

從三環評量理論探討融合教育 之學習評量

林沛穎

加拿大莎佳奇萬大學

林昱成

美國德州大學心理系

摘要

學習評量是學習過程中不可或缺的一環，隨著時代的變遷，評量的概念也應隨之改變。本文探討三環評量的基本概念與迷思（促進學習的評量、評量即學習、學習成果的評量），並更進一步提出一項新的整合型評量模式。我們認為，除原先的三環評量方式外，更應納入教學和考試的調整策略，以幫助接受融合教育的身心障礙學生，適性地學習與發展。

Assessment of, for, and as Learning in Inclusive Education

Pei-Ying Lin

University of Saskatchewan, Canada

Yu-Cheng Lin

University of Texas, United States

Abstract

Educational assessment is indispensable for student learning and instruction across grade levels and subject domains. As it is critical to address all students' learning needs in inclusive classrooms, the definitions and concepts of conventional assessment should be revisited and modified. This paper begins by laying out three major assessment concepts (assessment for learning, assessment as learning, and assessment of learning) and discussing myths of classroom and large-scale assessments. We propose an alternative assessment model which also includes the use of assessment and instructional accommodations for students with special needs placed in inclusive classroom settings. This alternative assessment model is aimed to promote equity in assessment for students with special needs.

教學評量是教學中相當重要的一環，無論學生的年級或學科，學習評量也是幫助學生學習不可或缺的一部份。更重要的是，教師對於教學或學習評量的信念或觀念 (beliefs or conceptions) 扮演舉足輕重的角色，希望透過本文的論述，提供給職前師資培育和在職教師參考。融合教育已為國際特殊教育的重要發展趨勢，有鑒於此，本文旨在探討融合教育情境下的

學習評量，目的為探討如何將基本的學習評量的概念應用於教學及融合教育的情境中。此外，作者也將舉例在北美相關的教學經驗，一併進行討論，希望能拋磚引玉，以供台灣特殊教育的參考。

壹、緒論

教育評量的概念與理論不斷的更新，從過去所謂的填鴨式考試，為因應趨



勢，逐漸從傳統的教師導向 (teacher-transmitted) 轉變為學生導向 (student-centered) 的評量方式。然而，在這樣的轉變下，師資培育所提供的課程及教師訓練對現任教師與職前教師的影響深遠，因為師資培育的過程逐漸形塑教師的信念，而這些信念被視為教師行為和做決定的最佳指標(Haney, Czerniak, & Lumpe, 1996; Pajares, 1992)。學者 (Brown, 2004; Volante & Fazio, 2007) 認為了解教師信念有助於未來相關政策的訂定、師資培育和專業進修。我們必須深入探討一些最根本的問題，例如：為何評量？如何評量？再者，我們也進一步探究身心障礙學生在融合教育情境下，所使用的評量是否公平客觀，並能幫助他們適性發展，這些重要的問題將在本文逐一探討。

貳、評量目的與三環評量理論

在教師決定如何評量學生前，必須釐清評量的目的。目前研究指出學習評量包含三個主要的概念：促進學習的評量 (assessment for learning; AFL)、評量即學習 (assessment as learning; AAL)、學習成果的評量 (assessment of learning; AOL) (甄曉蘭, 2008; 吳碧純 2009, 2013; Black & Wiliam, 1998a, 1998b)。所謂「促進學習的評量」意指教師持續使用多元評量策略和回饋來了解學生各個方面的學習成長、進步與學習需求，並可作為教師教學的依據。再者，評量的結果可幫助教師、學生本身、同儕來思考決定教學及學習的下一步(Black & Wiliam, 2009; Wiliam, 2011)。此外，「評量即學習」主要是鼓勵學生培養和應用自省及反思的能力，隨時思考自己的教學狀況，並作調整。也因此，學生在自己的學習過程扮演主動的角色，而非過去傳統被動的方式(Earl, 2003; Earl & Katz, 2006; Marshall & Drummond, 2006)。「學習成果的評量」則是用於評量學生的某項知識、能力和技能，以進一步了解學生是否達到課程期望或標準。這類

型的評量是為了提供不同時間、不同的資訊給教師、學校、甚至是教育主管單位、政策決策者 (Stiggins, 2006)，並可作為教育安置 (placement)、升學或畢業的依據。標準化測驗或是美國和加拿大各州的教育測驗常被歸類為此，例如加拿大安大略省 (Ontario) 每年為三、六年級的數學、閱讀及寫作，九年級數學，及十年級舉辦的讀寫能力測驗。其中，十年級學生必須通過此測驗，以達到畢業的要求，若未通過，可選擇參加第二次測驗或選修讀寫課程。

此外，「促進學習的評量」常被聯結到形成性評量，而「學習成果的評量」則是常被視為總結性評量，因此有學者認為這兩種形態的評量，具有截然不同的功能，以達到不同的目的(Birenbaum et al., 2006; Black et al., 1998a; Harlen & James, 1997)。而且，相當多的學者支持或批評形成性 (促進學習的評量)、總結性評量 (學習成果的評量) (例如 Birenbaum et al., 2006; Brindley, 2001; Teasdale & Leung, 2000)。這些爭論主要是一般印象均認為總結性評量強調考試，用於報告學生的學業成績 (如月考、期末考) 和學校績效，可能造成所謂的考試領導教學，為人所詬病，且引起相當大的教學和學習壓力。相對地，形成性的評量著重學習的過程和成長，持續提供回饋及調整學習的方式或進度，此過程較為複雜，具有彈性。Black 和 Wiliam (1998a, 1998b) 曾研究形成性評量的效果值 (effect sizes) 比其他許多教育介入或方法來得顯著，雖然，它的效果值在不同的研究中也有所不同(例如 Kluger & DeNisi, 1996; Wiliam, Lee, Harrison, & Black, 2004)。這些爭論和要求教學績效的壓力下，引起加拿大相當多教師的反彈，反對州政府的年度學測，雖然這些學測已在北美行之多年。然而，我們也必須反思並比較兩者間的優、缺點，並考慮整個教育環境的條件和限制。我們認為無論何種

評量模式，不應只拘泥在字面文意的解釋（例如 *assessment for learning* 或 *assessment of learning*），而是要能妥善應用各種，以利精確地了解及診斷學生的學習需求，可以作為有效改善教學的依據，規劃明確的未來學習目標。最近也有一些學者陸續指出一些盲點（Black, McCormick, James, & Pedder, 2006; Marshall & Drummond, 2006; Torrance, 2007），他們認為應該要能掌握評量的精神（spirit）而非字意（letter）。所謂的「精神」，鼓勵學生主動學習，具體教學策略可包括引導學生共同建立評量標準，經由小組討論和批判思考去修改學生自己的想法。而「字意」則較侷限在一般的通則和教法，例如跟學生解釋老師的評量標準，或老師提供回饋給學生。除此之外，我們也認為應打破對總結性或「學習成果的評量」的迷思，認為測驗的結果無法促進學生學習的觀念，譬如教師也可分析學生在年度學測或總結性評量的表現，進一步分析題目，找出經常出現錯誤的概念，與教學團隊進行研究和討論，以做為下學年度教學的參考。這種教學診斷的模式，較能提供有意義的資訊給學生與教師，也較能建立年度學測與學習之間的正面意義及關係。同樣，此診斷模式也可應用於一般在教室使用的總結性評量。綜合以上的討論，我們認為不論是形成性或總結性評量，亦或是促進學習的評量、評量即學習、學習成果的評量，都不應是截然不同、分離的理論，而必須是整合性的概念（Bennett, 2011; Wiliam, 2011）。因此，研究或教學重點應放在如何整合，靈活運用三個概念，不僅只是爭論、比較三者的優劣。

參、考試或教學調整在特殊教育的應用

以目前大多數相關的文獻，我們很難單從這三環理論的架構去了解身心障礙學生在融合教育情境下是否也能像一般

同儕受益，這是我們所應該關心的議題。我們認為考試或教學的調整（*instructional and test accommodations or adaptations*）也必須納入前述之三環評量概念中。目前，這些調整引起越來越多的關注，其所包含的策略相當廣泛，也因應學科的不同，而有所改變，例如延長時間、座位調整（*setting*）、個別測驗、唸讀（*read aloud*）、繕寫（*scribe*）、電腦應試、輔助科技、手語、點字、放大字體等。教學或考試調整近年來逐漸受融合教育者的重視，因為這些調整能有效地促進融合教育的施行。Pullin (2008) 指出過去在尚未提供這些調整的年代，相較於同儕，特殊教育學生的學習權益常被忽略。他們常必須和同儕以同樣的方式接受相同的評量，然而，評量的結果卻無法精確地量測他們的學習能力和進步，評量的公平和適切性易受質疑。或者，因為各種原因，他們不須接受學測或一般評量，這類的做法，直接影響到這些接受融合教育學生的受教權。北美國家已明文規定保障身心障礙學生的受教權及參與學測的機會（如美國 *the Individual with Disabilities Education Act of 2004 [IDEA]*, *No Child Left Behind Act of 2001 [NCLB]*），必須將他們的評量結果納入學校績效考核（*school accountability*）（例如 Niebling & Elliott, 2005; Stretch & Osborne, 2005; Thurlow, Lazarus, Thompson, & Morse, 2005）。也因此，這些調整受到相當大的重視，因為適當的調整可幫助接受融合教育的身心障礙學生，在較少的限制下，展現他們所學的知識和技能（例如 Bolt & Thurlow, 2007; Fuchs et al., 2000a, 2000b）。這是相當重要的一點，因為老師或學生才能有效地了解學習的狀況，並做適當的調整，訂定下一步學習目標。

值得一提的是，任何調整都必須維持測驗的效度，因為調整如果是不適當的，將會改變原本要評量的目標或構念，因此量測便不正確，連帶也會影響教師解釋評



量的結果，因而會破壞構念效度 (American Educational Research Association et al., 1999; National Research Council, 2004)。相當重要的是，這些調整也必須與學生的學習需求相輔相成，提供多餘的調整反而會造成混淆或是讓學生分心，對評量結果有負面影響 (Helwig & Tindal, 2003; Ketterlin-Geller, Alonzo, Braun-Monegan, & Tindal (2007)。為因應此情形，許多學生建議平時學生所使用的調整，應該和記錄在個別化教育計劃的建議相符 (Cook et al., 2010; Johnson & Monroe, 2004; Pitoniak & Royer, 2001)。此外，為參加學測評量所做的調整，也必須與教室評量所用的調整一致，畢竟學生必須已經學會如何使用且能適應各種調整方式 (Cox, Herner, Demczyk, & Nieberding, 2006; National Research Council, 2004)。舉例說明，如果一位教師想要深入瞭解學生的閱讀理解程度，但是考量識字困難學生需要更多的時間才能完成測驗，可考慮延長時間，因為可以反應學生的需求，且不改變原本要評量的目標，但是，如果教師想評量的是閱讀流暢度，延長測量測驗並不適當，因為此調整混淆了原本所要測量的目的。

目前已經有相當多學者贊成師資培育機構必須重視並提供在職及職前教師對評量和教學調整的認知和訓練 (Fuchs & Fuchs, 2001; Hodgson, Lazarus, & Thurlow, 2011)。知名美國學者 Fuchs et al. (2000a) 研究發現教師在評量學習障礙學生的閱讀測驗能力時，多提供過度的評量調整。更重要的是，他們發現教師難預測學生是否能受益於評量的調整，而且教師所建議的調整方式，實際上對學生並無幫助，甚至不使用調整的學生卻比使用教師所建議調整的學生表現佳。非但閱讀測驗如此，他們在另一項數學研究也發現類似的結果 (Fuchs et al., 2000b)。Tindal, Lee, 和 Ketterlin-Geller (2008) 也指出教師在判斷調整所可能帶來的正面影響缺乏一致

性，大約只有百分之五十的機會是一致的，因此教師對這方面的判斷並不可靠。另一項研究 Ketterlin-Geller, Alonzo, Braun-Monegan 和 Tindal (2007) 也支持這樣的發現，這些學者進一步指出，教師所建議的調整方式和學生的個別化教育計畫內的調整模式多有所不同。根據研究發現，學者 (例如 Fuchs et al., 2000a, 2000b) 建議教師必須根據評量的結果 (data-based measurements)，來客觀選擇適當的調整方式，而非依賴主觀的判斷。此外，雖然少數的台灣學生為非母語學習者，但是，他們的學習需求，不應被忽略，評量調整對這些學生的影響也需要被檢視，越來越多針對非母語學習者的研究，也指出評量的調整是必要的 (例如 Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Hofstetter, & Lord, 2004)。

本文討論三環評量概念間的關係，提出我們的看法，促進學習的評量、評量即學習、學習成果的評量等三者，必須整合，而非爭論孰優孰劣。再者，我們也建議評量調整也須納入三環概念和實踐於教學中，因為這攸關評量的公平性以及評量是否能提供教師或學生有價值的訊息，並且幫助融合教育情境下所有學生包含身心障礙學生有效地學習。本文的論述，提供給職前師資培育和在職教師的參考，期望建立一個多元且適性的融合教育環境。

參考文獻

- 甄曉蘭 (2008)。促進學習的課堂評量概念分析與實施策略。**中等教育**，59，92-109。
- 吳璧純 (2009)。注重並增進學童心理歷程能力的課程多元評量。**教育研究與發展期刊**，5，112-157。
- 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。
- Abedi, J., Lord, C., Hofstetter, C., & Baker,

- E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*, 16-26.
- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*, 1-28.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., et al. (2006). A learning integrated assessment system. *Educational Research Review, 1*, 61-67.
- Black, P., McCormick, R., James, M. & Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education, 21*(2), 119-132.
- Black, P., & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*, 139-148.
- Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 7-73.
- Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.
- Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention, 33*, 15-28.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language testing, 18*, 393-407.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*, 301-318.
- Cook, L., Eignor, D., Sawaki, Y., Steinberg, J., & Cline, F. (2010). Using factor analysis to investigate accommodations used by students with disabilities on an English- language arts assessment. *Applied Measurement in Education. Special Issue: Testing Students with Disabilities, 23*, 187-208.
- Cox, M. L., Herner, J. G., Demczyk, M. J., & Nieberding, J. J. (2006). Provision of testing accommodations for students with disabilities on statewide assessments: Statistical links with participation and discipline rates. *Remedial and Special Education, 27*, 346-354.
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind*. Winnipeg, MB: Western Northern Canadian Protocol.
- Fuchs, L. S., & Fuchs, D. (2001). Helping



- teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, 16(3), 174-181.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C., Binkley, E., & Crouch, R. (2000a). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67(1), 67-81.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000b). Supplemental teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29, 65-85.
- Haney, J., Czerniak, C. & Lumpe, A. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. *Journal of Research in Science Teaching*, 33(9), 971- 993.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379.
- Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69(2), 211-225.
- Hodgson, J. R., Lazarus, S. S., & Thurlow, M. L. (2011). *Professional development to improve accommodations decisions-A review of the literature* (Synthesis Report 84). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Individuals with Disabilities Education Act of 2004, Pub. L. No. 108-446, 20 U.S.C.
- Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention*, 29, 35-45.
- Ketterlin-Geller, L. R., Alonzo, J., Braun-Monegan, J., & Tindal, G. (2007). Recommendations for accommodations: Implications of (In)consistency. *Remedial and Special Education*, 28, 194-206.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Marshall, B., & Jane Drummond, M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, 21(2), 133-149.
- National Research Council. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Judith A. Koenig and Lyle F. Bachman, Editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Niebling, B. C., & Elliott, S. N. (2005). Testing accommodations and inclusive assessment practices. *Assessment for Effective Intervention*, 31, 1-6.

- No Child Left Behind Act of 2001*, Pub. L. No. 107-110, 115 Stat. 1425.
- Pajares, M. F. (1992). Teachers' beliefs and education research: Cleaning up a messy construct. *Review of Education Research*, 62, 307-332.
- Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research*, 71, 53-104.
- Pullin, D. C. (2008). Individualizing assessment and opportunity to learn: Lessons from the education of students with disabilities. In P. Moss, D. Pullin, J. P. Gee, Haertel, E. H., & Young, L. J. (Eds.), *Assessment, equity, and opportunity to learn* (pp. 109-135). New York: Cambridge University Press.
- Stretch, L. S., & Osborne, J. W. (2005). Extended time test accommodation: Directions for future research and practice. *Practical Assessment Research & Evaluation*, 10 (7), 1-8.
- Stiggins, R. J. (2006). Assessment for learning: A key to student motivation and learning. *Phi Delta Kappa Edge*, 2(2), 1-19.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.
- Tindal, G., Lee, D., & Ketterlin-Geller, L. (2008). *The reliability of teacher decision-making in recommending accommodation for large-scale tests* (Technical report 08-01). Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post - secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281-294.
- Thurlow, M. L., Lazarus, S. S., Thompson, S. J., & Morse, A. B. (2005). State policies on assessment participation and accommodations for students with disabilities. *Journal of Special Education*, 38, 232-240.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30, 749-770.
- William, D. (2011). What is assessment for learning?. *Studies in Educational Evaluation*, 37, 3-14.
- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49-65.