

從澳門某所中學的科學教育實踐案例 探討學校多元評量實施的過程

麥瑞琪^{1,*} 張國祥² 楊文佳¹ 薛寶嫦²

¹澳門大學 教育測驗與評核研究中心

²澳門大學 教育學院

摘要

本研究旨在以澳門一所進行校本科學課程開發且以素養教育為導向的中學為研究案例，探討素養導向之科學教育多元評量的關鍵特徵，以及在現有的評量基礎上如何基於評分規準(rubrics)的評量設計，來驗證學習指標與學生基本能力的評量方法和具體步驟，使能客觀地評價學生之學習進程是否達到課程標準所訂定之學力水平，從而為改進學校教學、完善教育政策提供依據，並豐實學校多元評量的內涵。同時，通過研究案例學校實施多元評量的過程，本文分析歸納出教師實施多元評量時須兼備「量」和「質」的特點，即評量要能重視對師生雙方的回饋、具有前瞻性、可為學生學習架設鷹架，這樣才能促進學生更好地往下一學習階段發展，終極目的是邁向學習成功。

關鍵詞：Rubrics評分規準、多元評量、科學教育、核心素養、學習進程

壹、緒言

進入21世紀，世界各地視培養學生核心素養或21世紀關鍵能力(21st century competencies)為重要的教育目標(張華，2016；國家教育研究院，2014；褚宏啟，2016；蔡清田、陳伯璋、陳延興，2014；Rychen & Salganik, 2001)，各個教育系統透過學習內容、教學方法及學習評量三者的綜合運用，重塑各個學科課程內涵與核心素養的關係。其中科學教育一直是學校教育中的重要領域，一方面，科學能力的習得有助學生在未來取得成功，對各國人材儲備和中長遠發展起正面作用；

另一方面，科學學習更是培養學生核心素養的搖籃。

值得注意的是，由於資訊通訊科技(Information Communication Technology, ICT)日趨成熟，以及科學、科技、工程和數學(Science, Technology, Engineering, and Mathematics, STEM)教育在國際上的急速發展，新世紀的科學教育目標，從基本的教育理念、價值觀和全球知識經濟與資訊科技社會的需求著眼，而做整體的考量和通盤的規劃(郭重吉，2012)。近年來，一些先進國家或亞洲鄰近地區，包括中國內地、美國、英國、澳洲、日

*通訊作者：麥瑞琪，skmak@um.edu.mo

(投稿日期：民國109年6月30日，修訂日期：民國109年10月22日，接受日期：民國109年10月26日)

本、韓國、香港、新加坡等因應社會需要的轉變，陸續針對科學與科技在當今人類社會與個人生活中所擔當的角色而革新科學教育的目標(郭重吉、張惠博，2005)。因此，當前世界各國課改的焦點，正是積極推展科學教育以培育科技人才和提升全民科學素養，強調運用科學資訊解決問題和處理數據，並做出合宜決策的能力(Thier & Daviss, 2002)。

澳門教育當局為了適應澳門社會和世界科學技術的發展需要，並在世界科學教育改革浪潮的影響下，於2014年開始逐步實施《本地學制正規教育課程框架》(以下簡稱《課程框架》)(澳門特別行政區政府，2014)及《本地學制正規教育基本學力要求》(以下簡稱《基力要求》)的行政法規(澳門特別行政區政府，2015)。在《課程框架》及《基力要求》的架構下，澳門學校依據基礎教育的學習目標，以及世界各地科學教育發展的趨勢而進行課程改革，以培養學生的科學素養作為核心，貫徹以科學探究作為串連各個學習階段並整合各個科學領域的重要手段，學生透過有系統的探究活動，發展所需的科學知識和技能，以幫助他們評鑑科學與科技發展帶來的影響。與此同時，澳門教育當局也意識到傳統的紙筆測試，重視背誦及低階認知性評核的模式已不能配合未來社會培育人才的需要，於是在2016年底推出了《本地學制正規教育學生評核制度》的諮詢工作，指出現今教育應當以重視培養學生的高階思維能力、解決困難能力、人際技巧及協作能力為重要目標，提倡這些教育目標需要透過多元評量方式才能對學生做出適當的評估；希望透過多元評量，學校得以全面評估學生，從而提升學生的學習成效，逐步降低學校留級率，促進學生學習成功(教育暨青年局，2017)。

然而，學校要系統化整合科學素養的培

育目標，就需要確定科學素養核心概念在不同年級或學段的授課內容和程度，解決這一問題的重要工具是借助先進的評量手段設計並構建中小學生對不同科學概念的學習進程(learning progression)模型，以促進學生科學素養的連貫發展(李川、劉克文，2016)。事實上，評量學生是課程實施的一個重要環節，亦是學習與教學循環中的重要部分，它必須配合學與教的過程(課程發展議會，2002；Black & Wiliam, 1998, 2006, 2010; Earl, 2003, 2013)。

循此，本研究旨在以澳門一所進行校本科學課程開發且以素養教學為導向的中學為關鍵案例，研究學校如何開展多元評量的實施。選擇這一案例將使我們看到學校在大學教育評核研究單位的指導下，如何處理「最有可能被實施好的」評量策略，進而啟示澳門學校在日常的課堂中，透過促進學習的評量元素，以評量結合教學，提升學與教的品質。

貳、文獻探討

一、科學素養與科學教育整體課程規劃目標

科學素養為一般民眾生存於當代科技社會中不可或缺之基本知識與能力。Yore, Pimm與Tuan (2007)認為科學素養乃涵蓋個體的認知、後設認知、推理／思維、運用特定科學用語的能力、情意態度，以及使用ICT的能力和策略等層面，這些素養元素與個體適應現實生活更是息息相關。國際學生能力評估計畫(Programme for International Student Assessment, PISA)亦定義科學素養為一種能力，指個體以一位反思公民的身分，參與跟科學相關的議題並具有科學的見解(Organisation for Economic Co-operation and Development [OECD], 2016)。換言之，具有

科學素養的人能夠閱讀與科學相關的一般報導；能與他人就科學性的議題進行溝通，並獲致有效的結論；且能夠發問、發現或決定問題的解答，而這些問題往往源於個人對日常生活事物的好奇(National Research Council [NRC], 1996)。

現時，因應21世紀知識經濟和資訊科技時代的需求，許多國家課程規劃和學者強調基礎教育的科學教育目標，乃是培養學生的科學素養。比如中國教育部《基礎教育課程改革綱要(試行)》提出學校科學課程的實施要宣導學生主動參與、樂於探究、勤於動手的能力，培養學生蒐集和處理資訊、獲取新知識、分析和解決問題，以及交流與合作的能力(劉堅，2013)。美國教育部指出科學教育包含對物質科學、生命科學、地球科學等領域學科內容的理解；同時，個人亦須理解科學的本質、科學社群的活動，以及科學、社會與個人生活之間的角色；強調科學教育的目的是讓人獲得從事決策、參與公民社會及文化事務，以及經濟活動中所須具備的科學概念與過程之相關知識與理解(NRC, 1996)。當中，美國更把科學和工程整合在幼稚園至高中三年級《新一代科學教育標準》裡，對科學素養的培養強調「科學反思和社會應用」(NRC, 2012)。日本透過立法《科學技術・イノベーション基本法》(即《科學技術基本法》)，訂定科學教育學習綱領及國家施政計畫，整合不同領域、機構與行政層級間科學教育目標的一致性與相互統合，以確保科學教育政策的落實與資源的有效配置；新加坡則以培養資訊與思考創新能力作為國家科學教育政策推動之重點(郭重吉、張惠博，2005)。香港基於國際研究和當地調查所反映香港學生在科學上的實力，修訂了香港《科學教育學習領域課程指引(小一至中六)》，強調要讓學生瞭解科學的本質和掌握科學過

程技能，掌握綜合和應用跨學習領域的相關知識與技能的能力，重視STEM之間的關係，以促進學生成為21世紀更有效的終身學習者，為當今科學和科技世界做出貢獻(課程發展議會，2017)。上述各國科學教育政策與改革的動向，固然各有其不同的社會文化與教育背景，但是具前瞻地審視當前科技與社會發展的趨勢，比如：知識社會所重視的創新能力、科技社會中科學教育對象的全民化、課程內容和教學方法要與生活相結合，以及培養學生學習科學興趣與終生學習的能力等，各國改革的方向則是相當一致的(郭重吉、張惠博)，即科學教育應以培養學生具有創造力和理性思考的能力為目標。這些變革既成為當前中小學科學教育的研究熱點，同時也呈現出科學教育發展的未來趨勢(李川、劉克文，2016)。換言之，提升學生科學素養正成為國際和國家科學教育改革的主流與發展趨勢。

靳知勤(2007)指出要提升臺灣國民的科學素養，需要培養學生對科學的好奇、興趣、態度，以及人文價值觀；強調教學內容可藉實驗、動手操作及媒體教學，讓學生能夠體驗動手操作與探究，思考蘊含於學習歷程中的道理(know-why)與方法(know-how)，使學習內涵具體化，從而發展解決與生活實踐結合問題的能力。李春密與趙芸赫(2017)梳理美國、芬蘭、韓國、新加坡、英國、法國、中國，以及加拿大安大略省、澳大利亞維多利亞州等基礎教育階段STEM相關學科課程內容，建議將跨學科概念融入中國的科學課程設計，建立各學科內及學科間的有效聯繫，以表現期望的形式呈現知識內容，培養學生通過理性客觀的方式解決問題，能對社會生活中的科學、技術和環境問題有批判性的認知及合理的判斷能力。

從以上各國的科學教育整體課程規劃目標可歸納出世界各地科學教育發展的趨勢，

皆以培養學生的科學素養作為核心，學生透過有系統的探究活動，發展所需的科學知識和技能，以幫助他們評鑑科學與科技發展帶來的影響。學生除了具備執行科學探究的能力外，還要能反思科學探究的意義和科學探究的趨勢是什麼，即認識科學的本質，養成科學思維的習慣，以及掌握科學過程技能。

澳門與其他國家或地區一樣，正面臨著全球化時代學生對傳統學科教育不感興趣的窘境，因此學校課程有必要朝向更為積極且以學生為中心的方向發展。根據澳門現時的法律法規，澳門學校在《課程框架》及《基力要求》的架構下，可以依據自身辦學特色，開發各自的校本課程(澳門特別行政區政府，2006)，課程規劃的相對自由度讓教師有更大的空間來因應教育目標、教育理念和學生的需要，以學校的教育情境和資源為基礎，進一步規劃、主導設計科學課程，彈性剪裁教材，為澳門學校實施科學教育改革提供了十分有利的條件。

二、多元評量在素養為本課程的定位

可以想像的是，課程改革往往是人們的焦點，相對來說，評量改革卻未必得到相應的重視。從學生為本的角度來看，評量活動不應該只是簡單地讓學生展示學習成果，而是要結合於學與教之中。不同的評量方式各有所長，可以有不同的深度或廣度，以反映不同的學習面向。在教學現場裡，一系列不同模式的評量能讓大量的不同學習成果得以評估，應根據教學目的而採取靈活的評量模式。但無論如何，評量最主要的目的還是為了幫助和改善學生的學習(Black, Harrison, Lee, Marshall, & Wiliam, 2004)。進一步來說，評量與教學是一體兩面，評量不但要能

回饋教師的教學效能，也同時要兼備回饋學生的學習效能，所以評量並不獨立於課程外，反而是與課程及教學緊扣在一起(房湘雲，2013；Herman, Aschbacher, & Winters, 1992; Ramsden, 2003)。學校在部署評量政策改革時，首要工作是理順不同教育評量之間的關係，去除不合理和不必要的評量，合併功能重疊的評量，使教育各個環節與評量能夠有效銜接，讓教育評量能高效地發揮作用，進而系統化地推進教育評量的改革(辛濤，2017)。

近來的評量理論著眼於三大考量：一是評量能力而非只是知識，二是重視學生學習的歷程而非只是成果，三則是確保學生學習成效的水準與質量。這樣的評量趨勢，除有教育心理學理論上的支持，也因應21世紀公民素質的養成，兼顧「質量」與「公平」兩大理念(吳璧純，2013)。因此，學者不但倡議「對學習成果的評量」(assessment of learning)，更進一步發展至「促進學習的評量」(assessment for learning)及「作為學習的評量」(assessment as learning)。事實上，上述三種評量皆與教師所定的教學目標息息相關(Stiggins, Arter, Chappuis, & Chappuis, 2004)，而如何達成教學目標又與學習領域及學生特質息息相關，教師必須透過多次教學實踐的經驗來轉化評量實施原則並學習評量技能，以實施這三種形式的多元評量。

評量多樣化本身並不是目的，評量採用的形式與教師的教學目標和評量內涵密切相關(吳璧純，2017)。林曉雯與陳佩君(2005)探討國小自然科學教師實施多元評量所遭遇的問題和對應策略時，認為教師首應釐清多元評量的目的，準確解讀能力指標，並體認教學目標、課程和評量交互檢核的角色，進而

建構科學探究歷程的教學評量模式。林怡呈與吳毓瑩(2008)指出多元評量要發揮協助學習或教學的功效，就必須緊扣教學目標，否則就會流於空泛而削減支持學習的力量。然而，也有不少教師在接受多元評量相關研習後，只停留在對多元評量特質與內涵詮釋的組合運用，並陷入為多元而多元的迷思之中(曾淑惠，2015；歐用生，2003)。此外，當教師學科教學知識不足，評量規劃便出現貧乏僵化，對於教學無所助益，有時甚至誤導學生學習(林怡呈、吳毓瑩)。因此，不少學者建議教師使用多元評量，應重視學生個人內在的學習改變，瞭解個人與環境的互動關係，著重個體差異而給予適當的評量程序。再者，由於素養導向的評量活動與學生各種學習活動相呼應，教學與評量歷程同時連續累進，多元評量便須要在教師一連串的目標性教學與學生的學習系統中進行規劃與實施，關注學生素養形成的歷程與結果，區分形成性與總結性評量，或善用促進學習的評量、對學習成果的評量和作為學習的評量，以提供學生在學習歷程中所需要的協助和鷹架(吳璧純)。另外，還需注意的是，評量的改革需要新的評分規則和學生學習表現成績報告的方式，且要能正確的反映出所學習的課程的豐富性及多樣性，而非只是打一個分數(張美玉，2012)。因此，學校對學生的評量也應該重新被定位，而且要與新的教學觀點相一致。

三、Rubrics (評分規準)為促進學習成功所起的作用

教師在統整課程、進行教學及對學生施測時，必須留意教學目標是否與能力指標相對應，而且學生能夠藉由有意義的學習經驗而達至某項能力指標。當中，教師除了要顧

及知識與能力之間的連貫和一致外，也須藉由適切的評量工具來瞭解課程與教學所達至的效果，以及學生的學習成效，並以此作為調整教學的依據。

針對各地發展運用核心能力或科學素養界定學生科學學習目標和成效的做法，學者提出運用rubrics (在本文中，後續皆以rubrics來指稱評分規準此方案)來進行評量。Brookhart (2013)指出rubrics可用作評量複雜的學習任務，尤其當預期學習結果需要以表現呈現，當中有不同等級的表現情形，或不同的表現向度。一般而言，rubrics評分量尺包括四個部分：任務描述、表現向度、表現等級和表現描述(Taggart, 2005)。Rubrics的好處在於幫助授課者思考課程內涵的知識層次、應達成的學習目標與外顯行為、達成的能力面向，它除了評量當下的學習狀態外，也提供明確的努力方向(陳琦媛，2017；黃淑玲，2013；Stevens & Levi, 2005)。因此，吳璧純(2017)認為素養導向教學的學習評量工具主要乃依據「評量的基準」(criteria)來判斷學生的表現品質，因此評量工作必須事先建置評分基準表。

另外，教師將rubrics作評量工具時，還可進一步把它量尺化，轉換成學生科學素養精練水平等級，同時將不同科學領域的學習目標進行分類，依據共同的能力指標給予向度名稱，成為rubrics的素養表現向度(Stevens & Levi, 2005)，這些表現向度可簡單且完整呈現學習任務，讓學生能容易地瞭解到自己的學習進展(陳琦媛，2017)，也有利師生共同管理學習期望，有一致的學習成果目標(劉曼君，2014)。然而對利用rubrics評量學生能力的做法是否能達至符合教師當初預期的評量設計框架，上述學者卻未進一步驗證。

綜合上述，評量要有積極意義，該是

一種為促進學習成功而做出的評量，並且是回饋整體教學的重要策略。因此，在素養導向的教育環境下，學校教師需要運用正確、多樣化的評估方式，彼此交叉使用，從而瞭解學生投入學習過程的情形與努力的程度。多元評量的設計重點乃聚焦於發掘學生學習成就，創造出不同的學習及被評量的機會，它正好為師生提供了多樣化的檢核方式。此外，rubrics評量工具的特色是能提供各表現向度之不同等級的詳細描述，有助把學習目標融入不同向度的表現期望中，澄清學習和能力指標的關係。老師透過設定目標基準、利用rubrics評分量尺來檢視、記錄學生在學習過程中是否能達成預期目標；根據學生的學習表現及進步，評量其學習狀況、困難或落差，進而提供必要的協助及引導，以做下一個教學的判斷，協助學生做好學習管理，促進學生學習成功。

參、研究方法

一、研究目的

近年來，澳門學校在執行多元評量的具體操作上，教師普遍感到較陌生，也缺乏相關的專業培訓(澳門中華教育會，2016)。儘管澳門教育當局頒布了《課程框架》和《基力要求》行政法規，然而，上述法規只涵蓋課程必須學習的領域及能力指標，並未涉及對教學成效的具體評估方案，許多教師尚不清楚如何針對《基力要求》的學習指標或能力進行檢核，對於基本學力指標與學生評量之間的關係驗證也不明確。

本文透過對以素養教育為導向的學校作研究案例，向教師展示如何在學校現有的評量基礎上，引入基於rubrics的評量設計來驗證指標與學生能力的評量方法，進而指引澳門教師

在日常的課堂中開展多元評量的具體方案和步驟，以客觀評價學校科學教育的狀況，瞭解學生是否達到課程標準所訂定之學力水平，為改進學校教育教學、完善教育政策提供依據和參考，俾能豐實學校多元評量的內涵，並促進學生往下一學習階段的發展。

二、研究問題

基於上述研究背景和文獻分析，本文的重點即在探討下述兩點：

- (一)以澳門某所學校作案例探討融入素養導向之科學教育的多元評量有何關鍵特徵？
- (二)以澳門某所學校作案例探討教師如何基於rubrics的設計理念建構評量系統以瞭解學生是否達到澳門基本學力和校本科學課程所訂定之素養水平？

三、個案的選擇與背景資料

本文以澳門一所進行校本科學課程開發且以素養教育為導向的中學作為關鍵案例，因此，選擇個案研究方法。個案研究方法允許研究者透過多種資料蒐集途徑，探究案例學校教師如何基於rubrics的設計取向，針對澳門基本學力指標來驗證指標與學生能力的評量方法，從中剖析澳門學校在日常的教學中開展多元評量的具體方案和步驟，以客觀評量學生是否達到澳門基本學力要求所訂定之水平，並豐實學校多元評量的內涵。與此同時，為進一步瞭解運用rubrics驗證學力指標與學生能力的評量方法，研究者報導和分析案例學校教師和當地大學教育評核研究單位合作實施評量的實際情況，希望從中淬鍊出教師的實踐知識與經驗，而得以與實務者和研究者分享。

案例學校是澳門一所創立於1889年且設

有幼稚園、小學和中學學部的私立學校，目前全校74班，共超過3,400名師生，在澳門屬於大型學校。自2016年起，案例學校重組及逐步實施其校本科學課程，把學校常規的科學課程分為共同學習的「核心內容」及「個別化內容」兩部分。其中，共同的「核心內容」主要參照國際科學教育研究的發展趨勢、最新的兩岸四地科學教育課程標準和指引、澳門基本學力要求及學校發展情況後得出，著力於學生科學素養的習得，重視培養學生應用科學知能解決問題的能力；「個別化內容」則讓學生選擇符合自身興趣的科學主題和方向學習，配合分層教學的理念，讓學生在不同教育階段或年級因應自身興趣、能力及未來發展取向選擇學習內容，具有很大的彈性。除「核心內容」外，科學教師可根據自身和學生特點設計課程內容和教學活動。一般來說，符合學校要求和課程架構，能適應學生發展的內容均會被接納並允許施教。為了確保教學質量，中學一年級與二年級各班還須定期進行科學素養評核，以瞭解學生是否掌握課程的「核心內容」，即評量學生在基礎科學內容、科學探究的基本能力、各種與探究相關的高階思維、「基本學力」所列明的能力等之情況。

四、本研究資料蒐集方式

本研究係屬個案研究，因此採用文件資料蒐集、隨堂觀察、訪談和數據分析等多種證據來源產出個案的資料，研究時間橫跨2018／2019整個學年。研究者首先蒐集案例學校的科學校本課程規劃、課程設置、教學檔案與評量政策等文件，主要從檔案中獲取與校本課程開發和科學素養教學相關的具體安排。同時，研究者進入學校現場進行了隨堂觀察，記錄科學教師實施校本多元評量之情況，藉以闡述素

養導向科學教育的多元評量之關鍵特徵。為驗證rubrics的評量架構設計之有效性，研究者透過與學校課程領導人員和科學教師的會面訪談，與案例學校教師們共同梳理校本素養導向的科學課程教學目標，也就是該門課程欲傳授的內容，以及所期待的學生學習成果為何，繼而確定這些知識、技能和態度所對應之澳門基本學力指標，也就是教師所欲評量學生的內容，同時跟據rubrics的架構設計擬定要評量的項目及指標，和評量尺度的表現等級。最後，研究者收集教師於2018／2019學年第一學期針對二年級218位學生進行的「科學素養評量」測試結果，以Rasch模型對評量題目參數進行校準(calibration)，把學生能力和測試題目置於單維度的同一量尺上標定比較，藉以瞭解教師運用rubrics驗證學力指標與學生能力的評量方法，協助教師客觀評估課程實施的整體成效。

五、本研究資料分析方法

本文探討以素養教學為導向的學校如何開展多元評量的實施，以及運用rubrics驗證學力指標與學生能力的評量方法。研究聚焦在學校日常的課堂中教師如何處理「最有可能被實施好的」評量策略，藉以逐步歸納素養導向科學教育的多元評量之關鍵特徵。為釐清基於rubrics的評量設計取向來驗證澳門基本學力指標與學生科學能力的做法，研究者同時採用Winsteps 3.92.1版數據分析軟件以Rasch模型對評量項目參數進行校準，教師根據學生能力和題目難度在同一量尺上的相對位置，得以客觀評估學生是否達到澳門基本學力和校本科學課程所訂定之水平。本文的資料分析主要採取如下步驟：(一)組織各類資料，根據研究問題進行初步分類；(二)閱讀各類原始文本，尋找相互聯繫；(三)根據評量架構和實際評量數據資料建立比對檢證，並歸納總結成果。

肆、研究結果與討論

一、素養導向之科學教育的多元評量關鍵特徵

訪談和觀察是個案研究採用的重要資料蒐集方法之一。為深入瞭解個案學校(下面簡稱A校)的科學課程理念和評量實踐方法的決定過程,研究者對該校課程領導人員和教師進行了重點訪談和隨堂觀察。其中訪談工作共安排了三次,每次時間約兩小時,訪談內容均整理成逐字稿,並經A校課程領導及教師過目校正,以提高研究資料的準確性並作為分析之用。本節乃根據A校校本課程規劃、課程設置、教學檔案與評量政策等原始資料與教師訪談內容整理而成,聚焦於領導層和教師如何在日常教學中根據《課程框架》、《基力要求》及A校校本科學課程的指引,逐步實施課程改革和擬訂二年級科學科課程及多元評量框架,藉以分析素養導向科學教育的多元評量之關鍵特徵。值得注意的是:這種以能力導向和診斷、改進為目的的多元評量,其測量項目的研發思路、方法、計分方式、結果的呈現,皆與傳統考試評價不同(辛濤、張彩,2018)。

(一)素養導向的科學課程設計理念

圖1顯示A校根據澳門教育當局頒布的科學教育《課程框架》和《基力要求》,以及當前世界科學教育目標而設定的校本科學課程,其特點是以發展學生的科學思維和探究能力為重點,著力於培養學生對科學的興趣,鼓勵學生綜合和應用科學知識與技能以解決實際情境下的科學問題,並能根據科學證據做出明智的判斷和決定,成為自主的科學學習者。A校依據圖1之課程規劃,在校本科學課程框架的設置中著意加強學生在生活脈絡中學習科學知識和探究技巧,重視學生科學思維和能力之培養,期望他們透過課程瞭解科學的過程與本質;同時涵蓋了澳門教育當局《基力要求》的四個科學學習範疇,即物質科學、生命科學、地球與太空科學,以及科學探究(圖2)。學生習得一定程度的科學知識後,便可進行科學探究及問題解決活動,探究和解難的內容與現實世界和生活脈絡有關,並涉及物質科學、生命科學,以及地球與太空科學三個範疇。當中的科學思維與能力是指學生解決科學問題過程中所涉及的認知歷程和能力,而相關歷程和能力及其培養目標乃參考國際數學與科學教育成就趨勢調查(Trends in International Mathematics

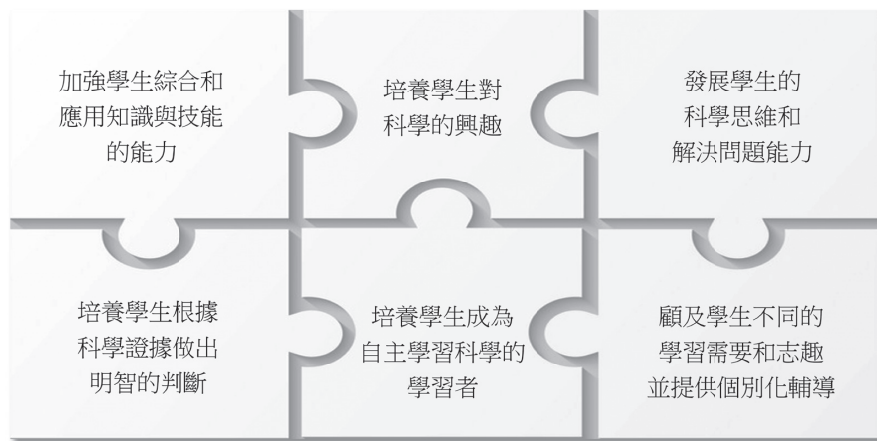


圖1：A校科學教育整體課程規劃宗旨

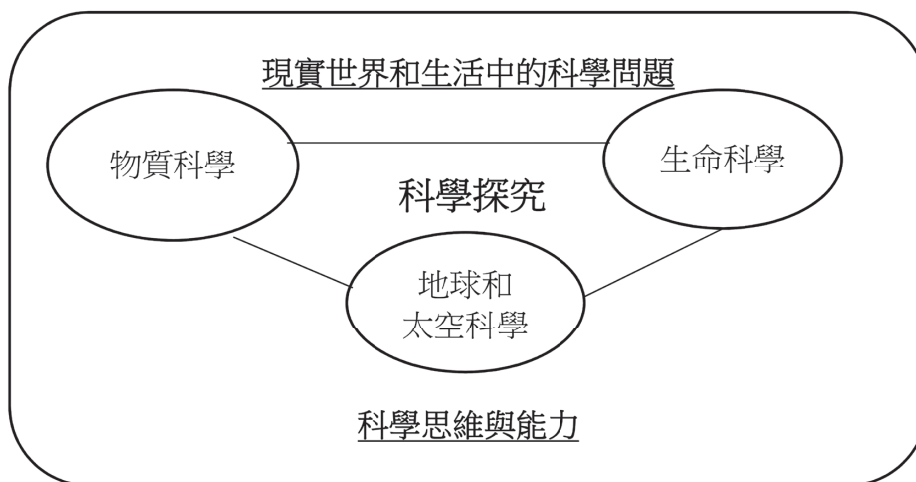


圖2：A校科學課程框架

and Science Study, TIMSS)、PISA的科學素養框架、香港和澳門的科學課程文件，以及根據A校的校本發展需要編訂而成，包括「理解和探究」、「詮釋與應用」、「推理與論證」三個維度，涵蓋澳門「基本學力」對中學科學科相關範疇的學力要求(詳見澳門特別行政區政府，2017; Mullis & Martin, 2013; OECD, 2016)。

誠如前文所述，在素養導向的教育環境下，學校教師需要運用正確、多樣化的評量方式，才能更客觀和有效地瞭解學生學習的過程與投入努力的程度。多元評量的目的是透過多樣化的檢核方式來創造出不同的學習經驗，以及被評量的機會，提供以實證為本的回饋，來反映學生的學習進展，同時幫助教師依照評量結果調整教學設計和規劃課程。另外，為兼顧「促進學習的評量」及「作為學習的評量」的多元評量理念，A校改變過去純以分數高低來評量學生學業成就的做法，更新校本科學課程的教育目標時，也同步啟動校本多元評量的改革，教師依照教學目標設計學習目標，並選擇適合的作業類型、評量方法或學習活動，來綜合評量學

生參與學習後達成學習目標的程度。A校課程領導主管認為：

學校採取證據為本、逐步實施的課程改革策略，在全面推行課程改革至其他年級和學科之前，需要準確地評估初中科學課程改革的效果。課程改革的另一特色，是為不同需要學生提供機會和施展潛能的平臺，也符合國際上關於促進每位學生取得成功的呼籲。此舉對於不擅傳統模式下學習、多動、記憶力不佳、讀寫能力稍遜的學生效果尤佳，有助其找到合宜的學習方向，取得應有的成功經驗。學校倡議「適才教學」，老師採用適合自己的課程和教學，學生獲得適合自己的學習與經驗。當然，為了確保具備基本的教學水平，各級各班均需定期進行素養測試和考試，以瞭解學生是否掌握課程的核心內容，如科學探究的基本能力、各種與探究相關的高階思維、「基力」所列明的能力等。(晤談紀錄，01/12/2018)

(二)素養導向的多元評量實踐框架

A校實施校本多元評量時，考慮需要平衡進展性評量和總結性評量對教學的作用，同時為增加評量回饋學與教的成效，因此重訂學校和教師在校內與課堂中實施專業評量和決策的角色。A校引入校外評量作為參考資訊，協助教師客觀診斷學生的學習和學校的課程質量。整個過程既要得到學校領導層的支持，也需要課程執行團隊的互相磨合，以至家長和學生的充分理解。為了讓校內不同取向的評量可以共融協作，課程改革由學校行政和教學團隊共同定案。A校科學教師總結在課程改革和推動多元評量的過程中，校內各持份者共同參與尤為重要。

開始時，部分家長未能明白新課程的理念，學校通過一系列的講解會和個別跟進以釋除家長的疑慮；在校內，部分任教科學「個別化課程」和「核心課程」老師亦未能馬上適應新課程的要求。難處在於，新課程沒有配套現成的教科書，老師須預備生物、化學和物理等學科的教學內容，也須從日新月異的科學知識技術中篩選合宜的內容整合至運作課程[operational curriculum]之中。過程中，學校的老師需捨棄傳統課程的包袱，不能沿用教科書主導的刻板教學模式，更應摒棄考試導向、分數至上的評核方式。總的來說，雖然缺少教學經驗，但新畢業老師在適應新課程方面具備一定的優勢；通過學校投放的資源和參與校內外培訓，初中科學課程的執行團隊正逐步實施和完善新課程的各個環節。(晤談紀錄，01/02/2019)

圖3為A校實施多元評量的框架，A校以「促進學習的評量」理念連結整個科學課程的進展性和總結性評量活動，而且「對學習成果的評量」偏重總結學生學會了多少，以評定學生的學習進展水平。在總結性評量的技巧上，是在「促進學習的評量」的思路下，採用進展性評量方法來診斷學生的學習困難及監察其進程，以提升學習成效，當中的評量設計均緊扣促進學生的學習為首要核心任務，重視對學與教的發展和調整。值得注意的是A校以素養教育為導向的多元評量特色，是強調學與教過程中對師生雙方的回饋，除評量活動的多元化外，也著意引導多方的參與和校外評量。例如學校藉著參加國際評量持續追蹤科學教育質量，鼓勵學生參加國際和當地科學競賽，希望師生在此類學習活動中吸取經驗，形成良性互動，促進動手實作能力。另外在應用電子化系統方面，A校也將校內聯網e-Class作業系統運用到測評上，並把相關數位系統融入日常課堂的即時隨堂測驗和學習，例如學生應用平板電腦或電腦進行學習，尋找和分析資料，參閱不同文本和課業視頻，以雲端和e-Class作業系統呈交及評改作業和報告；好處是能有系統地記錄評量所得資料及數據，讓教師、學生可隨時追蹤進展，制訂改善學習的方案。

(三)素養導向的實作表現評量

由於A校以科學探究和問題解決的過程為重點，對科學實驗的評量也有別於過往做法。教師採用引導式探究的思路設計實驗課，讓學生在設定的主題下探索和選擇實驗方法，就是學生有既定的實驗目標，在已有資源下通過考慮和比較不同方法，控制不同變量等，設法實現目標。比如研究者於研究期間觀察A校兩位教師的科學實驗課為例，其中小詩老師(化名)以「燃燒食物」為題，開

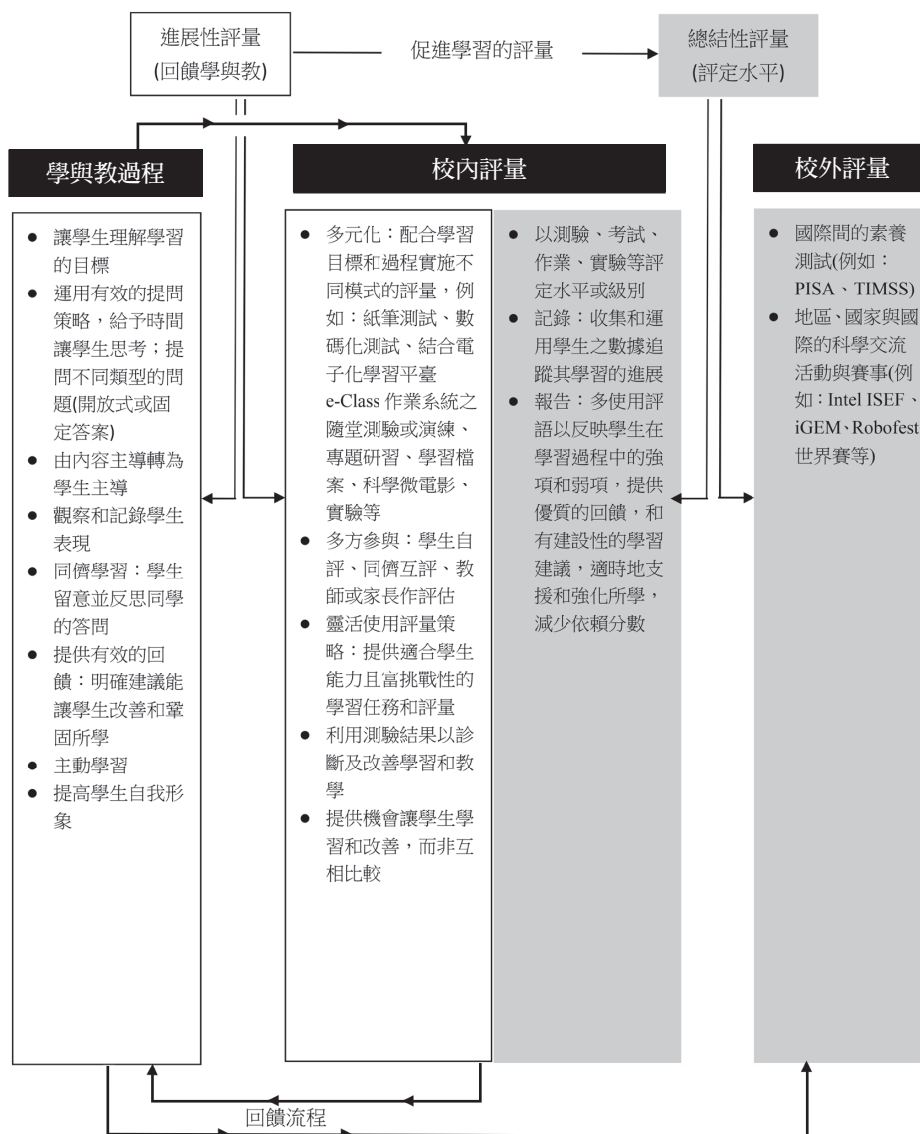


圖3：A校實施校本多元評量的架構

註：PISA：國際學生能力評估計畫(Programme for International Student Assessment)；TIMSS：國際數學與科學教育成就趨勢調查(Trends in International Mathematics and Science Study)；Intel ISEF：英特爾國際科技工程大獎賽(International Science and Engineering Fair)；iGEM：國際基因工程機器大賽(International Genetically Engineered Machine Competition)。修訂自教育局(2017)。評估素養及學校評估政策。查詢日期：2020年5月1日，檢自https://www.edb.gov.hk/tc/curriculum-development/renewal/guides_SECG.html。

展實驗數據處理的探究，她一改學生以往只按老師給予的步驟進行實驗並得出結果，老師檢查數據結果後便完成課堂目標的做法，要求學生充分借助校園的電子化學習系統，

同步使用平板電腦或電腦記錄重覆實驗得出之數據，利用電子表單製成數據表格，以便進一步計算數據，並檢視實驗結果是否可靠，誤差是否過大等問題，以及通過實驗結

果來驗證研究假設。教師亦要求學生以電子方式完成和遞交實驗報告，報告還可以以電子海報形式呈現，這樣的處理能讓不同組別的學生都可互相參考和提出改進意見。又如大雄教師(化名)在講解液體溶解度時，首先講解溶解度的基本概念，其對日常生活的意義和影響，之後通過實驗與理論結合的方式，讓學生自行探究與溶解度相關的若干問題。其中一個設計是以製作棒棒糖為情境，先求出多少白砂糖才可讓容積固定的水達到過飽和狀態；然後蒸發多少水可讓糖結晶；最後讓學生製成一根完好的棒棒糖。此時，教師對學生的實作表現評量，主要集中在學生執行或製作需要高階思維或複雜問題解決的過程，觀察實驗表現和作品呈現的方式；著重審視學生在模擬情境中應用知識和技能的學習過程，從而更具體地指出他們在學習中哪些部分已經達到學習目標和要求；分析其表現中的不足或缺憾，並提出改進的建議。教師指導的重點在於為學生提供描述性反饋而非判斷性的訊息，過程中可以讓學生參與自評和同儕互評，進而學習自我監控學習進度。此外，由於課程設計刻意加強了學生動手的能力，希望學生能在遊戲中學習到科學知識，A校亦善用社區網絡和大學科學研究單位所提供的實驗場地，讓師生進行更加多元化的科學學習活動。

綜合上述的研究發現，A校素養導向的科學教育多元評量關鍵特色在於：1.師生皆理解評量是教與學的一個重要部分，共同回顧和反思有關評量的資料及數據，分享學習的成果；2.重視評量結果的回饋作用，讓學生知道並確認他們所達到的學習水平，以確認如何做下一步學習；3.通過自評和互評的過程，讓學生學習從師生的答問中進行反思，增強自我監控學習進展的能力。A

校這種融入素養導向的科學教育多元評量，與提倡「促進學習的評量」的學者所關注之焦點一致，理念基礎是相信每一個學生都能進步(張淑賢，2014；陳健生、甘國臻、霍秉坤，2010；Black, Harrison, Lee, Marshall, & Wiliam, 2003, 2004)。如同Black與Wiliam (2006, 2010)、Black等(2004)、Earl (2003, 2013)認為不論是「促進學習的評量」還是「對學習成果的評量」，A校所實施的多元評量，其所指向的最終目標是透過評量的過程幫助和改善學生的學習，而非問責、排名或認證能力。

二、Rubrics評量設計在驗證基本學力指標與學生能力關係之關鍵作用

澳門教育當局於2014年開始逐步實施《課程框架》及《基力要求》的行政法規，許多教師不清楚如何針對《基力要求》的學習指標或能力進行檢核，基本學力指標與學生評量之間的驗證也不明確。再者，教師對於現時多元評量中如何貫徹能反映學生學習進展水平的「對學習成果的評量」，亦未有深入掌握。因此，研究者與案例學校之科學教師引入基於rubrics的評量設計來驗證指標與學生能力的評量方法，以進一步豐實學校多元評量的內涵。

Rubrics是與學習目標相呼應的一組指標，依照一定的參照規準以衡量學生在學習上的表現，包括衡量表現的規準(如一個或數個面向)、描述(所要評量的特質的定義、範例、說明)，以及等級(評分的尺度)等。由於rubrics具有較為一致性的規準，因而有利於師生都能掌握複雜的評量規準，一方面學生得以進行自我評量、預先反省，或進行同學互評，另一方面教師也能根據評量的結果，

規劃進一步的教學重點(Little, 2006)。因此，rubrics非常適用於針對《基力要求》的學習指標又或素養為本的能力檢核之用。

為便於瞭解rubrics的具體應用，茲以案例學校教師如何基於rubrics的設計取向建立檢定學力指標與學生能力的評量系統為例，逐一分述步驟和其特徵。圖4為案例學校教師利用rubrics的評量特色設計建構二年級學生科學素養評量系統的整體流程。

(一)決定學科維度與向度內涵及培養目標，聚焦於測量可敘明的指標、表現、行為或量

如上所述，A學校的科學課程乃著重學生解決科學問題過程中所涉及的認知歷程和能力，當中的科學思維與能力包括「理解和探究」、「詮釋與應用」、「推理與論證」三個維度，涵蓋澳門「基本學力」對中學科學科相關範疇的學力要求。因此，教師首先針對A學校中學二年級科學課程之科學思維與能力的教學目標，整理編製各項對應的能力維度，其中，作為示例，「詮釋與應用」之內涵及培養目標詳如表1所示。

(二)按照學科維度及擬評量的向度，編製總體評量規準表和與其對應之基本學力指標，利用間距衡量表現

研究者與科學教師擬訂課程的具體培養目標後，按照陳琦媛(2017)及Brookhart (2013)的建議編定二年級科學思維與能力各維度的評量規準表，即把相應的培養目標、表現等級，以及各等級表現行為之描述轉換為總體評量規準表(general scoring rubrics)，其後教師可對總表進行修訂，以反映具體的課堂學習目標及評量要求。承上續以「詮釋與應用」維度為例，表2為「詮釋與應用」維度中的兩個向度的總體評量規準表，在每個向度下包括三個表現等級，以及相關的等級描述，即學生達到相關等級時應具備的能力和／或能完成之科學活動與行為。

(三)按照規準表設計測驗題目以評量學生科學素養表現

研究者與科學教師同時把上述評量規準表用於學生的總結性評量及進展性評量上。例如以學生在科學考試或「科學素養評量」上的表現作為評定表現等級之參考依據；又

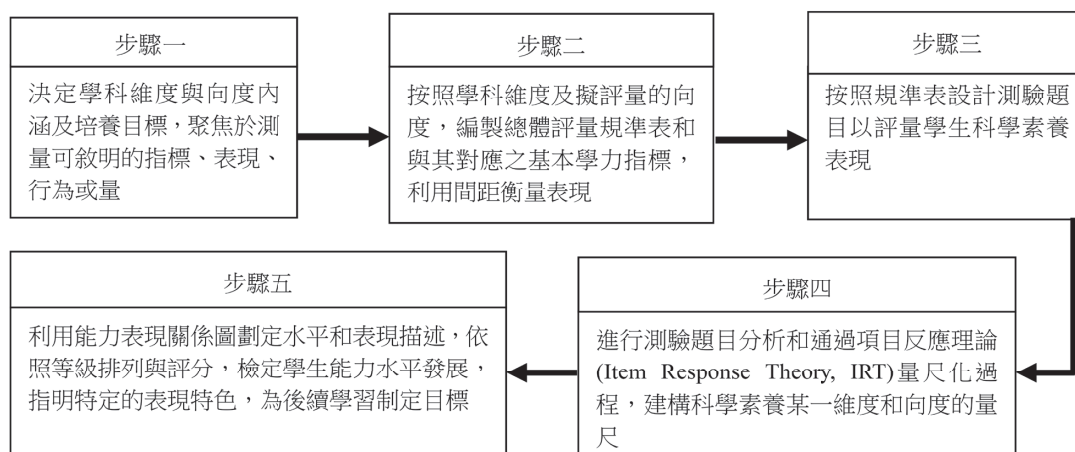


圖4：基於rubrics建構學生科學素養評量系統的整體流程

表1：案例學校中學二年級科學課程科學思維與能力的「詮釋與應用」維度及其六個向度內涵和培養目標

向度	內涵及培養目標
一	能使用合宜的科學知識，並且辨識、使用或生成科學模型，以解釋科學現象和解決現實情境下的科學問題。
二	能使用科學知識來解釋特定文本、表格、圖片和圖形訊息，並能以這些訊息解釋特定的科學現象。
三	能把科學數據和資訊轉換成不同的表達形式。
四	能識別或描述生物、材料、過程或其他特定對象之間的相似點和不同點，能根據給定的特徵和屬性對特定對象進行區分和分類。
五	能按給定的步驟／程序／方法進行科學探究和／或實驗，以完成探究和／或實驗任務並收集相關數據／證據。
六	在科學探究和實驗過程中，能夠合宜地選擇和運用儀器、做好預防措施等。

註：依澳門A校校本發展情況而調整。編訂參考自課程發展議會(2017)。科學教育：學習領域課程指引(小一至中六)。查詢日期：2020年5月1日，檢自https://www.edb.gov.hk/attachment/tc/curriculum-development/kla/science-edu/SEKLACG_CHI_2017.pdf；澳門特別行政區政府(2017)。第55/2017號社會文化司司長批示。查詢日期：2020年6月18日，檢自https://bo.io.gov.mo/bo/i/2017/26/despsasc_cn.asp#56；“TIMSS 2015 assessment frameworks,” by I. V. S. Mullis & M. O. Martin, 2013 (http://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf); *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy* (pp. 26-27), by Organisation for Economic Co-operation and Development, 2016, Paris, France: Author.

表2：「詮釋與應用」維度中向度二與向度三之總體評量規準表及與之對應的學力指標

向度	基本學力指標	表現等級		
		待改進	發展中	已具備
向度二：科學地詮釋資訊及科學現象	A-2-4初步學會對研究資料進行分類與整理，並運用科學術語進行表達	未能使用科學知識來解釋特定文本、表格、圖片和圖形訊息，或時常出錯，並且未能準確地使用這些訊息解釋特定的科學現象。	能使用科學知識來解釋特定文本、表格、圖片和圖形訊息，但偶有出錯，並且未能準確地使用這些訊息解釋特定的科學現象。	能準確地使用科學知識來解釋特定文本、表格、圖片和圖形訊息，並能以這些訊息解釋特定的科學現象。
向度三：科學數據和資訊的表達	A-2-5能初步運用文字、圖表等多種形式撰寫簡單的研究報告	無法把科學數據和資訊轉換成其他的表達形式。	能把部分科學數據和資訊轉換成更直觀易懂的表達形式(例如各式圖表)，但未能轉換較艱深或抽象的科學資訊。	能準確地把科學數據和資訊轉換成多種表達形式(例如各式圖表、簡報等)。

如教師於每個課題結束後皆會進行課堂評量，一般而言，教師將使用電子化教學和管理系統實施電子化測驗，並即時給予學生回饋。而學生的課堂表現和電子化測驗結果，也是判斷其表現等級的重要依據。因此，編訂總體評量規準表和／或根據具體學科要求

調整規準表後，教師按照課程及評量框架、培養目標及評量規準表擬訂評量試題。

為協助教師客觀評估課程實施的整體成效，同時進一步驗證rubrics的評量架構設計之有效性，A校科學教師於2018／2019學年第一學期針對二年級218位學生進行「科學素

養評量」，以瞭解學生是否達到課程標準所訂定之學力水平。是次評量涵蓋了學生當時所接觸的「電」、「化學反應」、「健康」等科學議題，並在科學學習範疇的內容知識及科學思維與能力兩個維度做重點評量。當中，評量所選取的大部分題目均為雙重維度題，即這些題目的背景脈絡涉及最少一門學科內容，並主要測量學生的科學思維和能力。學生須運用在學校及生活中掌握的科學知識，活用理解和探究、詮釋與應用、推理與論證等能力解決真實情景下的問題。題目類型分為單選題、配對題、排序題、開放式問答題。問答題需編訂評分指引，並據評分指引對學生答案進行編碼。

(四)進行測驗題目分析和通過項目反應理論(Item Response Theory, IRT)量尺化過程，建構科學素養某一維度和向度的量尺

根據對二年級學生科學素養測試題目之古典測驗理論(Classical Test Theory, CTT)分析結果，顯示平均試卷分數為13.78，標準差4.68，試卷的整體信度為.76，說明測試結果的準確度或是精確度都沒有嚴重測量誤差，即測試結果具內部一致性、可信賴的特點。研究者其後將相關指標和教師對學生的評量，使用Winsteps 3.92.1版以Rasch模型對評量題目參數進行校準，把學生能力和測試題目置於單維度的同一量尺上進行標定，根據學生能力和題目難度在量尺上的相對位置，進行學生之間、題目之間以至學生能力與題目之間的比較，也就是學習階段完結時與學生的成就連結起來，既讓學生知道自己的學習成效，瞭解自己的學習進度在哪一項的核心能力上學習得較好、哪一項較弱，同時也提供教師未來教學應改善的方向。

經題目分析，是次測試的題目區分度介

於0.06 ~ 0.51。檢視全部25道題目，只有題號為Q2之區分度不足0.10，須審視題目並視情況(例如答題通過率是否很高或很低所引起)予以微調修訂。另外，各測試題目的難度係數平均值為500，難度最大者題號為Q22(難度係數為813.93)，難度最小為Q4(難度係數為252.40)。題目卡方擬合指標(Infit Mean Square, MNSQ)均滿足0.8 ~ 1.4之擬合標準(Wolfe & Chiu, 1999)，適合使用Rasch模型將學生能力和測試題目難度並列在單維度的同一量尺上進行標定。

(五)利用能力表現關係圖劃定水平和表現描述，依照等級排列與評分，檢定學生能力水平發展，指明特定的表現特色，為後續學習制定目標

圖5顯示科學教師針對二年級學生進行的「科學素養評量」的題目難度與受試學生的科學素養能力表現關係圖(item-person map)。圖中左邊代表學生能力水平的分布情況，右邊代表題目難度水平的分布情況。從下往上，學生的能力水平增高，題目的難度水平增大。越靠近圖頂端，學生的能力水平越高，題目越難。相反，越靠近底部，學生的能力水平越低，題目越容易。若根據試題難度、難度之標準誤及Rasch模型特點等，可把題目及學生能力初步劃定四個水平類別，由低至高分別為水平1、水平2、水平3及水平4。學生科學素養水平級別的評定是根據學生在總量尺的表現而定，說明學生處於某一特定水平時所掌握的科學能力，以及所能夠做到的一系列與科學相關的活動，此數據分析方法是當前科學素養國際測試(例如：PISA, TIMSS)經常採用的。

教師分析學生在評量任務中的表現，可以協助釐清課程或教學中需要做改善的地方。如圖5所示，被評為水平1的學生有16

位，水平2學生有103位，水平3學生有86位，水平4學生有13位。從圖5可見，題目難度與學生能力的關係方面，整體測量項目對二年級學生而言十分適中，達標學生(水平2或以上)有93%，水平1學生只占該級樣本7%。教師可根據評量結果，針對學生能正確回答問題的比率，發現學生的科學能力問題所在，在後續的教學中協助學生釐清相關之科學知

識和概念，掌握相關的能力，以達成科學學習的教育目標。尤其是針對位處水平1的16名學生的科學能力發展，若根據rubrics總體評量規準表，其表現屬待改進。因此，未達水平2的學生，其基本學力指標也可被視為未達標，教師必須適時進行補救教學，幫助這些學生儘快趕上其同齡同學的科學水準，讓學生得以順利過渡到下一個學習階段。

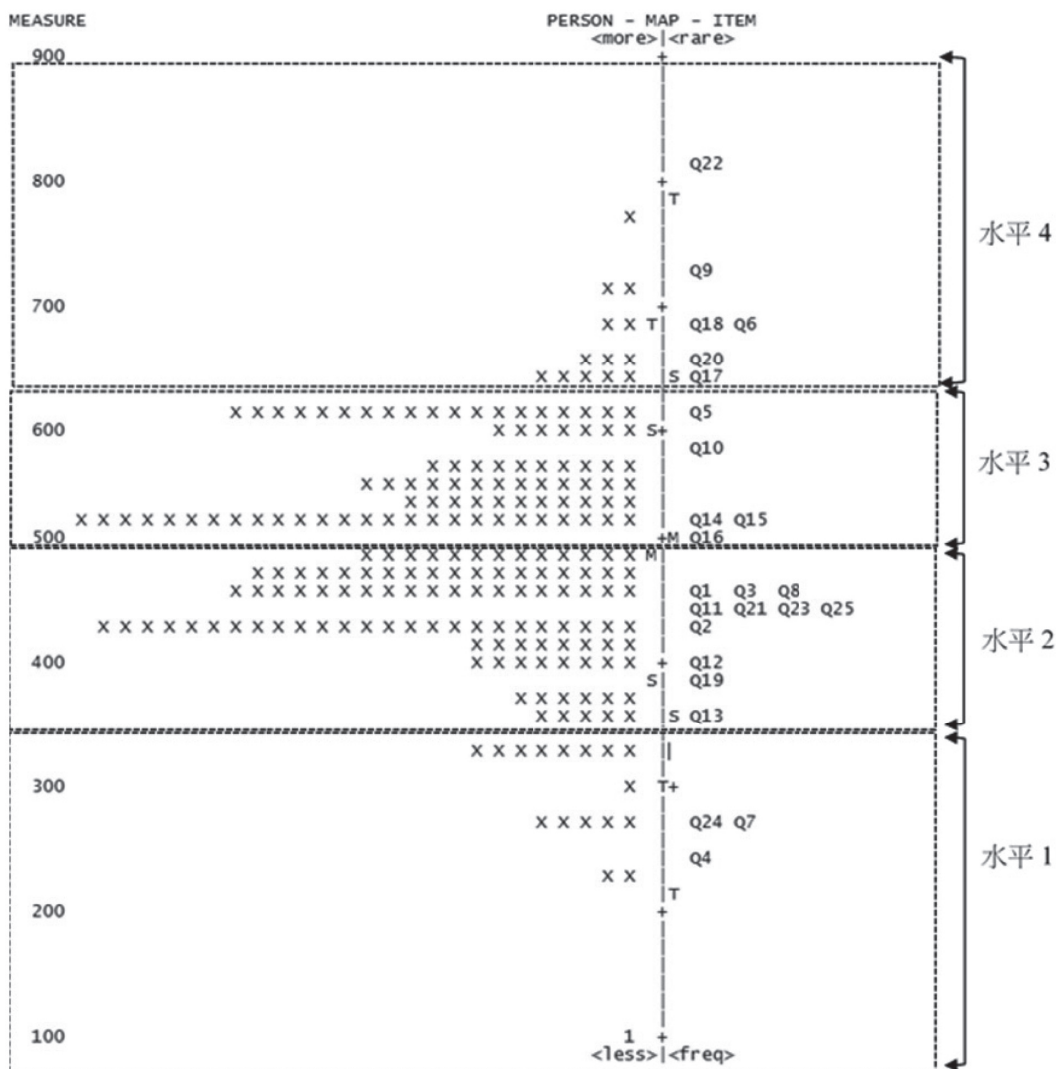


圖5：學生科學素養的能力水平和題目的難度水平分布

長久以來，教師們經常談到測驗／評量系統使他們站在一個教師對學生(上對下)評量的地位，而不是站在與學生一起(平等)評量的地位(張美玉，2012)。Bredekamp (1987)認為對學生的學習進行評量是為了選擇更適合的學習活動，然而，學校內傳統的紙筆測驗多為「量」的評量，依賴信度和效度都很好的測驗題目，它只能對學生的學習提供一個大概的瞭解，這種「量」的評量缺少「質」的訊息，無法提供學生是否學會了某種技能或某種概念的訊息。A校基於rubrics的設計取向針對澳門基本學力指標來驗證指標與學生能力的評量方法，同時兼備「量」和「質」的特點。過程中，教師透過設定基準，把學習目標結合rubrics而設計題組，一方面能從測量結果中檢視、記錄學生在學習過程中是否能達成預期目標，掌握整體的教學效果；另一方面對處於某一水平的學生，也能從相關的等級描述中瞭解到其學習進度。例如對教師而言，能從學生所位處的水平瞭解到其學習狀況、困難或落差，進而提供必要的協助及引導，以做好下一個教學的判斷，協助學生管理好學習。又例如對學生而言，一個被評定為水平2的學生也可根據量尺上的題目位置知悉自己大致掌握水平1的問題和知識點，只有一半機會能夠做到水平2的題目要求，較少機會做到水平3的題目要求，不太可能做到或達到水平4的題目要求，亦即意味著學生尚未具備最高及中上水平題目所測量的能力指標。再者，由於上述的測試題目是在rubrics總體評量規準表的基礎下設計，因而測試結果經rasch模型量尺化後所劃分的四個水平便可對應rubrics總體評量規準表上的三個表現等級，即Rasch量尺上的水平1可視為rubrics總體規準表上的待改進表現等級，Rasch量尺上的水平2至水平3可視為rubrics總體規準表

上的發展中表現等級，Rasch量尺上的水平4可視為rubrics總體規準表上的已具備表現等級。換言之，透過Rasch量尺化的過程可以進一步豐富rubrics評量的訊息回饋，因此兼備「量」和「質」的特點。此時，教師便可針對各個水平題目所指明的能力和／或能完成之科學活動與行為的要求，為學生的後續學習計畫提供更加明確的方向。

在教學過程中，要將「能力指標」轉化為「學習目標」時，教師有需要隨時參照相關的能力指標，清楚界定出學生應該學習到哪些知識與能力，進而在教學中逐步引導並協助學生習得。上述A校教師運用適切的評量工具來瞭解課程與教學所達至的效果，以及學生的學習成效，並以此作為調整教學及瞭解學生學習的依據，為教育工作者展示了只要把評量運用得宜，是可以幫助教師和家長瞭解學生到底學會了什麼，以及還有哪些需要學習。由此觀之，A校教師利用rubrics的評量特色設計建構學生科學素養評量系統，蘊含著「促進學習的評量」中重視回饋、具前瞻性、提供學習鷹架的關鍵特徵(莫慕貞，2016)。上述研究結果向教師展示了如何在學校現有的評量基礎上，引入基於rubrics的設計來瞭解學生能力的評量方法，有助教師從素養為本的角度來思考在日常的課堂中開展多元評量的具體方案和步驟，以客觀評價學校教育的狀況，本研究為豐實學校多元評量的內涵提供了可堪借鏡的研究實例。

伍、結論與建議

「Hill (2005)指出實踐校本評量通常面對兩種挑戰：技術性和改革管理；其中技術方面包括真實性評量任務的設計、成績調整的方法、測評協調的方式；而改革管理方面則

涉及公眾的接納度和教師在評量方面的專業能力與發展」(轉引自羅耀珍, 2008, 頁22)。再者, 要讓教育評量發揮積極的效果, 很大程度取決於如何進行評量和如何運用評量結果(侯傑泰, 2015)。本研究結果呈現案例學校在素養教學為本的教育環境下, 如何藉由學科主任、任課教師逐步有組織地規劃和發展全校科學教育課程及評量的改革, 協調運用「對學習成果的評量」及「促進學習的評量」方案以實踐多元評量。

一、結論

本研究對案例學校素養導向的科學教育多元評量實施過程的研究表明, 在進行課改和評量革新的過程中, 教師藉由對科學教育目標的反思和探討, 逐步明確了對相關能力指標的理解與詮釋, 進而改變評量觀念和評量方式, 建構以探究學生學習和思考歷程為重點的教學評量模式。更重要的是多元評量並非單向、自上而下的評量, 而是通過各種評量讓師生共同回顧和反思評量有關的資料及數據, 分享學習的成果; 學生知道並確認他們所達到的學習水平, 以決定如何做下一步學習, 並從反思中增強自我監控的能力。過程中, 多元評量活動始終緊扣教學目標和評量內涵, 並以促進學生學習成功為目的。

此外, 為把「能力指標」轉化為「學習目標」, 讓教師運用適切的評量工具來瞭解課程與教學所達至的效果, 以及學生的學習成效, 本研究探究案例學校基於rubrics的設計取向, 針對澳門基本學力指標來驗證指標與學生能力的評量方法, 發現有助教師從素養為本的角度來思考評量本身的目的, 加強教師對能力指標的解讀, 以及理解教學目

標、課程和評量交互檢核的角色, 進而審視評量任務的設計、成績調整的方法和測評協調的方式, 以發揮多元評量協助學習或教學, 提供指導和援助, 引領後續課程與教學的計畫與施行。

二、建議

Colorin (n.d.)認為多元評量的優勢是教師可以定期或隨時追蹤學生的進展, 透過多元評量, 教師能針對學生具體的學習狀況提早修正教學, 不須等到教學單元結束以後才思考對策(蘇淑英, 2014)。唯有必要指出的是評量的改革確實需要新的評分規則和學生學習表現成績報告的方式, 且要能正確的反映出所學習的課程的豐富性及多樣性。本研究探討教師利用rubrics來設計驗證指標與學生能力的評量方法, 未來也可以進一步擴展到不同的關鍵年級或學科領域; 並因應課程教育目標的持續更新, 教學過程的多元化, 朝向以素養為本的評量方式發展, 著重評量學生理解與高層次的分析、應用與批判能力; 同時, 透過實作評量、專題研習、主題報告、探究實驗等多元的方式和組合, 給予學生選擇的自由和空間, 讓學生能根據個人的興趣和能力, 實現個性化的學習。此外, 為方便相關持分者能隨時掌握學生的進展狀況或能力達標基準, 在條件許可或系統成熟時, 還可利用大數據(big data)龐大的數位化資料庫, 記錄學生在電子學習平臺上的學習歷程。這樣的學習歷程資料庫在有系統的管理及分析下, 能為學習者及指導者(教師或家長)提供十分有價值的資訊, 並為後續開發智慧型的適性教育評量系統奠定基礎。

參考文獻

1. 李川、劉克文(2016)。中小學科學教育改革的熱點與趨勢。中小學管理，5，4-7。
[Li, C., & Liu, K.-W. (2016). Zhongxiaoxue kexue jiaoyu gaige de redian yu qushi. *School Administration*, 5, 4-7.]
2. 李春密、趙芸赫(2017)。STEM相關學科課程整合模式國際比較研究。比較教育研究，39(5)，11-18。
[Li, C., & Zhao, Y. (2017). An international comparative study of STEM—Related curriculum integration models. *International and Comparative Education*, 39(5), 11-18.]
3. 辛濤(2017年2月4日)。建立科學評價體系 營造良好育人生態。查詢日期：2020年5月1日，檢自http://www.moe.gov.cn/jyb_xwfb/moe_2082/zl_2017n/2017_zl06/201702/t20170205_295750.html。
[Xin, T. (2017, February 4). *Jianli kexue pingjia tixi Yingzao lianghao yuren shengtai*. Retrieved May 1, 2020, from http://www.moe.gov.cn/jyb_xwfb/moe_2082/zl_2017n/2017_zl06/201702/t20170205_295750.html]
4. 辛濤、張彩(2018)。中小學教育質量綜合評價改革的現況與前瞻。中國教育學刊，8，37-41。
[Xin, T., & Zhang, C. (2018). Status quo and prospects of the reform in comprehensive assessment of the quality for the primary and secondary education. *Journal of the Chinese Society of Education*, 8, 37-41.]
5. 吳璧純(2013)。從三種評量類型看多元評量的意義。新北市教育，8，20-24。
[Wu, P.-c. (2013). Cong sanzong pingliang leixing kan duoyuan pingliang de yiyi. *New Taipei City Education*, 8, 20-24.]
6. 吳璧純(2017)。素養導向教學之學習評量。臺灣教育評論月刊，6(3)，30-34。
[Wu, P.-c. (2017). Suyang daoxiang jiaoxue zhi xuexi pingliang. *Taiwan Educational Review Monthly*, 6(3), 30-34.]
7. 林怡呈、吳毓瑩(2008)。多元評量的活化、迷思、與神話——教學歷程的個案研究。課程與教學，11(1)，147-172。doi:10.6384/CIQ.200801.0147
[Lin, Y.-C., & Wu, Y.-Y. (2008). Activation, misconception, and myth of classroom assessment—A case study. *Curriculum & Instruction Quarterly*, 11(1), 147-172. doi:10.6384/CIQ.200801.0147]
8. 林曉雯、陳佩君(2005)。科學教學評量的反思與實踐：教師行動研究。科學教育學刊，13(1)，1-27。doi:10.6173/CJSE.2005.1301.01

- [Lin, S.-W., & Chen, P.-C. (2005). Reflective practices on science assessment: A teacher's action research. *Chinese Journal of Science Education*, 13(1), 1-27. doi:10.6173/CJSE.2005.1301.01]
9. 房湘雲(2013)。促進中文科讀寫教學的評估實踐。《評估與學習》，2，94-99。
[Fong, S. W. (2013). Cujin zhongwenke duxie jiaoxue de pinggu shijian. *Assessment and Learning*, 2, 94-99.]
 10. 侯傑泰(2015年11月18日)。侯傑泰：監測或不監測？——上海臺灣經驗。查詢日期：2020年5月1日，檢自<http://www.fed.cuhk.edu.hk/~kthau/web/articles/Hau2015TSAShanghaiTaiwanMingPaoDailyNews20151118.pdf>。
[Hau, K.-T. (2015, November 18). *Kit-Tai Hau: Jiance huo bujiance?—Shanghai Taiwan jingyan*. Retrieved May 1, 2020, from <http://www.fed.cuhk.edu.hk/~kthau/web/articles/Hau2015TSAShanghaiTaiwanMingPaoDailyNews20151118.pdf>]
 11. 教育局(2017)。評估素養及學校評估政策。查詢日期：2020年5月1日，檢自https://www.edb.gov.hk/attachment/tc/curriculum-development/renewal/Guides/SECG%20booklet%204_ch_20180831.pdf。
[Education Bureau. (2017). *Assessment literacy and school assessment policy*. Retrieved May 1, 2020, from https://www.edb.gov.hk/attachment/tc/curriculum-development/renewal/Guides/SECG%20booklet%204_ch_20180831.pdf]
 12. 教育暨青年局(2017年6月)。《本地學制正規教育學生評核制度》諮詢總結報告。查詢日期：2020年6月1日，檢自http://202.175.82.54/dsej/edulaw/201706/concluding_report_c.pdf。
[Education and Youth Affairs Bureau. (2017, June). *Bendi xuezhi zhenggui jiaoyu xuesheng pinghe zhidu zixun zongjie baogao*. Retrieved June 1, 2020, from http://202.175.82.54/dsej/edulaw/201706/concluding_report_c.pdf]
 13. 張美玉(2012)。國民中小學多元評量案例。查詢日期：2020年5月1日，檢自<https://www.tcavs.tc.edu.tw/upload/103090495921.pdf>。
[Chang, M.-Y. (2012). *Guomin zhongxiaoxue duoyuan pingliang anli*. Retrieved May 1, 2020, from <https://www.tcavs.tc.edu.tw/upload/103090495921.pdf>]
 14. 張淑賢(2014)。促進學習的評估原則及指標。《評估與學習》，3，15-21。
[Berry, R. (2014). Cujin xuexi de pinggu yuanze ji zhibiao. *Assessment and Learning*, 3, 15-21.]
 15. 張華(2016)。論核心素養的內涵。《全球教育展望》，45(4)，10-24。
[Zhang, H. (2016). On the connotations of a key competence. *Global Education*, 45(4), 10-24.]
 16. 郭重吉(2012)。中小學科學教育的推展。《科學發展》，474，66-70。
[Guo, C.-J. (2012). Zhongxiaoxue kexue jiaoyu de tui zhan. *Science Development*, 474, 66-70.]
 17. 郭重吉、張惠博(2005)。從政策層面評述國際間科學教育的改革。《科學教育月刊》，

284, 23-35。

[Guo, C.-J., & Chang, H.-P. (2005). A review of worldwide science education reforms from policy perspective. *Science Education Monthly*, 284, 23-35.]

18. 國家教育研究院(2014年11月28日)。十二年國民基本教育課程綱要：總綱。查詢日期：2020年5月1日，檢自<https://www.naer.edu.tw/upload/1/16/doc/288/%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%B8%BD%E7%B6%B1.pdf>。

[National Academy for Educational Research. (2014, November 28). *Curriculum guidelines of 12-year basic education: General guidelines*. Retrieved May 1, 2020, from <https://www.naer.edu.tw/upload/1/16/doc/288/%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%B8%BD%E7%B6%B1.pdf>]

19. 陳健生、甘國臻、霍秉坤(2010)。「促進學習的評估」改革：學校的實施與挑戰。《課程研究》，4(4)，69-88。

[Chan, J. K.-S., Kennedy, K. J., & Fok, P.-K. (2010). The reform of “assessment for learning”: School implementation and challenges. *Journal of Curriculum Studies*, 4(4), 69-88.]

20. 陳琦媛(2017)。運用rubrics評量核心素養。《臺灣教育評論月刊》，6(3)，87-90。

[Chen, C.-Y. (2017). Yunyong rubrics pingliang hexin suyang. *Taiwan Educational Review Monthly*, 6(3), 87-90.]

21. 莫慕貞(2016年，6月)。促進學習的評估：理念與實踐。發表於校董專題講座2015/16學年——如何持續提升學與教效能？香港：教育局。

[Mok, M. C. (2016, June). *Cujin xuexi de pinggu: Linian yu shijian*. Paper presented at Xiaodong zhuanli jiangzuo 2015/16 xuenian—Ruhe chixu tisheng xue yu jiao xiaoneng? Hong Kong: Education Bureau.]

22. 黃淑玲(2013)。從知識到可觀察的能力：評估學習成效的策略與建議。《評鑑雙月刊》，44，16-23。

[Huang, S.-L. (2013). Cong zhishi dao keguancha de nengli: Pinggu xuexi chengxiao de celue yu jianyi. *Evaluation Bimonthly*, 44, 16-23.]

23. 曾淑惠(2015)。翻轉教學的學習評量。《臺灣教育評論月刊》，4(4)，8-11。

[Tseng, S.-H. (2015). Fanzhuan jiaoxue de xuexi pingliang. *Taiwan Educational Review Monthly*, 4(4), 8-11.]

24. 褚宏啟(2016)。核心素養的國際視野與中國立場——21世紀中國的國民素質提升與教育目標轉型。《教育研究》，11，8-18。

[Chu, H. (2016). The international perspective and China's stance of key competencies—The enhancement of national qualities and the transformation of educational goals in 21st century

- in China. *Educational Research*, 11, 8-18.]
25. 靳知勤(2007)。科學教育應如何提升學生的科學素養——臺灣學術精英的看法。《科學教育學刊》，15(6)，627-646。doi:10.6173/CJSE.2007.1506.02
- [Chin, C.-C. (2007). A reflection on the science education of Taiwan—The voice from the elites in Taiwan. *Chinese Journal of Science Education*, 15(6), 627-646. doi:10.6173/CJSE.2007.1506.02]
26. 歐用生(2003)。《課程典範再建構》。高雄市：麗文文化。
- [Ou, Y.-S. (2003). *Kecheng dianfan zai jiangou*. Kaohsiung, Taiwan: Liwen Cultural.]
27. 劉堅(2013)。新世紀課程變革：親歷者的視角。《北京大學教育評論》，11(4)，2-20。
- [Liu, J. (2013). Curriculum reform in the new century: Review and reflection. *Peking University Education Review*, 11(4), 2-20.]
28. 劉曼君(2014)。學生學習成果之評量及評分量表rubrics之使用。《評鑑雙月刊》，48，54-56。
- [Liu, M. (2014). Xuesheng xuexi chengguo zhi pingliang ji pingfen liangbiao rubrics zhi shiyong. *Evaluation Bimonthly*, 48, 54-56.]
29. 蔡清田、陳伯璋、陳延興(2014)。十二年國民基本教育課程發展指引(NAER-101-18-A-1-02-00-2-15.16.17.18)。臺北市：國家教育研究院。
- [Tsai, C.-T., Chen, P.-C., & Chen, Y. H. (2014). *Curriculum development guidelines for 12-year basic education* (NAER-101-18-A-1-02-00-2-15.16.17.18). Taipei, Taiwan: National Academy for Educational Research.]
30. 課程發展議會(2002)。《學校評估政策——評估實施方式的轉變》。查詢日期：2020年5月1日，檢自https://cd1.edb.hkedcity.net/cd/TC/Content_2909/html/chapter05.html。
- [Curriculum Development Council. (2002). *School policy on assessment—Changing assessment practices*. Retrieved May 1, 2020, from https://cd1.edb.hkedcity.net/cd/TC/Content_2909/html/chapter05.html]
31. 課程發展議會(2017)。《科學教育：學習領域課程指引(小一至中六)》。查詢日期：2020年5月1日，檢自https://www.edb.gov.hk/attachment/tc/curriculum-development/kla/science-edu/SEKLACG_CHI_2017.pdf。
- [Curriculum Development Council. (2017). *Science education: Key learning area curriculum guide (primary 1–secondary 6)*. Retrieved May 1, 2020, from https://www.edb.gov.hk/attachment/tc/curriculum-development/kla/science-edu/SEKLACG_CHI_2017.pdf]
32. 澳門中華教育會(2016)。《本地學制正規教育學生評核制度》諮詢文本集思會意見書。查詢日期：2020年5月1日，檢自<https://www.edum.org.mo/feature/new--5>。
- [The Chinese Educators Association of Macau. (2016). *Bendi xuezhi zhenggui jiaoyu xuesheng pinghe zhidu zixun wenben jishui yijianshu*. Retrieved May 1, 2020, from <https://www.edum.org.mo/feature/new--5>]

org.mo/feature/new--5]

33. 澳門特別行政區政府(2006年12月26日)。非高等教育制度綱要法。查詢日期：2020年6月18日，檢自https://bo.io.gov.mo/bo/i/2006/52/lei09_cn.asp。
[Macao SAR Government. (2006, December 26). *Feigaodeng jiaoyu zhidu gangyaofa*. Retrieved June 18, 2020, from https://bo.io.gov.mo/bo/i/2006/52/lei09_cn.asp]
34. 澳門特別行政區政府(2014年6月30日)。本地學制正規教育課程框架。查詢日期：2020年6月18日，檢自https://bo.io.gov.mo/bo/i/2014/26/regadm15_cn.asp。
[Macao SAR Government. (2014, June 30). *Bendi xuezhi zhenggui jiaoyu kecheng kuangjia*. Retrieved June 18, 2020, from https://bo.io.gov.mo/bo/i/2014/26/regadm15_cn.asp]
35. 澳門特別行政區政府(2015年7月20日)。本地學制正規教育基本學力要求。查詢日期：2020年6月18日，檢自https://bo.io.gov.mo/bo/i/2015/29/regadm10_cn.asp。
[Macao SAR Government. (2015, July 20). *Bendi xuezhi zhenggui jiaoyu jiben xueli yaoqiu*. Retrieved June 18, 2020, from https://bo.io.gov.mo/bo/i/2015/29/regadm10_cn.asp]
36. 澳門特別行政區政府(2017年6月26日)。第55/2017號社會文化司司長批示。查詢日期：2020年6月18日，檢自https://bo.io.gov.mo/bo/i/2017/26/despasc_cn.asp#56。
[Macao SAR Government. (2017, June 26). *Di 55/2017 hao Office of the Secretary for Social Affairs and Culture sizhang pishi*. Retrieved June 18, 2020, from https://bo.io.gov.mo/bo/i/2017/26/despasc_cn.asp#56]
37. 羅耀珍(2008)。促進學習的評估。香港：香港大學出版社。
[Lo, Y. C. (2008). *Assessment for learning*. Hong Kong: Hong Kong University Press.]
38. 蘇淑英(2014)。落實多元評量於課堂教學中。《語言之道》，3，78-81。
[Su, R. (2014). Luoshi duoyuan pingliang yu ketang jiaoxue zhong. *The Way of Language*, 3, 78-81.]
39. Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham, UK: Open University Press.
40. Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21. doi:10.1177/003172170408600105
41. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74. doi:10.1080/0969595980050102
42. Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp. 9-26). London, UK: Sage.
43. Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. doi:10.1177/003172171009200119
44. Bredekamp, S. (Ed.). (1987). *Developmentally appropriate practice in early childhood*

- programs serving children from birth through age 8*. Washington, DC: National Association for the Education of Young Children.
45. Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
46. Colorin, C. (n.d.). *Using informal assessments for English language learners*. Retrieved May 10, 2020, from <http://www.colorincolorado.org/educators/assessment/informal>
47. Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
48. Earl, L. M. (2013). Assessment for learning; assessment as learning: Changing practices means changing beliefs. *Assessment and Learning*, 2, 1-5.
49. Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
50. Little, D. (2006). *Grading with rubrics: Developing a fair and efficient assessment tool*. Retrieved May 10, 2020, from https://ira.virginia.edu/sites/ias.virginia.edu/files/trc_gradingrubric.pdf
51. Mullis, I. V. S., & Martin, M. O. (2013). *TIMSS 2015 assessment frameworks*. Retrieved May 10, 2020, from http://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf
52. National Research Council. (1996). *National science education standards*. Washington, DC: National Academic Press.
53. National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
54. Organisation for Economic Co-operation and Development. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. Paris, France: Author.
55. Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: Routledge.
56. Rychen, D. S., & Salganik, L. H. (2001). *Defining and selecting key competencies*. Göttingen, Germany: Hogrefe & Huber.
57. Stevens, D., & Levi, A. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Sterling, VA: Stylus.
58. Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—Using it well*. Portland, OR: ETS Assessment Training Institute.
59. Taggart, G. L. (2005). *Rubrics: A handbook for construction and use*. Lanham, MD: Rowman & Littlefield.

60. Thier, M., & Daviss, B. (2002). *The new science literacy: Using language skills to help students learn science*. Portsmouth, NH: Heinemann.
61. Wolfe, E. W., & Chiu, C. W. (1999). Measuring change across multiple occasions using the Rasch Rating Scale Model. *Journal of Outcome Measurement*, 3(4), 360-381.
62. Yore, L. D., Pimm, D., & Tuan, H. L. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education*, 5, 559-589. doi:10.1007/s10763-007-9089-4

Through a Case Study of Science Education of a Macao School Examining Processes of Implementation of Multiple Assessment

Soi-Kei Mak^{1,*}, Kwok-Cheung Cheung², Man-Kai Ieong¹ and Pou-Seong Sit²

¹Educational Testing and Assessment Research Centre, University of Macau

²Faculty of Education, University of Macau

Abstract

Based on a case study conducted in a secondary school in Macao that practices literacy-based school education and promotes school-based science curriculum development, this study sought to investigate the essential characteristics of multiple assessment in literacy-oriented science education. Based on the current foundation of assessment, the methods and procedures of how rubrics assessment scales can be designed and utilized to verify learning objectives and students' basic competencies in science were the focus of examination in this study. These methods and procedures are able to evaluate objectively whether a student's learning progression can achieve the academic standards as laid down in the curriculum standards, resulting in enrichment of multiple assessment in school and providing evidence for informed policy making regarding school improvement and instructional effectiveness. This case study renders readers opportunities to understand the processes pertaining to the implementation of multiple assessment. Additionally, analysis of these processes allows investigators to understand that multiple assessment possess dual characteristics of "quantity" and "quality." Suggestions forwarded in this study are that investigators using multiple assessment should pay attention to the reciprocal feedback between students and their teachers, and be forward-looking and provide scaffolding support for students to learn better. The ultimate aim of multiple assessment is to guide students to progress to the next stage of learning or development, and migrate closer to success in learning.

Key words: Rubrics Assessment Scale, Multiple Assessment, Science Education, Core Literacy, Learning Progression

* Corresponding author: Soi-Kei Mak, skmak@um.edu.mo