案评量的定义众说纷云，庄明贞认为档案评量的主旨是在有目的、有计划地搜集与组织学生的作品，以呈现作品的质量和进步情形。简茂发主张档案评量是从学生所学的众多项目中，找出重要的评量项目，做为适切的组合，进行有系统的评量，而非零碎地进行各个项目评量。詹宝菁强调档案评量是在一段时间内，以个别学生为单位，有目的地从不同的角度和层次收集学生在学习上的参与、努力、进步和成就的一种检验方法，从中有组织地汇集，再经由师生之间的合作，根据评分标准评量学生的表现。

Airasian指出档案评量是一种收集学生的表现或作品的教学评量方式，它要求学生汇集一系列的表现或作品以展现他们的学习能力或进步的情形，档案内容应包含足以代表学生朝向课程目标前进和发展的作业，因此，档案评量是教学与学习相结合的评量，若教学与学习无任何关系，则档案评量将不具任何意义。

Stecher和Herman则认为档案评量是学生持续在一个时段内，收集各式各样的代表性学习成果，以呈现其学习的情形，而非突发地简单地收集资料。所以，档案评量具有以下四项特质：（1）累积，一段长时间的收集与扩充；（2）深耕，是从计划性、规划性教学衍生的成果，而非临时要求的任务；（3）自我选择，学生能从某些作品中作出自己的选择；（4）反馈，学生自我评量所选择的任务、质量和成果。（李坤崇 2006：302-303）综上所述，李坤崇认为档案评量的定义为：教师依据教学目标与计

划，请学生依据特定的目的持续一段时间主动且有系统的收集、组织与反思反映学习成果的档案，用以评量其在学习上的努力、进步、成长的情形。（李坤崇 2006：304）

3、变通性评量（Alternative Assessment）：变通性评量又称为“另类评量”，是一种与传统课堂教学评量有所不同的评量方式，大体上是以实作做为评量的内容。变通性评量深受认知学习理论的影响，强调知识是建构的，学习是个体从新知识和原先的知识建构有意义的学习。简而言之，知识的学习是一种生活技能的应用，因此，变通性评量重视知识的实性应用和自我评量标准。这是Herman于1992年针对变通性评量提出的观点。

庄明贞则主张变通性评量应看重课程本位的能力评量（Curricularbased Competency Testing）及筑建不同社会文化常模的建立。她将变通性评量的内涵和发展趋势做了归纳和整理，如下表所示：

表4：变通性评量的内涵与发展趋势（陈湘媛 1997：8）

变通性评量的内涵	变通性评量的发展趋势
1. 学生需要在学习活动中表现、制造、产生某些行为。	1. 从行为主义转变为认知学习理论的学习与评量观点。
2. 学生必须能完成有意义的教学活动作业。	2. 从传统的纸笔测验转化为有意义、真实性评量（Authentic Assessment）。
3. 学生能将所学做真实生活的应用。	3. 从单次评量转化为多次作品集合式的评量。
4. 评分必须以人的判断为主，而非着重于机械部分的使用。	4. 从单一评量转化为多元化评量。
5. 教师在课堂教学评量的角色必须要调整。	5. 从强调个别式评量转化成群体式评量。

庄明贞同时认为变通性评量需要教师与学生的互动，教师客观地观察并记录才能真实地评量学生的学习历程和学习成果，意即评量学习的过程与结果是同等重要的。下图有助于了解实施变通性评量的流程：

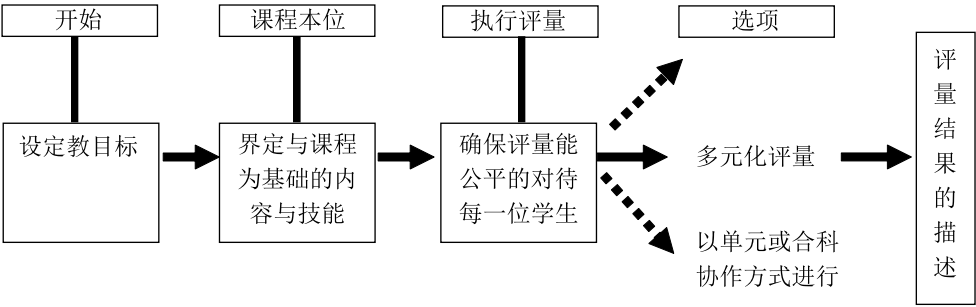


图2：变通性评量的流程（综合李坤崇 2006：218、陈湘媛 1997：7）

综观以上的论述，变通性评量在本质上是一种课堂教学评量改革，旨在改善修正传统教学评量的缺失，促使教学与评量能更紧密的联系，与多元化课堂教学评量的本质要求更相符，因此可将之视为课堂教学评量改革中一种必要参考的方式。

4、高层次思维认知评量：王为国认为高层次思维是人类认知能力中较上层、较精致的能力，它涉及思维中反思（reflecting）的历程，通过这段历程，人类可运用策略进行更有效的学习、可创造新的知识、有能力选择可信的知识，以及能更有效地解决问题。

Craver提出高层次思维的特质有：（1）非线性，不一定要遵循旧有的实施方式；（2）多面向，不能以单一观点来理解思考的过程；（3）产生不只一种解决问题的方法，而且每个方法都各有其优缺点；（4）所得的答案并非只有黑或白，应带有灰色地带；（5）运用多元标准，可能和其他观点相冲突；（6）包含不确定的要素。（王为国 2006：167）

Lazear于2004年提出认知领域的层次应包含搜集及理解基本知识、信息分析与处理、高层次思维与推理层次等。（王为国 2006：168）由此观之，课堂教学评量不应只停留在知识的搜集与理解的基本认知层次，应让学生的思维层次往上提升至信息的分析与处理，进而达致高层次思维与推理层次。其中，高层次思维与推理层次包括以下几个要点：（1）应用：探索所习得的信息在个体生活上的应用，对原有生活产生何种变化？（2）改变：如何运用所习得的信息改变个体原有的理解或生活？（3）对重要信息做出个人的判断；（4）计划每日如何运用所习得的信息；（5）将知识与信息做整理。（王为国 2006:169）Lazear将认知领域分类的层次于语文能力表现上的要求，有助对高层次思维认知评量的理解，兹分述如下表：

表5：认知分类层次在语文能力上的要求（王为国 2006：171）

认知 分类层次	语文能力				
	创意写作	诗的表现	正式演说	语言幽默	倾听与报告
3. 高层次思维 与推理 ↑	运用多样化的写作方法去表达个人的想法及感觉	把诗当作是表达思想、感觉和理解的媒介	探索教材较大的应用及和其他学习及知识的链接	能运用主题或概念来达成幽默机智地运用语言的习惯	展现深度倾听及能报告所学习的内容
2.信息分析与 处理 ↑	运用多样化的写作形式及风格去展现对教材的领悟	运用多样化的写诗方法去展现对教材周密思考的过程	经由考虑意义及提出问题来讨论基本的信息	运用教材或概念当作幽默说话及写作的出发点	能做良好的报告及亲自做结论
1. 搜集基本 知识	运用基本的写作形式去传达简单理解与理念	运用基本韵律或诗的韵律形式去展现与主题相关的简单事实及图形	针对基本的事情及信息做简短的演说	运用和学习概念相关的幽默表现	能够报导非常基本的信息以及和其他人沟通

就教学目标而言，目标是具有层次的，层次愈低者，目标就愈形具体，层次愈高者，就愈抽象。所以，教师在进行课堂教学评量时应多运用认知教学目标中的高层次要求，以促进学生在理解方面的理解和应用。教师在教学过程中要留意教学设计及课堂教学评量设计的安排与编制，避免流于较低层次的教学活动或教学评量，因为实施高层次思维认知评量的主要目的在于提高学生在更高层次思维的学习，有助于学生更全面性的发展。

5、口语评量：依据李坤崇、陈英豪、吴裕益、Airasian、Linn和Miller等人的观点，常用的口语评量有两种：（1）口试（oral examination），常用于总结性评量；（2）发问（questions），常用于形成性评量。传统课堂教学评量中以纸笔测验为主，忽略学生在口语方面的评量，所以为提升学生的口语表达能力，教师可在教学时进行发问，在特定时段内进行口试，以促进学生在口语方面的能力。教师在进行口试时，可设计适切的评量表，事前让学生知道评量的标准，用以评量学生的表现，提升学习的效能。

Airasian认为在课堂教学发问有以下几个方面的功能：（1）提高参与感：教师在课堂中发问不仅可让学生专心，亦可提高学生的参与程度；（2）加深思考过程：发问可指明学生的思考方向，同时可引导学生更深入地思考问题；（3）增强学生之间的互动与学习：发问可使学生听到同学对问题的见解，有助于彼此之间的互动，亦间接增进学习效果；（4）达“立即增强”之效：学生答对问题时，教师的肯定或赞赏，均可直接、立即地达到增强的效果；（5）利于掌握教学进度：教师在教学期间进行发问，有助于了解学生的学习情况，以作为调控教学进度的其中一项依据；（6）提供诊断信息：从学生对于问题的回答，教师可诊断他们在学习方面的问题，以作为进行补救教学的参考。（李坤崇 2006：358）

6、轶事记录：Linn和Miller对轶事记录作了以下的阐释，轶事记录乃教师观察到有意义事件的事实性描述，即教师观察学生日常生活的表现，详细写下重要且有意义的偶发个人事件和行为的记录。（李坤崇 2006：366）要强调的是，轶事记录并非课堂教学评量的唯一依据，而是评量过程中作为佐证的信息，因为学生在日常生活上的表现，很难在班级情境中体现出来，唯有通过教师的观察与记录，才能呈现出学生更全面性的表现，无形中亦可提升课堂教学评量结果的效度。

Linn和Miller提出，轶事记录的原则有以下五个方面：（1）叙述简单且具体：依据事件扼要具体的描述，切勿流于冗长；（2）描述事件情境：简明描述学生所陈述所经历事件的情境；（3）勿撰写评量式的轶事：教师不应进行价值判断，径行判断该行为的好坏或对错；（4）勿撰写类推性的描述轶事：教师应用一般话语进行描述具体的行为，避免个人认知上的类推；（5）勿撰写解释型的轶事：教师应描述具体的事件情境，不应解释行为的原因。（李坤崇 2006：367）由此可见，轶事记录的使用，教师首先要重视的是客观性，应客观地记录事件，不宜直接将个人主观的理解、价值判断投注在记录上。

承上所述，多元化课堂教学评量的类型主要有以上中外学者普遍论述和采用的六种，基于这六大类型，可延伸更多不尽相同但用意相似的课堂教学评量方式。总而言之，课堂教学评量的发展已趋向多元化、弹性化、人性化，这是不争的事实。

（二）多元化课堂教学评量的特点

高博铨指出多元化课堂教学评量是以教师、学生、师生间共同参与的课程与教学活动为对象，结合各种理论、评量工具、方法和资源等多样化的探讨途径，进而了解教师的教学效率、学生的学习成就，以及课程的设计与实施的情形。

吴毓莹认为课堂教学评量的开放，除了从一元丰富成多元外，尚需从成果导向走向历程导向，从点或平面的描述走向立体的呈现。田耐青则认为多元化的课堂教学评量具备以下五项基本原则：

1、评量长期化：长期评量可对学生的作品或表现进行连续性的观察，并可提供学生对学习做持续性的自我反省。

2、评量多元化：评量内涵含内容和技巧两个部份，评量人员包括了教师、学生和家長。

3、评量反馈化：评量的结果应能为教学提供信息，反馈予教师、学生和家長。

4、评量兼顾非正式评量：学生在课堂互动的日常表现，反映了他的学习态度，所以应给予关注和观察。

5、评量应激励主动的自我评量：学生有能力对自己学习的优缺点进行评量，能理解自己所学到的知识与概念，更能确认自己应如何运用适切的思考模式和学习历程。

张稚美提出了以纸笔测验、教师观察以及成长档案的三角检核法，用以评量学生的学习与成长。纸笔测验是以评量学生在认知层面的学习结果。教师观察则是教师对评量的知能和目标进行系统化的设计，再对学生进行观察和纪录，这部分主要是对学生在情意和技能方面的能力和学习成果进行评量。成长档案的评量可完整地呈现学生学习的整个历程，能为学生构筑学校与家長的沟通桥梁。（李坤崇 2010：2）

加德纳对于学生学习与成长的思维向度有所不同，他强调的是对未来的评量，他认为未来的课堂教学评量应具以下几个方面的特点：

1、评量重于测验：一般而言，测验只为几个目的而进行评量，故无法确实反映外在社会环境的实际需求；评量则可通过多元的方式了解个人在不同领域里的技能或潜能，提供个人更有效的反馈信息。

2、课堂教学评量乃教学历程的措施：评量是教学历程中的一部分，应与简单、自然的学习情境相结合，而非加附于教学或单独的个体。

3、评量应具生态效度（ecological validity）：评量宜在与真实情境相似的状态下进行，如此才能对个人的最终表现做出最佳的预测。

4、评量工具应具“智能公平”原则：大部分的测量工具着重于语文智能和数学逻辑智能，所以在这两项智能较强的人，其评量结果通常亦较好，所以评量工具应顾及及其他不同智能的强弱优劣。

5、利用多元评量工具：评量宜采用多元的测量工具，以评价不同能力的各种方面。

6、顾及个别差异：进行课堂教学评量时应考虑个体的个别差异、不同发展阶段及不同专业知识，所产生的评量结果的差异。

7、课堂教学评量素材趣味化：评量应利用一些有趣及能引发学生学习动机的素材，让学生乐于进行学习。

8、以学生的利益为前提：课堂教学评量的一个主要目的是帮助学生了解他们在学习上的优缺点，所以应以学生的利益为主要考虑点，将评量信息适切地反馈给学生。

（李坤崇 2010：2）

综合田耐青、李坤崇、张稚美、加德纳、Airasian、Lazear、Wolf、Bixby、Glean和Linn及Miller等人的观点，台湾学者李坤崇将多元化课堂教学评量涵义阐述如下：多元化课堂教学评量以教师教学与评量专业为基础，依据教学目标拟订适切的评量方式、评量内涵、评量人员、评量时机与过程，并呈现多元的学习结果，以提供更适性化的教学来增进学生的成长。（李坤崇 2010：4）

五、课堂教学评量多元化概念的误解

多元化课堂教学评量的提出，无疑为课堂教学评量旧有传统模式带来了冲击，也更贴近课堂教学评量改革的需要。在推动和执行多元化课堂教学评量的过程中，无论是学校行政单位或任课教师，在不同层面的概念上难免有些误解，兹澄清如下：

1、多元化课堂教学评量并非废除传统纸笔测验

有的人认为，实施多元化课堂教学评量就是要废除纸笔测验，这实为概念上的一大误解。事实上，多元化课堂教学评量乃以纸笔测验为基础，在测验的编制上注入多元化的元素，以达量化与质化评量相结合的效用。多元化课堂教学评量的实施其实是基于适性化和目标化的理念，既要让学生的能力得到充分的发挥，又兼顾教学目标，以较为适当的评量方式来进行教学成效的检视。

2、多元化课堂教学评量并非一种教学评量方式

有的人将实作评量、档案评量、变通性评量等单一的评量方式视为多元化课堂教学评量，这是另一种思维上的误区。多元化课堂教学评量，顾名思义就是在评量学生学习表现时，使用两种或两种以上评量方式，要强调并非在某一次的评量时使用两种或以上的评量方式，而是在一个学习阶段或一个课程单元内使用不同的评量方式来评量学生的表现，如此才能更全面地获取学生不同层面的学习成果。

3、多元化课堂教学评量并非为了多元而多元

多元化课堂教学评量的实施并非为了教育改革而改革，更不是为了课堂教学评量改革而多元，甚至认为没实施多元化的课堂教学评量就不是在改革评量，不是在进行教育改革。其实，多元化课堂教学评量的实施应依据教学评量目标，选取最适切的评量方式，如此才有利于更公正客观地评量学生的学习成就，若陷入为多元而多元或纸笔测验的窠臼，皆非多元化课堂教学评量的要旨。多元化课堂教学评量应考虑教师和学生负担，若评量的进行过于繁琐、复杂、费时费钱过多，都将造成教师和学生困扰，造成过度重视评量而轻视教学的乱象。因此，教师在实施多元化课堂教学评量时应以“简易可行”为基本准则。

4、多元化课堂教学评量并非课堂教学评量的目的

多元化课堂教学评量是了解和提升学生学习成效的工具，是一种教学评量的手段而非目的。所以，教师切勿将多元化课堂教学评量视为目的，一味的盲从附合，而应秉持教师的专业素养，考虑教学目标，选择适合的评量方式来评量学生的学习成果。

5、多元化课堂教学评量并非学习的终点

将评量视为学习的终点，是一些教师和学生多元化课堂教学评量在思维上的另一个误解。他们将评量结果仅作为奖惩的依据，而不能深入思考学生在学习方面的优劣得失，更未提出相应的补救措施。实施多元化课堂教学评量的其中一个要点是根据学生的学习情形，提出适性的教学策略，激发学生的潜能或补救学生在学习上的缺失，所以多元化课堂教学评量是学习的另一个起点而非终点。

6、多元化课堂教学评量并非主观的评量

多元化课堂教学评量常将纸笔测验、实作评量、档案评量、变通性评量等多种评量类型用于评量上。一般而言，除纸笔测验的客观性较少被质疑外，其他方式多会让人有主观有余客观不足的感觉。其实，严谨的多元化课堂教学评量会在评量前拟定评量的指标为计分依据，尽量减少主观因素的干扰，因此，多元化课堂教学评量未必是主观的，只要能拟定评分依据，评量仍是客观公正的。（李坤崇 2010：4-5）

六、结语

课堂教学评量是调控课堂教学的重要手段之一，对学生的学习，发挥着检测、诊断、指向和激励等功能。虽然长期以来，教师仍习惯于传统课堂教学评量方式的运用，认为传统笔试较能客观的评量学生的学习成效。然而，多元化课堂教学评量的提出，为的是优化课堂教学的质量，促进学生的自主发展和主动学习，无疑为课堂教学评量指明了一个革新的方向。为了培养能够面对未来挑战的学生，多元化课堂教学评量强调教师在实施评量时，以全面性的思维去看待学生的学习，在实际的操作中通过实作评量、档案评量、变通性评量、高层次思维认知评量等方式来进行评量，使学生能够面对未来剧烈变化的社会带来的考验。通过在学校进行模拟现实社会情境的多元化课堂教学评量，学生能够提前接受锻炼，做好准备，从中锻炼出面对未来挑战的能力与实力。

简而言之，本文旨在探析多元化课堂教学评量的基本原理，希冀为推进当前马来西亚华文教育课堂教学评量的改革，提升课堂教学水平，为保障课堂教学质量奠下基础。

参考文献

- 蔡斌祥，1997，《谈开放教育的教学评量》，邓运林主编《开放教育多元评量》，高雄：高雄复文出版社。
- 陈湘媛，1997，《谈开放教育教学活动的多元评量》，邓运林主编《开放教育多元评量》，高雄：高雄复文出版社。
- 崔永华编，2008，《对外汉语教学设计导论》，北京：北京语言大学出版社。
- 李坤崇，2006，《教学评量》，台北：心理出版社。
- 李坤崇，2010，《教学评估》，上海：华东师范大学出版社。
- 覃兵，2010，《课堂评价策略》，北京：北京师范大学出版社。
- 日本筑波大学教育学研究会编，2003，《现代教育学基础》，上海：上海教育出版社。

- 王财印、吴百禄、周新富，2012，《教学原理》，台北：心理出版社。
- 王道俊、郭文安主编，2009，《教育学》，北京：人民教育出版社。
- 王为国，2006，《多元智能教育理论与实务》，台北：心理出版社。
- David Lazear，2000，《落实多元智慧教学评量》，台北：远流出版社。
- Donald R. Cruickshank等，2007，《教师指南》第4版，南京：江苏教育出版社。
- Howard Gardner著、沉致襄译，2008，《重构多元智能》，北京：中国人民大学出版社。
- Linda Campbell, Bruce Campbell, Dee Dickinson著、郭俊贤、陈淑惠译，1998，《多元智慧的教与学》，台北：远流出版社。
- Weber, E. 著、中国基础教育课程改革“促进教师发展与学生成长的评价研究”项目组译，2003，《有效的学生评价》，北京：中国轻工业出版社。
- Miller, M.D., Linn, R.L. & Gronlund, N.E. 2009. *Measurement and Assessment in Teaching* (10th ed.). Upper Saddle River, NJ: Merrill.