

学业评价

李坤崇◎著

多种评价工具的设计及应用

XUEYEPINGJIA

DUOZHONGPINGJIAGONGJU

DESHEJIIYINGYONG



华东师范大学出版社

学业评价

多种评价工具的设计及应用

XUEYEPINGJIA

DUOZHONGPINGJIAGONGJU

DESHEJIJIIYINGYONG

李坤崇◎著

图书在版编目(CIP)数据

学业评价/李坤崇著. —上海:华东师范大学出版社,2016.

ISBN 978-7-5675-5920-2

I. ①学… II. ①李… III. ①教育评估
IV. ①G40-058.1

中国版本图书馆 CIP 数据核字(2016)第 291907 号

简体中文版由作者李坤崇先生授权出版发行。版权所有,盗印必究。
上海市版权局著作权合同登记 图字:09-2010-681 号

学业评价

多种评价工具的设计及应用

著 者 李坤崇
项目编辑 蒋 将
审读编辑 蒋 将
责任校对 张多多
装帧设计 俞 越

出版发行 华东师范大学出版社

社 址 上海市中山北路 3663 号 邮编 200062

网 址 www.ecnupress.com.cn

电 话 021-60821666 行政传真 021-62572105

客服电话 021-62865537 门市(邮购)电话 021-62869887

地 址 上海市中山北路 3663 号华东师范大学校内先锋路口

网 店 <http://hdsdcbbs.tmall.com>

印 刷 者 浙江省临安市曙光印务有限公司

开 本 787×1092 16 开

印 张 22.25

字 数 534 千字

版 次 2016 年 11 月第 1 版

印 次 2016 年 11 月第 1 次

书 号 ISBN 978-7-5675-5920-2/G·9965

定 价 40.00 元

出版人 王 焰

(如发现本版图书有印订质量问题,请寄回本社客服中心调换或电话 021-62865537 联系)

序

学业评价是当前课程教学研究领域的热点话题,也是我国新课程推进过程中的重大难题。学业评价的核心理念旨在改进与促进,而不再只是鉴定或选拔;旨在让学生学会自我评价,而不再只是充当“受评人”;旨在提供改进或促进的依据,而不再只是分出三六九等;旨在描述与交流评价结果,而不再只是告知对错或订正;旨在评价本身的不断完善,而不再只是习惯与重复运用原有的工具……经过十多年新课程的传播,大部分老师们可能都知道了这些理念,但苦于不知道符合上述这些评价理念的工具与技术是什么样的。

本书就是这样的一本书。它清晰地阐述评估学生的学习成果的理念,说明如何精准掌握教学目标,并详细介绍多种评价工具及其应用实例,包括纸笔测验、表现性评价、档案袋评价、口语评价、系列表现性评价等。另外,本书也提供了教师安排学期学业评价的实例、学业评价结果通知的方式等,提供了日本、澳大利亚、美国、英国等不同学业评价通知单的范例,以供读者在形成本土化的评价计划时作参考。

该书的作者是李坤崇教授。他是台湾地区中小学九年一贯新课程的推手,也是台湾高中新课程的擘手,更是目前如火如荼地进行的大学课程改进的教育专家。李教授堪称台湾新课程革新的活字典、学业评价的实践者,他是秉持丰厚的学术涵养与参与台湾教育革新的丰硕经历来撰写本书的。

由于机缘,本人与李教授交往多年,得以有幸提前阅读本书,深感此书涉及评价领域的前沿,论点条理分明,辅以各种不同的实例,体现实务与理论的一致性。而且,本人在阅读过程中,经常会产生“似曾相识”的感觉。因此,本书是本人最喜欢阅读的评价类图书之一,本人愿意在此与各位读者分享阅读体会并竭力推荐。

崔允漷

自序

教育改革虽为课程、教材、教学、评价的整体工程，然而就走访与观察两岸四地教育的经验，深感其顺序应系先教学、再评价、再次课程、最后教材，即从与班级教学最直接相关的创新教学着手，再改善学业评价，再扩及学校课程计划，最后自编或改编教材。近十年来观察两岸四地中小学教育现场，发现教学、课程、教材的进展远大于评价。

为提升教师学业评价素养，于1999年出版《多元化教学评量》一书后，感谢各界的关怀与支持，获得出乎意料的回响。为回馈各界的支持，乃持续收集国内外文献与发展实例，并于台湾中小学推广学业评价的理念，与带领主任、教师进行纸笔测验与多元评价实作，于2006年出版《教学评量》一书。

2008年后走访大陆、香港及澳门，得以与两岸四地的中小学教师进行学术交流，分享研究成果。在交流分享过程中有深深体会：两岸四地学生的学业评价仍有待加强。

在与两岸四地中小学校长、主任、教师进行实际互动的过程中，发现教师仍普遍缺乏编制纸笔测验、表现性评价、档案袋评价、口语评价、轶事记录的基本素养，尤其是编制段考的纸笔测验多剪贴自出版社光盘片，或直接引自出版社提供的试卷，未绘制双向细目表，使得纸笔测验未必呼应教学目标、未能评价学生的真正学习成果。

教育改革强调“多元评价”，却让中小学教师误以为多元评价就是废除纸笔测验，误以为多元评价就仅仅是工作表现评价或档案袋评价，误以为多元才对、一元不对，误以为多元评价是主观的评价，误以为多元评价可以解决所有教学与评价的问题。为厘清误解，特提出“多元化”、“人性化”、“精确化”、“统整化”评价的趋势。“多元化”学业评价强调专业多元、内涵多元、过程多元、时机多元、情境多元、方式多元、人员多元、计分多元及结果呈现多元；“人性化”学业评价着重于评价中心学生化、学业评价计划化、评价内涵同理化、结果呈现增强化及结果解释正向化；“精确化”学业评价强调编制历程标准化、题目编拟通则化、题目分析质量化及结果解释定锚化；“统整化”学业评价则重视课程与评价计划化与统整化、教学与评价统合化与适性化及学期评价计划化与精致化。

笔者有感于两岸四地中小学教师对学业评价基本理念仍存误解，学业评价专业素养仍待增进，及多元评价实作能力仍待加强，乃着手重新整理数年的研究经验与实作成果，厘清学业评价新理念，并纳入两岸四地试卷实例，再度汇整成书。在华东师范大学崔允漦所长的引荐下，2011年于华东师范大学出版社出版《教学评估：多种评价工具的设计及应用》。

感谢两岸四地中小学教师、主任、校长的支持，《教学评估：多种评价工具的设计及应用》一

书已售罄。为了让书名更容易被两岸四地教育界伙伴了解,并纳入近几年实践的案例,特更名为《学业评价:多种评价工具的设计及应用》。再度感谢两岸四地中小学教师、主任、校长持续提供宝贵的教学与评价经验,得以让理论与实务更为契合;感谢教育行政机关继续提供学习成长的机会,得以让笔者跟着实际推动的伙伴共同学习与成长;更感谢华东师范大学课程与教学研究所崔允漦所长、华东师范大学出版社协助出版本书。笔者期待能持续激起更多两岸四地专家学者与中小学、教师、主任、校长的回响,进而改善学业评价,落实教育改革。若有任何指教,请以 Email(dr. lklkc@gmail.com)联系。

李坤崇 谨识

于2013年4月 台湾嘉义

第一章 学业评价理念	1
第一节 多元化学业评价	2
一、多元化学业评价的意义	2
二、多元评价的迷思	4
三、多元化学业评价的特质	5
第二节 人性化学业评价	10
一、人性化学业评价的意义	10
二、人性化学业评价的特质	10
第三节 精确化学业评价	11
一、精确化学业评价的意义	11
二、精确化学业评价的特质	11
第四节 统整化学业评价	13
一、统整化学业评价的意义	13
二、统整化学业评价的特质	13
第五节 改善学业评价的实施建议	14
一、实施学业评价应掌握的原则	15
二、实施学业评价应具备的条件	15
三、实施学业评价的准备工作	16
四、推动学业评价之策略	18
五、结语	19
第二章 教育目标、教学目标及能力指标	20
第一节 教育目标	20
一、认知教育目标	20
二、情意教育目标	26
三、动作技能教育目标	28
第二节 教学目标	30
一、单元目标	31
二、具体目标	31
第三节 能力指标	33
一、能力指标的功能、特质	34
二、能力指标解读、转化的整合理念	34

第三章 纸笔测验编制与实例 41

第一节 纸笔测验编制步骤 41

- 一、决定测验目的 42
- 二、设计双向细目表 43
- 三、编拟测验试题 52
- 四、审查与修改测验试题 61
- 五、编辑测验试题 62

第二节 各类试题类型的编制原则 64

- 一、选择题编制原则 64
- 二、是非题编制原则 77
- 三、配对题编制原则 80
- 四、填充题编制原则 85
- 五、解释性练习题编制原则 89
- 六、论文题编制原则 95

第三节 纸笔测验试卷实例 100

- 一、台湾小学自然与生活科技学习领域试卷 100
- 二、台湾中学语文学习领域试卷 100
- 三、香港语文科拟题计划及试卷 113

第四章 表现性评价理念 123

第一节 表现性评价的内涵 123

- 一、表现性评价的意义 123
- 二、表现性评价的特质 125
- 三、表现性评价的优点与限制 129

第二节 表现性评价的类型 131

- 一、限制反应实作作业 132
- 二、扩展反应实作作业 132

第三节 表现性评价的编制 133

- 一、呼应评量目的采取表现性评价 133
- 二、确定表现性评价的实作目标 134
- 三、研拟重要的实作层面及可观察的实作指标 135
- 四、规划评量规准、程序及计分 138
- 五、设计实作表 140
- 六、制作使用说明与评量准则 141
- 七、表现性评价的实例 141

第四节 表现性评价的信度与效度 146

- 一、进一步明确具体地界定评价目的 147
- 二、提高表现性评价效度 147

三、提高表现性评价信度	148
四、力求客观的学业评价通知单	149
第五章 评定量表与检核表理念与实例	150
第一节 评定量表	150
一、评定量表的种类	151
二、评定量表使用领域	153
三、评定量表的误差	156
四、有效运用评定量表的原则	157
第二节 评定量表实例	159
一、综合活动学习领域评价实例	159
二、健康与体育学习领域评价实例	159
三、自然科实验评价实例	166
四、自然科专题报告评价实例	166
第三节 检核表	173
一、检核表的功能	173
二、检核表的编制	173
三、检核表的计分	175
四、检核表的不足与应对之道	175
第四节 检核表实例	176
一、小学生活、学习领域检核表实例	176
二、小学数学领域检核表实例	179
三、中学垒球投球技能检核表实例	183
第六章 档案袋评价理念与实例	186
第一节 档案袋评价理念	186
一、档案袋评价的意义与特质	186
二、档案袋评价的优缺点	189
三、档案袋评价的类型与结构	190
四、档案袋评价的制作与实施	193
五、档案袋评价的运用原则	197
第二节 档案袋评价实例	199
一、综合活动学习领域学习档案袋评价实例	199
二、语文学习档案袋评价实例	200
第七章 口语评价与轶事记录理念与实例	218
第一节 口语评价理念	218
一、口语表达的优缺点	218

二、运用口语评价的原则	219
第二节 口语评价实例	221
第三节 轶事记录理念	221
一、记录的内涵	221
二、轶事记录的优缺点	225
三、轶事记录的改善原则	226
第四节 轶事记录实例	227
第八章 系列表现性评价理念与实例	229
第一节 系列表现性评价理念	229
一、系列表现性评价的优点	229
二、系列表现性评价的缺点	230
三、实施系列表现性评价的原则	231
第二节 系列表现性评价实例	233
第九章 学期学业评价理念与实例	251
第一节 学期学业评价理念	251
一、规划目标	251
二、规划原则	251
三、实施策略	253
第二节 学期学业评价实例	253
第十章 学业评价通知单	288
第一节 呈现学习结果	288
一、评定学习结果的范围	288
二、呈现学习结果的方式	288
三、多元化地呈现学习结果	291
第二节 学业评价通知单的功能与困境	292
一、功能	292
二、困境	293
第三节 世界主要发达国家学业评价通知单的实例与解析	294
一、日本打濑中学的学业评价通知单	294
二、日本打濑小学的学业评价通知单	300
三、日本绪川小学的学业评价通知单	301
四、澳大利亚小学的学业评价通知单	320
五、美英小学的学业评价通知单	323

第四节 我国台湾地区学业评价通知单的实例与省思 328

一、力求突破的学业评价通知单实例 328

二、改善学业评价通知单的省思 329

参考书目 338

美国教育测量学会(American Educational Research Association, 2009)认为,评价(assessment)是指经由观察、表现与成就等方式来了解学生获得学生学习信息,并对学生学习进展予以价值判断,评价要回答的问题为“这个人表现得有多好”,被测量比描述(测量)与质化描述(非测量)。测量为一种量化的过程,用以描述个人拥有特殊特征的程度,例若要回答的问题为“这个人读了多少”,重点在量化描述;测验(test)是指运用一种特定的工具或系统来测量学生(如:纸笔试题),在固定时间与合理可比较的情境下对所有学生进行测量,回答的问题是“这个人表现得有多好”,重点在个人表现与他人比较。

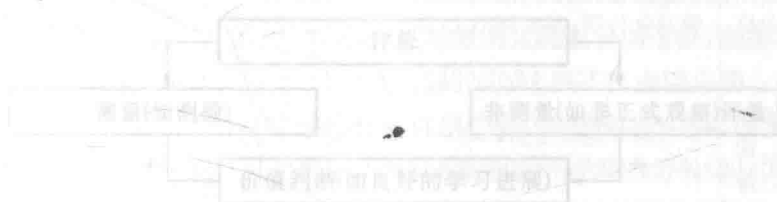


图 1-1 评价过程

资料来源: M. K. Miller, R. L. Linn & N. E. Gronlund, *Measurement and Assessment in Teaching* (10th ed.), Upper Saddle River, NJ, Merrill, 2004, p.29.3

评价(evaluation)乃评估学生学业的学习成果,以检验教师教学目标的达成情况。本书并不用词为教学评估(assessment in teaching),因易与“学业评价”混淆,更易被误认为“教师教学工作的评价,即英文的 teaching evaluation,经与华东师范大学课程与教学研究所研究员李其龙及华东师范大学出版社讨论后,决定改为“学业评价”。

Wiggins, 1982)认为,评价(assessment)不应只是用于了解学生的表现,更应作为改善学生的表现,林与米勒(Linn & Miller, 2005; Miller, Linn & Gronlund, 2006)认为,评价能提供学生目前的成就状态与进展及与教育质量相关的信息;学习中心课程,应注重要学习什么、内容标准,并决定何时应达成何种程度的标准(表现标准)与达成其应达到的标准本位评价,艾琳逊(Peter W. Altrichiani, 1980)指出教师依据学生不同的学习行为,教学,综合决策,“行政决策”是指打分数、分组、评价进度、解释考试分数与要求合格,评量需要特别安置或奖励的学生;“教学决策”是指教师思考教什么,如何教,教材,教材选择,课程如何讲解以及活动设计如何依学生反应而变化;“综合决策”是指了解学生的行为特征,对课堂的教学、沟通与合作及建立与维持有效的课堂气氛。

我国台湾地区学业评价制度,未来宜朝“多元化学业评价”、“人性化学业评价”、“精确化学业评价”、“程序化学业评价”等四个制度努力以赴,以期能提升教育品质,提高学生学习成效。

第一章 学业评价理念

米勒等(Miller, Linn & Gronlund, 2009)认为,评价(assessment)是指经由观察、表现与专题评定或纸笔测验获得学生学习信息,并对学生学习进展予以价值判断;评价要回答的问题是“这个人表现得有多好?”,兼顾量化描述(测量)与质化描述(非测量)。测量(measurement)是获得量化的过程,用以描述个人拥有特殊特征的程度;测量要回答的问题是“这个人表现了多少?”,重点在量化描述。测验(test)是指运用一种特定的工具或系统化程序(通常包括一组试题),在固定时间与合理可比较的情境下对所有学生进行测验;测验要回答的问题是“无论是与个人或他人的表现比较,这个人表现得有多好?”,重点在个人与他人的表现的比较。

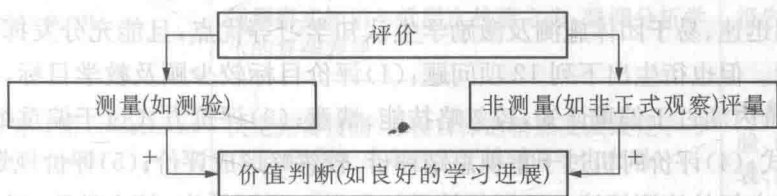


图 1-1 评价过程

(数据来源: M. D. Miller, R. L. Linn & N. E. Gronlund. *Measurement and Assessment in Teaching* (10th ed.). Upper Saddle River, NJ: Merrill, 2009, p. 29.)

学业评价(academic evaluation)乃评估学生学业的学习成果,以检核教师教学目标的达成程度。本书原本用词为教学评估(assessment in teaching),因易与“学业评价”混淆,更易被误认为以教师教学为主的评价,即英文的 teaching evaluation,经与华东师范大学课程与教学研究崔允漷所长及华东师范大学出版社讨论后,决定改为“学业评价”。

威金斯(Grant P. Wiggins, 1998)认为,评价(assessment)不应只是用于了解学生的表现,更应用于教育与改善学生的表现。林与米勒等(Linn & Miller, 2005; Miller, Linn & Gronlund, 2009)均强调评价能提供学生目前的成就状态与进展及与教育质量相关的信息;学业评价必须确认学生需要学习什么(内容标准),并决定何时应达成何种程度的标准(表现标准),方能落实绩效责任的标准本位评价。艾瑞逊(Peter W. Airasian, 1989)指出教师依据学业评价结果,来实行政治、教学、综合决策:“行政决策”是指打分数、分组、评价进度、解释考试结果、与家长会谈、评断需要特别安置或奖励的学生;“教学决策”是指教师思考教什么、如何教、何时教、用何教材、课程如何进展以及活动设计如何依学生反应而变化;“综合决策”是指了解学生学术、社交与行为特征,增强班级教学、沟通与合作及建立与维持有效的班级气氛。

观之两岸四地学业评价现况,未来宜朝“多元化学业评价”、“人性化学业评价”、“精确化学业评价”、“统整化学业评价”等四个向度全力以赴。

第一节 多元化学业评价

纸笔测验虽然能测量认知领域的学习结果,但在技能、情意领域则有其先天限制,如文章、图画、实验报告、自然科学展览设计、演讲、握笔、打字、问题的形成、想法的组织、创意和小组合作学习能力等,均难以运用纸笔测验评价(Wolf, Bixby, Glen & Gardner, 1991; Linn & Miller, 2005; Miller, Linn & Gronlund, 2009)。教师教学必须兼顾认知、技能、情意之学习结果,评价不仅应分析“应该怎样表现(知识)”与“真正表现行为(实作)”之间的差异,亦应兼顾“过程或程序效能”与“作业表现作品效能”,方能检讨教学成效与实施补救教学,故宜善用“多元评价”方能適切评价学生真正的各项学习表现。

一、多元化学业评价的意义

李坤崇(1999, 2001a, 2001b, 2002a, 2006)认为,传统的以纸笔测验为主的学业评价,具有计分客观、批阅迅速、易于团体施测及激励学生认知学习等优点,且能充分发挥公平、客观、省时、省钱的功能。但也衍生出下列 12 项问题:(1)评价目标较少顾及教学目标,导致教学与评价脱节;(2)评价内涵过于偏重认知,较忽略技能、情意;(3)评价方式过于偏重纸笔测验,较忽视其他评价方式;(4)评价时机过于重视总结评价,较忽略形成评价;(5)评价规划不是周延的,未能营造公平、良好的施测情境;(6)评价结果解释过于强调缺失,较少鼓励、激发学生;(7)评价结果解释过于依赖量化测量,忽略质化描述;(8)评价认知过于强调记忆层次,忽略高层次认知层次;(9)命题观念与技术延续仍有待加强;(10)测验试题编排以教师为中心,未以学生为中心;(11)评分缺乏客观标准与自省思维,流于主观;(12)家长分数至上观念难以消除,升学主义持续挂帅。可见,传统的学业评价与人性化、多元化的学业评价相去甚远。

米勒等(Linn & Miller, 2005; Miller, Linn & Gronlund, 2009)均强调评价过程包含各式各样的程序,依据其采用的参考架构,可将程序分类成许多不同的方式。有关班级评价程序的描述,详见表 1-1。

拉齐尔(Lazear D., 1999)认为,教师实施多元教学的意愿高于多元评价,其因有二:一为教师较缺乏多元评价经验与知能,实施时较不易获得家长与同学的支持;二为家长不习惯看自己与他人孩子的分数,更不习惯去分析与鉴定孩子的各项学习与关键能力。然而,身为教师应当为所当为,而非取决于个人意愿。教师当务之急在于提升评价专业素养,而非回避多元评价。

张稚美(2000)提出通过纸笔测验、教师观察(如真实或实作评价)、多元智慧成长档案的三角检核法,来评价学生成长历程与学业发展,且此三项各占学年成绩的三分之一。纸笔测验用来评价学生认知层面的学习结果;教师观察指教师先系统建立评价知能与评价指标,再对学生实施有目的的“观察”与“记录”,此方式颇能评价学生技能、情意层面的学习结果;成长档案评价可完整呈现学生整个学习历程,提供学生自我评价与习惯性省思的机会,若再将家庭纳入评价学生学习与成长历程之中,更能为学生搭起学校与家庭的桥梁。张稚美(2000)虽然发展出系列的评价方式与具体策略,颇能兼顾纸笔测验、教师观察、成长档案三种评价方式,然而评价方式与评价人员仍可更多元,且评价内涵不一定局限于多元智慧。

加德纳(Gardner H., 1993)从不同的思考向度去看学生的学习成长,强调未来评价将具

表 1-1 班级评价程序的描述

分类基础	评价类型	评价功能	工具举例
测验本质	最大表现(maximum performance)	决定个人“能做(can do)”的最佳表现,即个人全力以赴时,能做到的最佳表现。	性向测验、成就测验
	典型表现(typical performance)	决定个人在自然状态下“会做(will do)”的表现,即强调个人具有代表性的表现而非最佳表现。	态度、兴趣、人格量表、观察技术、同学评价
测验形式	固定选项测验(fixed-choice test)	知识、技能的有效测量,乃间接指标,即学生从既有的选项中选择问题的答案。如选择题、是非题或配对题,学生短时间能回答大量问题,具有计分客观、信度高与成本效益高等优点。	标准化选择题测验
	复杂表现评价(complex performance assessment)	测量学生在情境脉络的表现和学生自身正确评价问题的表现,即让学生建构延展性反应或响应复杂作业。旨在改善固定选项测验过于强调事实知识与低层次技能之失,强调分析学生的复杂表现。	实验室实作、专题计划、论文式测验、口头报告
用之班级教学	安置评价(placement assessment)	决定先备技能、课程目标的精熟度及最佳学习状况,即用于决定教学开始时的学生表现。	准备度测验、性向测验、课程前测、自陈量表、观察技术
	形成评价(formative assessment)	决定学习进展、提供回馈以促进学习和更正学习错误,即用于在教学期间监控学习进度。	教师自编测验、已发布的测验、观察技术
	诊断评价(diagnostic assessment)	决定持续性学习困难的原因(心智、生理、情绪、环境),即用于在教学期间诊断学习的困难。	已出版的诊断测验、教师自编诊断测验、观察技术
	总结评价(summative assessment)	分派等第以决定课程结束时的成就或认证目标达成的精熟度,即用于评价教学课程(或单元)结束时的成就。	教师自编调查测验、表现评分量表、产品量表
结果解释方法	标准参照测验(criterion referenced test)	以清晰定义学习任务的明确范畴来描述学生表现(例如个位数加法)。其解释依据清晰的定义或明确的成就的范畴(标准表现),强调不能达成学习任务。	教师自编测验、已发布的测验、观察技术
	常模参照测验(norm referenced test)	以学生在所处团体中的相对水平来描述学生表现。解释依据清晰界定的团体(常模表现),强调个体间的差异。	标准化性向测验和成就测验、教师自编调查测验、兴趣量表、适应量表

修改自: M. D. Miller, R. L. Linn & N. E. Gronlund, *Measurement and Assessment in Teaching* (10th ed.). Upper Saddle River, NJ: Merrill, 2009, p. 43.

有下列特色:(1)评价(assessment)重于测验(testing)。通常测验为几个目的而实施,无法准确真实反映出外在环境实际的需求,而测量更可获得个人技能或潜能,提供个人有效的回馈及外在环境有效的信息。(2)评价乃教学历程中简单、自然发生的措施。评价乃教学历程中的一部分,应与简单、自然的学习情境高度结合,而非外加。(3)评价应具生态效度(ecological

validity)。评价宜在“真实工作情境”的相似状态下进行,方能对个人的最终表现作出最佳的预测。(4)评价工具应具“智能公平”原则。大部分测验工具着重语文及逻辑—数学两种智能,然此两种智能较强者评价结果通常亦较佳,所以工具应顾及各种智能的人群。(5)评价应利用多元测量(multiple measures)工具。评价宜采取多元的测量工具,以测出不同能力的各种面向。(6)评价应顾及个别差异、发展阶段及各种不同的专业知识:评价时宜考虑个体的个别差异、不同发展阶段以及不同专业知识所产生的评价结果的差异。(7)评价素材趣味化:评价应利用一些有趣及能引起动机的素材,让学生乐于进行。(8)以学生利益为前提:评价旨在帮助学生了解其学习优缺点,教师应将评价信息回馈给学生。

田耐青(1999)提出,符合多元智能理念的评价应具备下列五项原则:(1)评价长期化。长期评价可使教师对学生的作品或表现进行连续性的观察,并可让学生对学习作持续性的反省;(2)评价多元化。评价内涵包含内容和技巧的评价,评价人员包括教师、家长与学生;(3)评价回馈化。评价结果应该能为教学提供信息,回馈学生、教师及家长;(4)评价兼顾非正式评价。应兼顾学生在班级互动的参与度与日常表现以及反映的学习态度;(5)评价激励学生进行主动的自我评价。学生应有能力评价自己学习的优缺点,能清楚说出自己学到的知识或概念,更能确认自己应如何运用适当的思考掌控学习历程。

在此综合田耐青(1999)、李坤崇(1999, 2001a, 2001b, 2002a)、张稚美(2000)、Airasian(2000)、Gardner(1993)、Lazear(1999)、Wolf, Bixby, Glen 与 Gardner(1991)、Linn 与 Miller(2005)及 Miller, Linn 与 Gronlund(2009)的观点,阐述多元评价的意义:多元评价是指以教师教学与评价专业为基础,依据教学目标研拟适切的评价方式、评价内涵、评价人员及评价时机与过程,并呈现多元的学习结果,以提供更适性化的教学来促进学生成长。

二、多元评价的迷思

从李坤崇(2002b)1996年参与台湾地区高雄市开放教育改革,1998年参与台南市教育改革,2000年起参与台湾地区九年一贯课程改革,以及四度走访25个县市的中小学研讨教育与课程改革等的经验,可发现中小学教师或行政人员对多元评价存在很多误区,遂提出下列解读。

(一) 多元评价并非废除纸笔测验

有极少数教师认为,实施多元评价,即废除纸笔测验。此种思维乃极端思维、谬误思维。实施多元评价乃因以往传统纸笔测验定于一尊,纸笔测验几乎掩盖所有评价方式。然而,实施多元评价乃基于适性化、目标化的理念,不仅让每位学生的各种能力都获得充分发挥,而不局限于纸笔书写能力,更让各项教学目标均能以最适当的评价方式,来检验学生达成目标的程度。

(二) 多元评价并非一种评价

有不少教师误将档案评价、实作评价或系列实作评价(游戏化评价)等单一的评价方式视为多元评价,此乃谬误思维。多元评价顾名思义乃评价学生的学习表现,使用两种或两种以上的评价方式,方能称之为多元评价。

(三) 多元评价不能为多元而多元

有些教师误以为没有实施多元评价就不叫课程改革,出现颇多为了多元评价而多元评价的现象。实施多元评价必须呼应评价目标,依据评价目标选取最适合的评价方式。陷入纸笔测验或多元评价的窠臼,均非理想。多元评价应考虑学生、家长与教师的负担,若过于复杂、耗

时过多,将使教师、学生在评价时花费更多时间,而干扰到教学,产生过分重视评价而轻视教学的乱象。实施多元评价亦不应造成家长困扰,应顾及家庭背景、时间,避免对家长造成不必要的负担。因此,教师设计多元评价应以简易可行为原则。

(四) 多元评价并非目的

有极少数教师将多元评价视为目的,一味地盲从附和,而忽略了多元评价的实质是了解学生学习成效的工具,是手段而非目的。教师应秉持专业素养,考虑教学目标,善用最适合的评价方式来衡量学生的学习成果。

(五) 多元评价并非学习终点

两岸四地有些教师或家长将评价视为学习的终点,仅作为奖惩的依据,而未深究其学习得失,更未提出补救或增强措施。教师实施学业评价并非是学习的终点,而是另一个更高阶段学习的起点。教师在实施多元评价、觉察学生在各种学习上的成果后,应提出适性的教学策略,激发学生潜能并补救学生的学习缺失。

(六) 多元评价并非主观评价

多元评价常用纸笔测验、实作评价、轶事记录、口语评价、档案评价、系列实作评价(游戏评价)、动态评价等多种评价方式。除纸笔测验较少受质疑外,颇多人误以为其他评价方式也是主观的评价。其实严谨的实作评价、轶事记录、口语评价、档案评价、系列实作评价(游戏评价)及动态评价,通常会拟定评分标准作为计分依据,尽量减少人为的主观干扰。因此,多元评价并非主观评价,一份详列评分标准的非纸笔评价,评分仍相当客观。

(七) 多元评价并非“万灵丹”

多元评价虽可避免单一纸笔测验之失,但多元评价并非万灵丹,仍有其先天的限制,如增加工作负担、经费负担。除了应不断充实与提升自己的评价专业素养,持续研发各种符合教学目标的多元评价方式外,教师亦应获得学校行政、家长、社会大众的支持,不断尝试与修正,共同面对挫折并度过调适期,这样方能落实多元评价理念。

三、多元化学业评价的特质

参考上述多元化学业评价的意义与解读,多元化学业评价必须具备下列特质。

(一) 专业多元

多元评价的专业素养,不仅包括学科专业素养与掌握教学目标,更应包括教学专业素养、评价专业素养。李坤崇(1999, 2001b)认为,为达到评价专业化、目标化水平,学业评价宜加强下列八项重点:(1)掌握教学目标:清晰明确地掌握欲评价科目单元的教学目标;(2)专业判断知能:有些教师质疑一般出版社提供的“习作”或“学业评价单”不完全适合班级学生,因无一套评价工具适用于所有学生,而补救之道在于教师能否以评价专业素养来分析目前普遍运用的各项评价工具的优劣以及其是否适用于班级学生,了解现成评价工具是否需稍作修改以适合班级学生;(3)兼顾技能与情意的评价:评价不应限于低层次的认知,而应顾及技能与情意的评价、高层次的认知;(4)剖析教材内容:逐一剖析教材单元的重心,了解学生可能的学习问题;(5)设计细目表:依据教学目标、教材内容设计完善的双向细目表;(6)规划评价方式或慎选题目类型:采取适当的评价方式,如行为检核表、态度评价表、观察报告等;若采取纸笔测验,宜依据认知领域教学目标的层次慎选题目类型;(7)善用命题技术:命题时必须符合命题原则,充分

善用命题技术来適切呈现学生的学习结果；(8)适当解释评价结果：有些教师解释评价结果时，对学生实施不必要的分类或贴标签，使学生遭受挫折，产生负向自我观念、悲观的处事心态。因此，教师必须适当解释评价结果，让学生能从失败中成长，能从挫折中重生。

(二) 内涵多元

教师实施学业评价时，评价内涵至少包括认知、情意、技能等领域内涵(Linn & Miller, 2005; Miller, Linn & Gronlund, 2009)，展现评价内涵的生活化、多样化。以往学生考试只要抱着教科书死背，当个“贝多芬”(背多分)就可得高分，乃是因为传统学业评价内容多以教科书中认知层面的记忆、理解能力为主，较少顾及技能与情意、认知层面之高层次思考能力，或生活化的题材。李坤崇(1998)认为，推动教育改革除考虑认知、技能、情意外，尚需兼顾学生的学习历程、生活世界与社会行为。“学习过程”包括学生的学习方法、习惯，求知历程或解决问题的能力。学生以往限于书本记忆而疏于应用，且家长的过度保护剥夺了学生的成长机会，使得学生自我解决问题的能力欠佳，故学习历程应被纳入评价之中。“生活世界”乃学生日常行为、待人处事的能力，或许有人会说现今教育培养了一群躲在象牙塔里的书呆子，缺乏走入人群的待人处事能力，若纳入评价，则可引导学生注重此方面的学习。“社会行为”乃学生人际关系的社交行为或社交技巧，时下学生常拙于情感表达或社交技巧，导致此类犯罪事件层出不穷。可见，教师宜透过评价来强化学生的社会行为(李坤崇，1999)。

九年一贯课程基本理念在于养成学生“终身学习”的能力，而传统通知单的制作停留于抄写、认知阶段，若未能突破将难以达成九年一贯课程基本理念。为激发学生自主学习和终身学习的能力，教师在设计通知单时宜特别着重学习过程的导引。设计通知单时应掌握四大重点：(1)告知学习目标；(2)引导学习过程；(3)告知评价重点；(4)记录评价结果。另外，实施九年一贯课程之多元评价内涵，宜将各阶段能力指标进行细项分析(概念分析结果)，再分析能力指标细项与认知、技能、情意的关系程度，最后决定评价内涵。分析能力指标细项与认知、技能、情意的关系程度时，认知层次着重“知识理解”、“思考批判”，技能层次着重“技能表现”，态度层次着重“意愿态度”。然而实际学习活动中的评价内涵，仍随各能力指标细项内涵的不同而不同(李坤崇，2001c)。

评价内涵亦可纳入加德纳的“多元智能理论”(multiple intelligence theory)。加德纳(1999)将智能(intelligence)定义为“个体处理信息的生理、心理潜能，此潜能可用以解决问题或创作某种文化环境的重要作品”。这里强调个体的主动性及文化价值。他对“智能”抱持的基本理念为：第一，智能并非与生俱来就是固定或静态的；第二，智能可教、可学、可提升；第三，智能乃多向度的现象，展现于大脑、心灵、身体等系统的多种层次。

加德纳(1999)强调人类智能至少包含下列八项半智能*：言语—语言智能(verbal-linguistic intelligence)、逻辑—数学智能(logical-mathematical intelligence)、肢体—动觉智能(bodily-kinesthetic intelligence)、音乐—节奏智能(musical-rhythmic intelligence)、视觉—空间智能(visual-spatial intelligence)、自然观察者智能(naturalist intelligence)、人际智能(interpersonal

* 在加德纳的多元智能理论中，普遍的观点是八项智能，之后也有学者提出补充“存在智能”这一项。本书作者李坤崇吸纳了各方观点后，认为增设的这项智能是以原先八项智能之一的“自省智能”析出名，故在此提出了此为半项智能一说。特此说明。——编者注