

Rasch Validation of the Inclusive Curriculum-Based SCALE Instrument in Hong Kong

Kwan-Lan Vicky Tsang¹, Andrew Chung-Yee Tse², Kai-Ming Lee³

¹Department of Special Education and Counselling, Hong Kong Institute of Education, Hong Kong

²Centre for Advancement in Inclusive and Special Education, University of Hong Kong, Hong Kong

³Graduate School, Hong Kong Institute of Education, Hong Kong

Email: vtsang@ied.edu.hk, atse222@gmail.com

Received: Jan. 18th, 2013; revised: Feb. 4th, 2013; accepted: Feb. 10th, 2013

Copyright © 2013 Kwan-Lan Vicky Tsang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Background: Past research has reflected the needs to develop measures in the area of assessment, curriculum and teaching strategies for pupils with special educational needs. The SCALE, an assessment instrument to align assessment with the central curriculum for pupils with intellectual disabilities was developed by a team of teacher leaders from nine special schools in Hong Kong. **Aim:** The objectives of this study are to critically evaluate the validity and reliability of the SCALE attainment scales for each strand in four key learning areas (KLA) of the central curriculum of Hong Kong using Rasch analysis. **Method:** The SCALE data was calibrated using the Rasch measurement model. By calculating the probability of a response by right or wrong answers in terms of a logistic function of the difference between the ability of the person taking the test, Rasch analyses on the assessment data were carried out in the form of model fit statistics, point-measure correlation coefficient, person / item reliability and principal component factor analysis on dimensionality. **Results:** The Rasch statistics indicate that SCALE possesses a high degree of validity and reliability. It can classify pupils into at least 13 levels of abilities, and that the range of item difficulty is capable of stratifying pupils' abilities into at least 8 to 11 attainment levels. The result also supports the uni-dimensionality of the SCALE. That is, SCALE solely measures the attainment levels of the pupils but not other aspects of their learning. **Conclusion:** The use and the impact of SCALE in providing evidenced-based outcome data to document progress of pupils across their years in education, and, to inform school-based curriculum planning are further discussed.

Keywords: Curriculum-Based Assessment; Pupils with Intellectual Disabilities; Rasch Analysis

以全纳课程为本的香港融通学习成效量表(SCALE)之等级量尺分析检视

曾君兰¹, 谢宗义², 李启明³

¹香港教育学院特殊教育及辅导学系, 香港

²香港大学属下融合与特殊教育研究发展中心, 香港

³香港教育学院研究生院, 香港

Email: vtsang@ied.edu.hk, atse222@gmail.com

收稿日期: 2013年1月18日; 修回日期: 2013年2月4日; 录用日期: 2013年2月10日

摘要: 背景: 过去的研究结果显示, 需要为有特殊教育需要的学生发展一些具实用性的学与教套件, 包括相关的评估、课程与教学策略, 藉此促进全纳教育的发展。为响应此号召, 香港9所特殊学校的教师团队为智障生研发了一套以连接特殊学校与主流学校的课程的评估工具, 称为“融通学习成效量表(SCALE)”。**目的:** 本研究目标, 在于以等级量尺模式检视融通量表的信效度, 验证近期研发的融通

量表在评估具特殊教育需要的学生于香港主流课程中的四个主要学习领域的学习能力表现水平时是否达到其研发目的。**方法:**本研究采用了等级量尺模式来检视融通量表的信效度。依据受试者个人能力差异的逻辑函数判断答案正误的机率,检测使用融通量表之观察数据与等级量尺模式的预期数据匹配的适合程度、点数测量相关系数、个人/题项信度及主成因素构念向度分析。**结果:**从统计数据显示出融通量表拥有极高信效度,能将学生分为13个不同的学习能力水平级别,且题项难度范围足够将学生能力分为8~11个学习能力表现水平级别。所有的题项分数都与预期总分具非常高的相关性,证明融通量表具单一构念向度,也就是说,融通量表仅仅测量了学生的学习能力表现水平,而不是测量他们学习的其他方面。**结论:**从评估学习的角度,本文进一步讨论课程评估测量法的使用如何在绩效责任上产生作用;以及对于协助教师指导学习能力水平各有差异的学生的影响。

关键词: 课程本位评量; 智障学生; 等级量尺分析

1. 引言

1.1. 以课程为本的全纳评估模式

目前,智障生的课程取向一直受功能性课程模式主导。在传统的心理-医学典范(Psycho-Medical Paradigm)影响下,功能性课程模式建基于发展理论之假设,认为智障生的心理年龄和婴儿或学前儿童一样,经历相似的发展过程轨迹。因此,功能性课程强调小区生活技巧的训练,即包括职能、家庭、社群和休闲技能领域的训练^[1]。

随着全球卷入融合教育的浪潮,教育模式趋势已从功能性课程模式转移至中央课程学习模式(Central Curriculum Access Model)。后者以“通用课程(A Common Curriculum for All)”为基础,强调所有学生都享有同等学习的机会。主流学校课程内容,大多包括“主要学习领域(Key Learning Area; 简称 KLA)”,或“核心科目”之元素,如语言、数学、科学、社会学习及其他^[1]。各个核心科目中之“科目范畴(Strand)”,又可分成多个学习能力表现水平,放在同一评量尺的连续标准在线彼此校准。这样,课程与评估便被连接起来,好让所有的学生,无论有智力障碍与否,都可以根据其学习能力表现水平进行教学;并可依据不同等级对他们的学习进展进行评估。

1.2. “通用课程”政策在香港的发展

自2005年,一项以实践“通用课程(One Common Curriculum for All)”为原则的课程发展计划正式开展。该计划被称为“融通课程计划”,英文全写为 Systematic Approach to Mainstream Education (SAME)。该

计划以院校合作形式进行,由香港大学与九所特殊学校共同参与^[2]。这些特殊教育学校的学生均存在不同程度的残障问题,如视觉受损、肢体伤残、及轻度,中度或重度程度之智力障碍。各学校委派的教师或科目主任老师组成主要学习领域(KLA)撰写小组,并运用他们的主要专业知识来建立学习能力表现量表(Attainment Scales),然后依据这些量表为各个“主要学习领域”设计评量项目和评估步骤。参加该计划的所有教师必须经过培训研讨会、讲习班及撰写班的培训。约有130多名教师参与了此项计划^[3]。

“融通学习成效融通量表”,英文全称是 SAME Curriculum Assessment for Learning Efficacy (SCALE),简称则为“融通量表(SCALE)”,是融通课程计划的一部分。融通课程计划的整体目标在于为所有特殊教育学校的学生取得以香港主流课程学习的同等机会。因而,基于“通用课程”的精神,融通量表(SCALE)的设计是以连接特殊学校与主流学校的课程为宗旨^[4]。所以,融通量表(SCALE)是课程不可或缺的一部分,在不同学校之间协助教师区分和追踪学生学习状况的工具,并为各科目中表现出不同学习能力水平的学生设定适切的学习目标。

在过去的功能性课程模式下,特殊学校的学生一直使用的是专门针对残障者的课程,包括一些功能性训练科目,例如自理技能、感知肌能力训练等等。评估与课程的设计主要以学校为本位,各个特殊学校所采取的准则不尽相同,教学模式也都有差异;主要特色是由教师主导教与学。教师计划教学的过程通常始于:1)从课表中选出一个课题;2)遵循课程大纲规定的顺序授课;3)决定学习活动内容;4)用分数或

信息回馈来衡量学生的学习水平；5) 开始下一课题^[5]。这些教师的教育计划中，对有特殊教育需要的学生的期望一般较低。而且，很可能在不同学习阶段中出现重复学习目标的情况。在此课程架构下，有特殊教育需要学生仅能被动地学习。

在通用课程的模式下，教师运用融通量表(SCALE)衡量学生在学科上的学习能力表现水平，藉此区分及预期学生的不同学习反应。在此模式下，教师不是按照学生的身体缺陷或能力局限来分类，而是与一般学生一样，根据通用课程中某学科的学习能力表现标准来衡量其学习能力水平；然后根据所评定出来的学生学习能力表现水平(Attainment Level)来编定学习目标及内容。这种教学模式恰好与传统模式相反，教师遵循的顺序为：1) 透过有效评估来区分学生学习能力表现水平；然后 2) 为每个学生量身设计学习目标；3) 按目标策画学习活动和策略；最后 4) 透过学习经历帮助学生达致预期学习成果^[5]。教师亦自然而然地采纳以学生为中心的课堂管理方案(Curriculum Management Approach)。这种教学范式的转移也同时改变了教师的思维心态；他们将注意力从学生的局限转移到对学生的期望。他们的注意力不再侧重于有特殊教育需要学生的能力缺欠方面，而是在发挥学生的潜在学习能力与相关学习证据上去^[6]。

1.3. 融通量表(SCALE)之特点

一个理想的评估工具应具备以下特质：

- 1) 评估内容应与主流课程具有一致性；
- 2) 评估内容应为学生能力状况，而非其残疾状况；
- 3) 评估具绩效(Accountability)，与评估数据结果一致；
- 4) 评量任务乃立基于学习表现指标(Performance Indicators)，并其所代表的学习与范围；
- 5) 能够建立良好的信度与效度、保持主流课程标准设计结构测量的一致性；
- 6) 评量是用来测量学习水平的；尽管学习表现指标不应直接被用于教学活动，但是它们却与教学息息相关，并能够提供教学相关的信息。

“融通学习成效量表(SCALE)”乃是依据上述理想评估标准研发出来的评估工具，它具有以下特点。

首先，在内容设计上，“融通学习成效量表”是

以主流课程为依据；融通量表(SCALE)包括有四个主要学习领域或核心科目：语言教育(L)，数学教育(M)，个人、社会与人文教育(P)，及科学教育(S)。各科目都有不同的学习重点，即学习范畴(Strand)。如语文教育(L)可分为四个范畴：聆听(L1)、说话(L2)、阅读(L3)及写作(L4)。数学教育(M)可再分为三个范畴：数与代数(M1)、形状、图形与空间(M2)及数据处理(M3)。科学教育(S)可再分为五个范畴：生命与生活(S1)、物料世界(S2)、能量与变化(S3)、地球与太空(S4)及科学、科技与社会(S5)。由此可见，课程架构及内容是取材于通用课程，而所采取评估标准及评分报告也适用于所有学校。

融通量表(SCALE)提供了评量活动和一系列针对所有学生的学习能力或表现水平的测试，不论他们所在的学校是特殊学校还是主流学校^[4]。每个学习范围都设有一组的学习能力表现水平。各科目之各范畴共设有十二项学习能力表现水平。这些学习能力表现水平标记在同一条具连续性的学习能力标准在线(A Continuum of Ability)，从最初级的学习能力开始－即皮亚杰(Piaget)所描述的认知发展理论中之感知肌能阶段，最高的四个能力表现水平相当于主流学校里一般小学三年级的学习能力表现水平^[7]。学习能力表现水平中 A1~A3 代表了一个儿童感觉肌能发展的早期水平，因而适用于所有学科。A4~A8 代表了普通儿童在学龄前的学习进程。A9~A12 则代表普通儿童在小学一至三年级学习进程。这种特殊教育与普通教育在测量方法上的重迭对于连接两者的课程具有深刻意义(见图 1)。

融通量表(SCALE)创建出一条针对所有学生学习能力表现水平的共通课程连续标准线(Common Curriculum Continuum)，好让学习能力水平各有差异的学生，都可使用这个共通连续标准来测试。这样，有严重残疾的学生亦能够依据其学习能力水平接受这个共通体系的测试。换言之，所有不同学习能力的学生都可以使用融通量表(SCALE)来测定他们在主流课程科目之学习能力表现水平^[3]。

融通量表(SCALE)的实用性，在于它能容许每位具有不同学习能力的学生都有机会按照自己的理解能力和学习速度去学习共同的科目。从融通量表(SCALE)取得学生之学习水平之后，按照学生的学习能力，教师即可以要求学生有相应的学习表现反应。

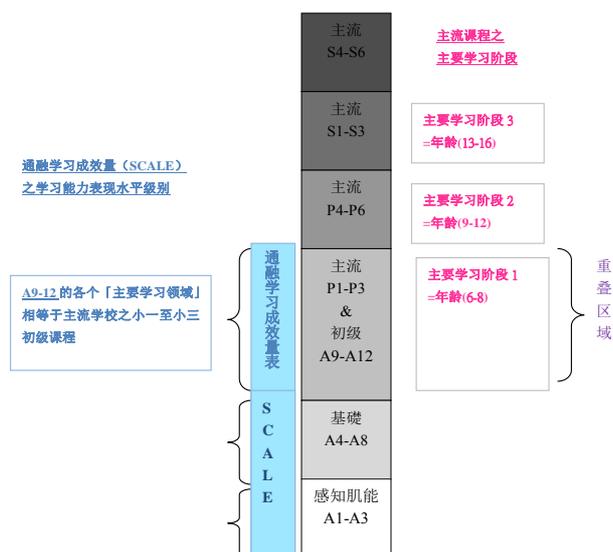


Figure 1. The assessment continuum between the SCALE attainment levels and the key stages of the mainstream curriculum
图 1. 连续融通学习能力表现量表的能力水平级别与主流课程的主要学习阶段之通用课程

老师们将不再视学习能力水平各有差异的学生为一视同仁，以致要求所有的学生都按照相同的速度学习。融通量表(SCALE)提供了一个通用学习表现标准体系，藉此统一了对不同学生的学习能力水平和学习进度的共通语言^[4]。

融通量表(SCALE)提供了学习能力表现水平的描述指标(Descriptors)，确保不同评量人员在评量过程中一致性。学习能力表现水平描述指标并非为了直接用于教学而设计的；如果教师硬要按照学习能力水平描述来进行教学，那么将会造成狭隘的零星技能培训，而非广深的课程学习^[4]。学习能力表现水平描述最大的价值，在于为每位学生在主流课程中各个主要学习领域之学习范畴内，用作各个学习能力水平进展的清晰指标。

1.4. 融通量表(SCALE)之绩效责任

课程本位的评估工具，一般都不会采用常模参照模式，且往往缺少信度和效度检定之证据，从而无法被广泛地应用于大规模的体系中。皆因这些大规模的评估体系，必须分别由很多工作人员在不同的地方各自施行，工作人员虽然已经接受了同一培训课程，但使用评估工具时，在操作测试和计分标准上也仍会因人而异，这局限对评估工具的信度会造成一定程度之影响。为保证课程评量的信度，每位教师须一致地操

作各题项测试过程和采纳相同计分标准，这一致性至关重要。融通量表(SCALE)对学习过程进行的测量属于总结性评量(Summative Assessment)，令其评量活动标准化及操作过程和计分步骤一致化。其中，需要为教师提供测试操作的培训和监督，以确保收集数据过程的一致性，以及在不同班级、课程，以及不同学校里收集的信息可呈现相同的意义。

本研究建议使用等级量尺模式(Rasch Model)对融通量表(SCALE)进行检定测试。传统的检测模式，通常建基于一个较大，且具常模代表性的样本数量(A Norm-Representative Sample size)，它依据常态分布假设(Normal Distribution Hypothesis)对信度和效度进行计算。等级量尺模式的不同之处，在于它是以题项反应理论为基础(Item Response Theory)。该理论依据个人能力和其对各题项难易度之间的相关反应的假设，来探究个别受试者的能力与相对各题项反应(Person Ability Versus Item Difficulty)之间出现的机率关系^[8]。此机率数据成为许多实务应用案例之实证基础，适切地调节教师对个体差异的应对策略。

此外，另一个有关课程本位评量法有效性的关注议题是：教师在为学生之学习能力表现水平评级时，是否存在教师判断的个人偏差(Teacher Bias)的问题，即评级结果比学生的实际学习能力表现水平为高(过于宽松)或为低(过于严格)。这不仅会造成低绩效责任(Low Accountability)的无效数据，还会减少评量工具在课堂计划上的实用性。当评量结果被应用于班级以外，如在不同班级间比较学生学习能力表现水平时，情况将更为严重，关于该评量方法信度和效度的担忧则更甚。鉴于对此项议题有这些绩效责任方面的关注，本文采用等级量尺模式来对融通量表(SCALE)进行验证。模式采用的洛基胜率对数计算(Logarithm Calculations)，不仅将受试者从常模参照标准的限制中解放出来，而且还考虑到教师作为评估员的个别差异性，从而使该个人偏差对评量工具的信度和效度的影响降至最低。

2. 研究目标

本研究目标，在于以等级量尺模式检视融通量表(SCALE)的信效度，验证近期研发的融通量表(SCALE)在评估具特殊教育需要的学生于香港主流课程中的四个主要学习领域的学习能力表现水平时是否达到

其研发目的, 通过统计学检视标准, 能得知其可信及有效程度。

3. 研究方法

3.1. 内容编制

融通课程计划(SAME)之工作队伍, 经过一番讨论, 针对四个主要学习领域(KLA)或各个核心科目中之科目范畴(Strand), 设计出相应的评量步骤。他们交换意见, 互相审视各学习范畴的不同学习能力表现水平之间排序的合理性。而那些被视为误置了的评量题项, 亦被重新评定级别、并被重新排列于更为恰当的序位上。

3.2. 先导性研究(Pilot Study)

受试者来自九所特殊学校的大约 600 名学生在 2010 年 6 月的下学期末接受了融通量表(SCALE)的评量。受试者涵盖了多类残疾状况: 包括视觉缺陷、肢体残疾、及中或重度智力障碍等。按照受试者的智力水平及学习能力表现水平挑选这些受试者(见表 1)。按照其智力水平分程度: 重度(n = 40), 中度(n = 36), 及轻度(n = 29); 按照其学习能力表现水平分等级: 感觉肌能级(A1~A3 = 30), 基础级(A4~A8 = 49), 及初级

(A9~A12 = 29)。

测量四个学习领域中的每一题项, 都按其完成程度计分。1 分代表“完成”; 0 分代表“未完成”。评量数据被记录于具有光学特性的扫描纸上, 然后使用数据管理系统将其转为电子表格之数据文件。再将电子表格数据文件导入 WINSTEPS 软件 3.70.0 版本^[9]进行数据处理。WINSTEPS 利用简单矩形数据来计算有关受试者对试题的反应机率关系数据。

评分者的培训导师是参与融通课程计划的资深教师, 参与了整个融通量表(SCALE)研发过程的研究。他们除了按照评量指引培训评分者, 也参与了该学习能力水平表现之评估过程。

数据分析程序采用之等级量尺模式, 是一个计算反应可能性的数学模式。有关学生之学习能力的信息被具体地分布于测试题项上, 依据受试者个人能力差异的逻辑函数判断答案正误的机率。从数学公式中可以看出, 相对于题项难度而言, 个人能力越强, 正确地答对题项的可能性就越大。若个人能力刚好等同于题项难度, 那么答对题项的机会率则有一半^[10]。等级量尺模式被用于分析评量工具之可信度, 告知教师如何为需要不同指导方式的有特殊教育需要学生编制今后的指导活动。

Table 1. Distribution of the learning attainment levels, intellectual ability levels and the number of pupil subjects across all participating schools
表 1. 受试学生样本的学习能力水平、智能程度及人数之分布

智能	学校	学习能力水平												班级总分	
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12		
重度	S1	3	3	2	1	1									
	S2	3	3	2	1	1									
	S3	2	2	3	2	1									
	S4	2	2	3	2	1									
中度	M1				1	2	2	2	1	2					
	M2				2	2	3	2	1	1					
	M3				1	1	3	3	1	1					
	M4*				0	0	1	2	0	0					
	M5				0	1	1	1	0	0					
轻度	L1								3	2	5	3	4		
	L2*								2	2	3	3	2		
总人数		10	10	10	10	10	10	10	9	9	8	6	6	total = 108	

*M4 and L2 是来自同一所服务中轻度智障学生的学校。

等级量尺模式的适合度统计数据(Rasch Fit Statistics), 是用于检测观察数据与等级量尺模式的预期数据匹配的适合程度。题项的内聚统计数据(Infit Statistics)反映了这些题项均在学生的学习能力范围之内的适合程度, 偏方误(Outfit Statistics)统计数据则显示测量结果中发现偏离学习能力范围的题项。这些题项对学生而言可能是太难或是太简单^[11]。各题项的适合度(Goodness-of-Fit)通过观察值与期望值的比率表示, 例如“均方值(MnSq)”与“标准化 Z 值(ZSTD)”的比率。高于 1.4 或小于 0.6 的均方值, 或超过 2 或小于 -2 的标准化 Z 值的绝对值通常被视为不匹配性(Misfit)之标准^[12]。

“点数测量相关系数(Point-Measure Correlation Coefficient)”, 指的是每题项得到的点数(分数)与评量下该构念(Construct)的总分之间的相关。本研究的构念正是学生在每个学习范畴内的学习能力表现水平。理论上, 该水平值应处于 -1 和 +1 之间。若水平值距离 +1 越近, 显示题项的分数与总分之间关系越紧密。负值则意味着受试者在某测试项上的分数较高, 但与总测试分数较低出现不一致之结果, 因此该测试题项有可能会被删除。如果有很多题项的分数测量相关系数在 0~0.4 之间, 则显示该测量工具可能评估超过一个构念向度(Dimension)^[13]。

等级量尺模式的“个人/题项信度(Person/Item Reliability)”指的是由融通量表(SCALE)测量的“个人/题项排序(Ordering)”的一致性。“个人/题项分离指数(Separation Indices)”显示的样本分布, 是以融通量表(SCALE)测量误差数值显示。分离指数越大, 显示该测量方法越精确, 信度越高^[11]。

为了检视融通量表(SCALE)题项对于反映构念之单一构念向度(Uni-Dimensional)程度, 可用“主成分分析法(Principal Component Factor Analysis”, 简称 PCA), 来分析其“残余差值(Residuals)”。当题项/个人互动的残余差值呈随机分布, 且与由残余差值负责解释的变量少于 5%时, 融通量表(SCALE)就可称具单一构念向度^[14]。

4. 研究结果

4.1. 等级量尺 Rasch 模式匹配适合度

对主要学习领域或核心科目而言, 融通量表

(SCALE)具有非常细微的个人差异值, 从 13.28~15.66, 及非常高的个人信度值, 从 0.85~1.00(见表 2)。同样地, 题项分离值也非常细微, 从 8.56~10.70, 及极高的题项信度值(Item Reliability), 从 0.93~0.99。这些统计数据显示, 融通量表(SCALE)能将学生分为 13 个不同的学习能力水平级别, 且题项难度范围足够将学生能力分为 8~11 个学习能力表现水平级别。所有融通量表(SCALE)题项分数之相关系数, 从 0.86~0.99, 均无负相关。这显示所有的题项分数都与预期总分具非常高的相关性, 是融通量表(SCALE)单一构念向度的有力证据。也就是说, 融通量表(SCALE)仅仅测量了学生的学习能力表现水平, 而不是测量他们学习的其他方面。

对于个别学习范畴而言, 所有的范畴统计数据都显示融通量表(SCALE)存在着很高的个人差异值和题项分离值, 除了语言科目的“阅读”和“写作”范畴、数学科目的“数据管理”范畴、个人、社会与人文教育科目的“文化与遗产”范畴, 及科学科目的“地球与世界”范畴外, 各学习范畴皆出现多于 6 个学习水平之分离值(见表 2)。

4.2. 构念向度

融通量表(SCALE)测量的四个主要学习领域的构念向度进行了题项/个人残余差值的主成分分析法检验^[8]。以特征值(Eigenvalue)为单位, 在观察到的 1787 个总原始变异单位(Total Raw Variance Units)中, 有 1513 个变异单位(即 84.7%)可被归因于语言科目主要学习领域; 在观察到的 1633 个总变异单位中, 1337 个变异单位(即 81.9%)可被归因于数学科目主要学习领域; 在观察到的 2516 个总变异单位中, 有 2069 个变异单位(即 82.2%)可被归因于个人、社会与人文科目主要学习领域; 最后, 在观察到的 2244 个总变异单位中, 有 1914 个变异单位(即 85.3%)可被归因于科学科目主要学习领域。融通量表(SCALE)测量的所有四项主要学习领域, 在没有显著残余差群的情况下都表现出极高的单一构念向度。

4.3. 题项之不匹配性

在所有融通量表(SCALE)题项中, 发现了极少的题项不匹配现象, 在全部 1355 个题项中, 仅有 104 个题项(7.6%), 与预期之等级量尺 Rasch 模式不匹配。

Table 2. Rasch summary statistics on the SCALE
表 2. 融通量表(SCALE)的等级量尺模式分析统计数据摘要

量表 (学生人数 n = 108)	题项 数目	受试者 分离指数	受试者 可靠程度	题项 分离指数	题项 可靠程度	点数测量 相关系数
核心科目	科目范畴					
语文教育(L)	280	14.5	1.00	10.70	0.99	0.95
	聆听(L1)	98	7.92	0.99	10.38	0.92
	说话(L2)	101	12.55	0.99	13.98	0.93
	阅读(L3)	37	2.45	0.86	3.60	0.93
	写作(L4)	44	3.88	0.94	5.86	0.97
数学教育(M)	297	13.28	0.99	9.95	0.99	0.93
	数与代数(M1)	106	9.72	0.99	10.99	0.99
	形状、图形与空间(M2)	109	8.07	0.98	10.37	0.99
	数据处理(M3)	82	5.94	0.97	7.75	0.98
个人、社会 与人文教育(P)	447	15.59	1.00	8.56	0.99	0.93
	个人发展(P1)	88	5.49	0.97	7.63	0.98
	时间概念(P2)	82	8.84	0.99	11.09	0.99
	文化与遗产(P3)	29	3.27	0.91	5.25	0.96
	地域及环境(P4)	82	6.83	0.98	9.32	0.99
	经济及资源(P5)	83	8.02	0.98	9.31	0.99
	社会及公民(P6)	83	6.25	0.98	8.46	0.99
科学教育(S)	331	15.66	1.00	8.97	0.99	0.92
	生命及生活(S1)	44	4.49	0.95	6.16	0.97
	物料世界(S2)	84	7.64	0.98	8.29	0.99
	能量与变化(S3)	95	7.97	0.98	8.09	0.98
	地球与太空(S4)	30	2.40	0.85	3.64	0.93
	科学、科技与社会(S5)	77	7.05	0.98	7.62	0.98

这些题项包括：语言科目有 18 项(6.4%)，数学科目有 25 项(8.4%)，个人、社会与人文教育科目有 31 项(6.9%)，及科学科目的 30 项(9.1%)。研究发现它们主要聚集在重迭区域，即学习能力表现水平于 A9~12 领域，相当于主流课程的小一水平。

5. 讨论

一个理想的评量工具的设计，主要围绕于一个关键课题 - 如何运用评量来提升学与教？一个理想的评量架构应该具有什么样的关键元素来支持教与学？在该评量架构下，又如何让学生准备好成为终生的学习者？课程评量模式在共通连续标准在线连接

了特殊教育与主流教育课程，也提高了学生学习能力，使学习者的自主学习能力透过评量与学习的整合而有所增加^[10]。

那么，什么样的评量最适合用于测量智障学生的学习能力表现呢？其关键源于主流课程的评量方法。也就是说，参照主流课程标准来校准该评量方法。如果某个评量方法未能与主流课程标准连接，那么教师需要使用不同的评量标准来各自衡量有特殊教育需要的学生与一般学生之学习能力表现，及比较两者之间的学业表现差异。这样一来，教师并没有共通的语言来参照主流课程标准，好使教师能够准确地区分有特殊教育需要学生的学习能力表现。因此，教师通常

对有特殊教育需要学生能够达到的一般主流学习水平的期望通常都会有所偏差,或会偏高、或会偏低^[15],这将影响到日后教师为有特殊教育需要的学生选择适切的授课内容^[6]。一个能够针对所有学生的共通课程评量方法,其全纳性(Inclusivity)在不同学习水平上都须有所体现;它让主流学校教师能够区分所有学生的学习成果,并且能为具有不同学习能力表现水平的学生设计出适切的教学活动,从而迈向适合该学生学习能力表现水平的学习目标。这种共通课程评量模式具有绩效责任,保证了所有学生,无论是否有特殊教育需要,都具有同样的权利来学习同样的课程^[6,16]。

一个理想的通用式课程体系,容许学校为其所有学生量身打造课程,不管哪是一所主流学校或特殊教育学校。事实上,许多国家仍然采用各自为政的校本课程体系模式,各个学校按照校本课程安排教学,以校本独特的授课方式完成课程目标。这种作法有两种结果:其一,特殊学校的学生被拒于教育文化与课程的大门之外,未能与他们同龄的伙伴一样可以自主学习;其二,由于学校之间不存在着对课程内容性质的通用鉴别标准,故此教师也未能站在同一个共通交流平台上,相互沟通如何评估其学生的特殊教育需要及提升他们的学习水平^[3]。

过去,曾有呼声要求在某个共通语言架构下为课程、指导活动及评量建立一个协调性质的校准方案(Coordinated Alignment),依此来决定教学和评量内容^[17-19]。因此,一个全纳式课程及其相应的评量过程和指导活动,均需要与主流课程保持一致性,这点是非常重要的。

6. 应用建议

在学生层面上,融通量表(SCALE)提供各个学生历年学习档案和学习进展的信息。一个学生的能力表现档案,包括各科目各范畴内的学习表现,可以被观察到并被记录下来,从而可对各个学生进行科目比较。教师、家长和学生能够清楚地看到哪一科是该学生的强项,哪一科是弱项。学生的表现可以横截面式(Cross-Sectional)地一次过被记录下来,也可以纵向式(Longitudinal)地多次记录,从而追踪该学生在一段时间内的学习进展(见图2)。个别学生表现差异的因素及其需要可被识别出来。然后,各个持份者可决定下一个行动方案,例如决定是否投入更多的资金来发展弱

项科目,或进一步提升学生的强项科目。此外,学生学习能力表现的逐年进度表现评量,有助于决定学生是否是在其学习水平范围内被测试,以及学生学习上是否有倒退现象。无论是哪种情况,教师必须根据评估结果,为已提供给学生的学习机会作反思,并留意是否需要更新学习目标、相关学习活动、指导策略等,藉以提高学生来年的学习能力表现。

在班级层面上,融通量表(SCALE)为教师提供了学生之学习能力水平资料,教师根据这些评估结果来设计授课内容。一个全班的学习能力水平分布图(见图3)能让我们理解到班内所有学生学习能力处于同一条连续标准在线作比较的概念,清晰地展示了班内所有学生在各科不同的学习反应。某个学生表现可与同班的其他同学作比较。教师同样可按照主流课程科目来衡量学生的水平,从而确认班级学习效率^[20]。如果学

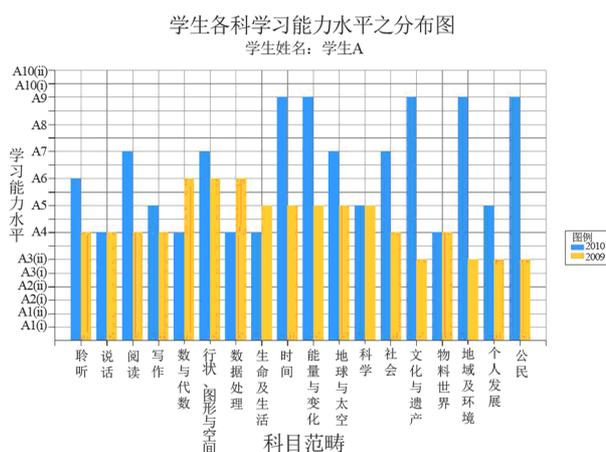


Figure 2. A comparison example of an individual pupil's learning attainment levels in two consecutive years
图2. 个别学生两年内各级学习能力水平之比较

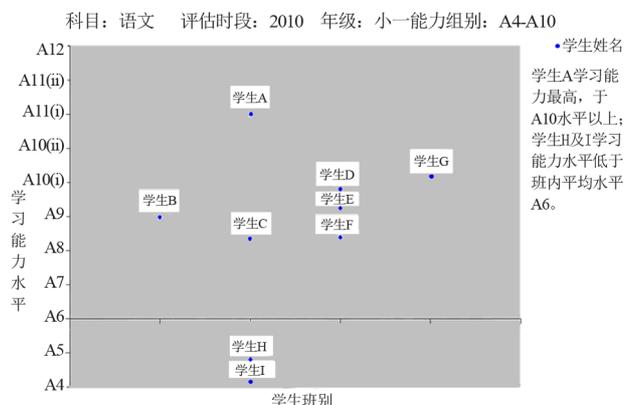


Figure 3. Distribution of attainment levels of students in an example of a Primary One class
图3. 小一班学生学习能力水平之分布

生的学习能力表现水平数据欠佳,学校科目小组可以征询不同专业人士的意见,衡量教师的专业水平是否到位,以及学校里是否有充分教学材料来支持该科目的教学。

在学校层面上,所有学生在学校内的测试分数可被整理出来,班级教师可透过分数数据来比较各班学生的学习能力表现水平。并可透过与不同年度评估结果之比较来衡量学生的学习能力强项^[20]。学校可多途径地使用评量结果来自我衡量学校的发展计划。如某科目需要优先投放更多教学资源,教师需要何种专业培训来提高的教学水平等等。这种课程管理计划不仅有助于提升学生的学习能力表现水平,而且还能提高整体学校的运作效率。

在区域层面上,融通量表(SCALE)提供同区域内各所学校之学生学习能力表现水平资料。同区域内,教师可以追踪学生上一年的学习进展来评量他们的学习能力表现是居50%水平之上或是之下。当一位就读特殊学校的学生大部分科目之学习范畴的学习能力表现均持续于50%水平以上时,校区行政负责人则应召开个别化教育计划会议来讨论是否应鼓励该生返回主流学校里去就读。

校区可比较各学校之间学生的学习能力表现水平分布,作为评审学校的课程指标,并为该校与其他学校同水平学生之间的比较提供了参考。此处,亦可依据共享之学校发展蓝图,更系统化地执行区内课程发展计划。同样地,学校可根据融通量表(SCALE)所提供之客观评量数据,为该校在其所属校区的表现作学校运作效率之绩效责任证据,即是该校是否确实提升了学校教学表现水平。

7. 研究限制

本研究的样本,仅限于特殊学校里的学生,并未包括主流学校里的智障学生。在未来的类同研究中,建议包括主流学校中的学生,包括有智能障碍及没有智能障碍之学生。

尽管融通量表(SCALE)的信度和效度检定数据良好,但仍然有改善空间。例如对那些配适度较低之题项,尤其是位于主要学习阶段一(Key Stage One)的初级层级之重选部分的不匹配题项,可进一步编改,然后再重新进行信效度检定测试,务求将融通量表(SCALE)与主流评量标准接轨得更完美。

8. 结论

本研究透过等级量尺模式验证法去检定融通量表(SCALE)的信效度。藉此证明要发展具信效度之课程本位评量是可能的,但所面对的挑战是严峻的,在应用于教与学之需要上却是合理的。事实上,世界上所有的教育者都面临同样的压力-他们须提供能通过科学监证的学生表现数据数据。使用融通量表(SCALE),可大大提高课程本位评量之绩效责任和维持平等教育参与,为较传统评量法夸进了一大步。

我们未来能有更多类似的尝试,把课程、评量与教学活动连接及校准,从而为有不同学习需要的学生提高教学质量,透过设计更多“通用课程”(A Common Curriculum for All),真正实践平等教育之理念。

9. 致谢

此研究乃由香港大学属下之融合与特殊教育研究发展中心与融通课程计划(SAME Curriculum Project)研究委员会赞助及支持。

参考文献 (References)

- [1] D. Browder, F. Spooner, L. Ahgrim-Delzell, C. Flowers, B. Algazzin and M. Karvonen. A content analysis of the curricular philosophies reflected in states' alternate assessment performance indicators. *Research & Practice for Persons with Severe Disabilities*, 2004, 28(4): 165-181.
- [2] The SAME Curriculum Project. Key learning area curriculum guide supplement. Centre for Advancement in Special Education, Faculty of Education, The University of Hong Kong, Hong Kong, 2008.
- [3] K. Humphreys. Developing an inclusive curriculum: "Every teacher matters". *International Journal of Whole Schooling*, 2009, 5(2): 43-54.
- [4] A. M.-C. Li, A. C.-Y. Tse and M.-G. J. Lian. The SAME Project: A Hong Kong experience in enhancing accessibility of the central curriculum. *Hong Kong Special Education Forum*, 2008, 10: 1-15.
- [5] K. Jametz. *The instructional demands of standard reform*. Washington DC: American Federation of Teachers, 2003.
- [6] R. F. Quenemoen, A. Carmilla, M. L. Thurlow and C. B. Masanari. Pupils with disabilities in standards-based assessments and accountability systems: Emerging issues, strategies, and recommendations (Synthesis Report 37). Minneapolis: University of Minnesota, National Center on Educational Outcomes, 2001. <http://eric.ed.gov/PDFS/ED452654.pdf>
- [7] H. Beilin. Piaget's contribution to developmental psychology. *Developmental Psychology*, 1992, 28(2): 191-204.
- [8] T. G. Bond, C. M. Fox. *Applying the Rasch Model—Fundamental measurement in the human sciences*. 2nd Edition, London: Lawrence Erlbaum Associates Publishers, 2007.
- [9] J. M. Linacre. WINSTEPS Rasch measurement software. Chicago: WINSTEPS, 2006.

- [10] M. M. C. Mok. Self-directed learning oriented assessment: Assessment that informs learning and empowers the learner. Hong Kong: Pace Publishing Limited, 2010.
- [11] C. W. Chien, T. G. Bond. Measurement properties of fine motor scale of Peabody developmental motor scales. 2nd Edition: A Rasch analysis. *American Journal of Physical Medicine & Rehabilitation*, 2009, 88(5): 376-386.
- [12] W. P. J. Fisher. Measurement-related problems in functional assessment. *American Journal of Occupational Therapy*, 1993, 47: 331-338.
- [13] M. M. C. Mok, Y. C. Cheng, P. J. Moore and K. J. Kennedy. The development and validation of the self-directed learning scales (SLS). *Journal of Applied Measurement*, 2006, 7: 418-449.
- [14] R. M. Smith. Rasch measurement models: Interpreting WIN-STEPS/BIGSTEPS and FACETS output. Chicago: MESA Press, 1999.
- [15] E. K. Kontu, R. A. Pirttimaa. Teaching methods and curriculum models used in Finland in the education of pupils diagnosed with having severe/profound intellectual disabilities. *British Journal of Learning Disabilities*, 2011, 38(3): 175-179.
- [16] M. M. Abell, D. K. Bauder and T. J. Simmons. Access to the general curriculum: A curriculum and instruction perspective for educators. *Intervention in School and Clinic*, 2005, 41(2): 82-86.
- [17] A. Kurz, S. N. Elliot, J. H. Wehby and J. L. Smithson. Alignment of the intended, planned, and enacted curriculum in general and special education and its relation to student achievement. *The Journal of Special Education*, 2009, 44: 131-144.
- [18] P. R. Parrish, R. A. Stodden. Aligning assessment and instruction with state standards for children with significant disabilities. *Teaching Exceptional Children*, 2009: 46-56.
- [19] L. B. Resnick, R. Rothman, J. B. Slattery and J. L. Vranek. Benchmarking and alignment of standards and testing. *Educational Assessment*, 2003, 9: 1-27.
- [20] K. Humphreys. Empowering schools in planning for effective learning diversity with SCALE (Feature article of keynote presentation at the SCALE launch seminar presented at the International Conference on Special Education). *CASE News*, 2010, 5(7): 4-7.