

數學後設認知量表之發展與信效度考驗*

涂金堂

高雄師範大學
師資培育中心

本研究旨在發展適用中小學生的數學後設認知量表 (Mathematical Metacognition Scale, MMS)，本研究分預試與正式施測兩階段，兩階段皆採立意取樣的叢集抽樣方式，各抽取中小學生 457 人與 861 人。預試資料進行探索性與驗證性因素分析，驗證性因素分析結果獲得三個因素：後設認知識、計劃與監控、評估，其內部一致性 Cronbach's α 係數分別為 .87、.87、.72，12 題的總量表 Cronbach's α 係數為 .90，顯示本量表具有良好的信度。正式施測資料，進行測量恆等性考驗，考驗結果發現不同性別具測量恆等性，顯示本研究發展的數學後設認知量表，具有良好的建構效度。

關鍵詞：探索性因素分析、測量恆等性、數學後設認知、驗證性因素分析

學習是複雜認知運作歷程，學習者須主動從長期記憶提取與新學習教材相關連的舊經驗，才能獲得有意義學習成果。在主動參與學習活動的歷程，學習者還需具備對學習活動進行計劃、監控、評估、調整的能力，才能獲得較有效的學習結果。此種對學習活動進行計劃、監控、評估、調整的運作機制，即是所謂的後設認知 (metacognition)。

「後設認知」一詞最早由 Flavell (1976) 提出，他主張後設認知是指「個體對自己認知歷程與結果的知識；與自己認知歷程與結果相關的知識；個體主動監控、整合和認知目標有關的歷程」。自從 Flavell 提出後設認知概念後，經過這 30 多年對後設認知的研究研究，顯示後設認知與學習具有密切關聯性 (Leutwyler, 2009)。Wang、Haertel 與 Walberg (1990) 以內容分析法，探討影響學生學習成果的 30 項相關因素，研究顯示後設認知是影響學生學習成果，非常重要的一項影響因素。已有許多研究結果顯示，後設認知對不同學習領域之學習活動，皆有很重大影響力，例如語文學習 (蘇宜芬、林清山, 1992; Cross & Paris, 1988; Mokhtari & Reichard, 2002; Pressley & Gaskins, 2006)、數學 (陳李綢, 1992; Carr, Alexander, & Folds-Bennett, 1994; Garofalo & Lester, 1985)、自然科學 (于富雲、陳玉欣, 2007; Kaberman & Dori, 2009; Rickey & Stacy, 2000)、社會科學 (Poitras, Lajoie, & Hong, 2012)、藝術 (Bathgate, Sims-Knight, & Schunn, 2012; Benton, 2013) 等。也因為後

* 本文通訊作者：涂金堂，通訊方式：tang@nkn.edu.tw。

設認知與學習兩者的密切關係，於 2006 年誕生一份新的學術期刊〈後設認知與學習〉(Metacognition and Learning)。

一、數學後設認知量表編製的重要性

對後設認知的研究，從 1980 年代開始，受到全世界數學教育界強調數學解題的重要性，因而數學教育學者著手探究後設認知在數學解題歷程中所扮演的角色 (Lester, 1982; Schoenfeld, 1985)。Desoete 與 Veenman (2006) 主張後設認知對數學解題的全部歷程，都具有重要的影響力。在一開始的數學解題歷程，不論是題意瞭解，或數學問題表徵，都需藉由後設認知的引導，才能順利完成。在數學解題歷程中，不論是擬定解題計畫，或執行解題計畫，也需要後設認知的監控，才能確定解題計畫是否合適。在數學解題最後階段，不論是解答驗算，或對解題成功之預測，同樣需要後設認知的評估，才能判斷解題是否成功。

有鑑於後設認知對數學解題的重要影響，新加坡教育部 (Ministry of Education, 2000) 在其公佈的數學課程中，提出數學解題是數學教學的核心目標，而數學解題則包含五項重要成份：後設認知 (metacognition)、歷程 (processes)、概念 (concepts)、技巧 (skills)、態度 (attitudes)，由此可知後設認知在數學解題中的重要性。

許多以後設認知所進行的教學實驗顯示，透過後設認知策略的訓練，可有效提高解題者的數學解題能力 (Kapa, 2001; Kramarski, 2004; Kramarski & Mevarech, 2003; Pennequin, Sorel, Nanty, & Fontaine, 2010)。這些研究結果顯示，不論是普通學生，或低成就的學生，皆可透過後設認知策略的訓練，達到提升數學解題的能力。

既然教導學生使用後設認知策略，能提升學生的數學解題能力。如何評估學生現有的後設認知能力，則成爲一項重要研究議題。Sperling、Howard、Miller 與 Murphy (2002) 提到兩個迫切需要測量學生的後設認知之理由：其一是許多以後設認知介入 (metacognitive intervention) 的教學實驗顯示有助於提升學生的學習表現，而有效評估學生的後設認知，是規劃適當介入活動的關鍵活動。其二透過對學生後設認知的測量，可以協助有關自我調整學習 (self-regulated learning) 的理論發展。

在評量學生的後設認知能力時，則須思考後設認知能力是屬於「領域普遍性」(domain-general) 或「領域特定性」(domain-specificity) 的問題。Veenman 與 Spaans (2005)，以及 Kelemen、Frost 與 Weaver (2000) 的研究結果顯示，後設認知能力應屬「領域特定性」。亦即學生面對不同學科領域，需採用不同的後設認知能力，例如閱讀理解的後設認知能力與數學解題的後設認知能力是不相同的。因而對後設認知能力之評估，應採「領域特定性」方式，才能更精準探究特定學科之後設認知能力。

雖然國內外學術界對後設認知的相關研究，已有三十多年時間，但對後設認知量表的發展，仍偏向以「領域普遍性」(domain-general) 方式編製量表 (Veenman & Spaans, 2005)。且在量表的效度考驗上，常只採用探索性因素分析 (林清山、張景媛, 1993; 涂金堂, 1996; Schraw & Dennison, 1994; Sperling et al., 2002)，而未能進一步採用驗證性因素分析。近年來，隨著結構方程模式 (structural equations modeling, SEM) 的快速發展，多位測驗統計學者 (Jackson, Gillaspay, & Purc-Stephenson, 2009; Noar, 2003) 主張探索性因素分析只能作為找尋量表較合適的因素個數，而驗證性因素分析則能檢驗透過探索性因素分析所獲得的因素結構模式，能否獲得實證資料支持。故後設認知量表的編製，不應只進行探索性因素分析，應更進一步進行驗證性因素分析，才能獲得更精準的建構效度考驗。

基於當前所編製的後設認知量表，在題目內容偏向採領域普遍性，在效度考驗缺乏驗證性因素分析等缺失。由於不成功的數學解題，常因解題者缺乏後設認知能力，且數學是中小學生較爲害怕的一門學科，故本研究主要以國內中小學生爲研究對象，針對特定領域的數學科，編製一份適用國中小學生的數學後設認知量表，期待能協助中小學數學教師，透過此份量表，可有效評量中小學生後設認知能力。可根據本後設認知量表的測量結果，考慮對學生進行後設認知策略的訓

練。本研究將透過探索性因素分析、驗證性因素分析與測量恆等性 (measurement invariance)、不同性別潛在平均數差異考驗、信度考驗等統計方法，來確定數學後設認知量表的適用性。

二、後設認知量表編製的理論基礎

(一) 後設認知理論內涵

有關後設認知的理論內涵，多位學者提出不同觀點，但大致可將歸納為兩大面向：認知的知識 (knowledge of cognition) 與認知的調整 (regulation of cognition)。認知的知識也被稱為後設認知知識 (metacognitive knowledge)，認知的調整也被稱為後設認知技能 (metacognitive skill) (Veenman, Van Hout-Wolters, & Afflerbach, 2006)。

1. 後設認知知識

Flavell (1987) 主張後設認知知識是指個人獲得與認知事物有關的世界知識，包括個人變項 (person variables)、工作變項 (task variables) 與策略變項 (strategy variables) 等三種後設認知知識。

個人變項的後設認知知識是指個體清楚瞭解個體內、個體間、普遍性等三種知識與信念。以數學解題為例，清楚自己比較喜歡算代數題目，比較不喜歡算幾何問題，即屬個體內的後設認知知識；瞭解與班上同學相比較，自己的數學解題能力，算是厲害的，即是個體間的後設認知知識；知道遇到不會算的數學題目，可搜尋腦海中是否有以前類似解題經驗，即是普遍性的後設認知知識。

工作變項的後設認知知識是關於認知作業的性質如何影響認知表現之知識。例如，知道數學應用問題比數學計算題更不容易得分。

策略變項的後設認知知識是指有關認知或後設認知策略的策略知識。例如，算幾何證明題時，知道可透過畫輔助線的策略，來協助順利完成幾何證明。

個人變項、工作變項與策略變項這三種後設認知知識，具密切的互動性，例如知道自己常容易計算錯誤 (個人變項)，一旦遇到較繁雜的計算題目 (工作變項)，會提醒自己放慢計算的速度 (策略變項)，此即為三種後設認知知識的交互作用結果。

Schraw 與 Dennison (1994) 從不同知識類型向度，主張後設認知知識的內涵，包含陳述性知識 (declarative knowledge)、程序性知識 (procedural knowledge)、條件性知識 (conditional knowledge) 等三種後設認知知識。陳述性知識是個體清楚自己身為一位學習者的知識，以及瞭解何種因素會影響個人認知表現的知識，例如知道學生的本份，就是上數學課時要專心聽講。程序性知識是指關於執行程序性技能的知識，例如遇到不會算的數學題目，會先透過找出題目重要訊息的方式，尋找解題方向。條件性知識是指知道何時與為何運用多種認知行動的知識，例如個體會根據不同的題目類型選擇不同的解題策略。

Larkin (2010) 綜合比較 Flavell (1987) 主張的後設認知知識和 Schraw 與 Dennison (1994) 主張的後設認知知識之異同，Larkin 認為 Schraw 與 Dennison 主張的陳述性知識，類似 Flavell 主張的個人變項與工作變項的後設認知知識；Schraw 與 Dennison 的程序性知識，類似 Flavell 的策略變項後設認知知識。至於 Schraw 與 Dennison 的條件性知識，則偏向個體對自己思考的監控與控制 (monitoring and control of our thinking)，是比較偏向後設認知技能的部分。

綜合上述討論，Flavell (1987) 和 Schraw 與 Dennison (1994) 主張的後設認知知識，兩者相近的部分是陳述性知識與程序性知識。本研究認同 Larkin (2010) 的主張，認為條件性知識比較偏向後設認知技能的監控。由於本研究認為後設認知技能的「監控」比「條件性知識」更強調「個體在知道什麼情況下，採取什麼活動」的動作，為避免產生區辨效度的問題，故本研究編製的後設認知量表，有關後設認知知識只包含陳述性知識與程序性知識兩部分。

2. 後設認知技能

Silver (1985) 認為在數學解題歷程中，若解題者若缺乏後設認知技能，則容易導致解題的失敗。例如在計算「全校有 300 位學生參加戶外教學活動，已知每輛遊覽車最多只能載 45 位同學，

請問學校至少應該租幾輛遊覽車？」的數學題目時，當解題者的答案為 6 又 $\frac{2}{3}$ 輛，顯示解題者只具備數字運算能力，缺乏後設認知技能的監控能力，無法判斷 $\frac{2}{3}$ 輛是不合理的答案，應該至少租 7 輛。

後設認知技能主要包含計畫 (planning)、監控 (monitoring) 與評估 (evaluation) 等三種技能 (Leutwyler, 2009; Schraw, Crippen, & Hartley, 2006; Schraw & Moshman, 1995)

計畫是指個體能選擇合適的策略與合理分配認知資源，例如在開始算數學題目前，解題者懂得先花一些時間思考如何著手，並懂得進行適當的時間分配。監控是指個體對所從事的認知活動之覺察歷程，例如上數學課時，學習者會隨時尋問自己是否聽懂老師所講的內容。評估是指對個體從事認知活動的結果，所進行的評判活動，例如解題者在算出數學題的答案後，會評判自己是否能答對該題。

綜合上述可知，學者專家對於後設認知的內涵，大多主張包含後設認知知識與後設認知技能等部分。後設認知知識比較強調個體對自身認知能力的瞭解、對認知作業類型的掌握，以及對認知策略的活用，這些後設認知知識是個體透過不斷進行數學解題活動所累積的，它是個體在進行數學解題活動時，已經事先具備的知識。而後設認知技能比較強調個體從事認知活動當下的所進行的計畫、監控、修正、評估等行爲。個體在進行數學解題活動時，若能伴隨出現計畫、監控、修正、評估等行爲，則越能協助個體獲得成功的解題活動。

(二) 後設認知的測量方式

數學領域的後設認知之測量，較常採問卷調查 (questionnaire)、晤談 (interview)、信心評量 (confidence ratings) 與放聲思考 (think-aloud) 等四種方式 (Jacobse & Harskamp, 2012)。

1. 問卷調查法

問卷調查法是透過事後回溯的技術 (retrospective techniques)，請受試者根據問卷題目內容，回憶其在數學學習或數學解題時，採用的後設認知策略。以受試者在問卷得分高低，評估其數學後設認知能力。此方法的優勢是實施方便，適合大樣本調查；缺點是採事後回溯法，易造成受試者的回答內容，與其在數學學習或數學解題時，所實際採用的後設認知出現不一致情形。

2. 晤談法

晤談法是被訪談者在解題結束後，由訪談者透過結構式問題，與被訪談者進行晤談。主試者藉由晤談方式，深入瞭解被訪談者在解題活動中，所產生的後設認知行爲。訪談者常針對被訪談者在解題歷程中，有關解題計畫擬定、執行、監控與解題結果評估等歷程，請被訪談者完整說明當下心理運作歷程，以探討被訪談者當時的後設認知表現。此方法優點是能較深入瞭解被訪談者後設認知運作情形，缺點為費時，且易造成被訪談者口述內容，超過實際後設認知運作情況。

3. 信心評量法

信心評量法是讓受試者先閱讀題目或完成解題工作，在尚未公布正確答案前，先行預測自己可能答對的題數。再比較受試者預測結果與實際作答結果，計算其預測正確性，由預測正確性來評估其後設認知能力 (Schwartz & Metcalfe, 1992)。此方法優點是可實際測量解題者有關預測與評估之後設認知能力，缺點是無法瞭解解題者解題歷程中的其他後設認知能力 (監控、調整、修正) 的運作情形。

4. 放聲思考法

放聲思考法是要要求受試者在解決數學題目時，將腦中所有思考運作情形，同步以語言大聲地口述出來，然後根據受試者的思考內容與歷程，評估其後設認知表現。採用放聲思考法時，通常會採錄音與錄影器材，錄下受試者的解題情景。將受試者錄下的口述內容，轉譯成書面資料的原案 (protocols)，最後根據原案內容，進行後設認知能力的評估。

此方法優點是能蒐集受試者解題歷程的後設認知運作情形，配合錄影帶的觀察，可獲得較客觀的分析結果，缺點是口述的同時可能干擾受試者後設認知運作 (Rosenzweig, Krawec, & Montague, 2011)。

上述四種評量數學領域後設認知的的方法，各有其優缺點。若想快速掌握大樣本受試者的後設認知能力，較適合採用問卷調查法或信心評量法，若想更深入了解受試者的解題歷程，則較適合採用晤談法或放聲思考法。本研究想編製一份協助快速評估學生在數學領域後設認知能力的量表，故採用問卷調查法進行後設認知的評量。

(三) 現有後設認知量表的探究

Schraw 與 Dennison(1994)編製 52 題的「後設認知覺知量表」(Metacognitive Awareness Inventory, MAI),是較常被採用的後設認知量表(Muis, Winne, & Jamieson-Noel, 2007)。Schraw 與 Dennison 以 197 位美國大學生,採用 100 公厘的兩極端量尺答題型態(左邊代表題目內容完全符合自己的情形,右邊代表題目內容完全不符合自己的情形),請受試者根據自己感受,在 100 公厘線上,畫出適當位置。該量表將後設認知分成後設認知知識與後設認知技能兩部分。後設認知知識包含陳述性知識、程序性知識、條件性知識等三分量表,後設認知技能包含計畫、訊息處理策略(information management strategies)、監控、除錯策略(debugging strategies)與學習評估(evaluation of learning)等五分量表。透過探索性因素分析,研究後設認知覺知量表到底是兩因素(後設認知知識、後設認知技能)或八因素(陳述性知識、程序性知識、條件性知識、計畫、訊息處理策略、監控、除錯策略、學習評估)較符合實證資料,研究結果發現兩因素之因素分析結果較合適,兩因素的 α 係數分別為 .88 與 .91,顯示具有良好信度。

雖然 Schraw 與 Dennison(1994)研究結果顯示兩因素較合適,但有些題目的因素負荷量過低,例如第 4、34、40、48 等四題,其最大因素負荷量分別為 .00、.36、.34、.00,皆未超過 Pett、Lackey 與 Sullivan(2003)建議因素負荷量高於 .40 的標準,但卻未刪除這些有問題的題目,因而易造成聚斂效度的降低,導致「後設認知覺知量表」的建構效度出現問題。並且只進行探索性因素分析,並未進行驗證性因素分析,無法確定兩因素之因素結構模式,能否獲得實證資料支持。

Sperling 等人(2002)從 Schraw 與 Dennison(1994)編製適用大學生的「後設認知覺知量表」(MAI),分別挑選並改寫 12 題與 18 題,以 144 位小學三到五年級學生,接受 A 版本(12 題)「少年版後設認知覺知量表」(Jr. MAI);以 200 位小學六到九年級學生,接受 B 版本(18 題)「少年版後設認知覺知量表」(Jr. MAI)。採用探索性因素分析,透過指定兩因素的因素分析法, A 版本與 B 版本皆抽取後設認知知識與後設認知技能兩因素,但同樣出現有些題目之因素負荷量過低, A 版本的第 4、5、8 題等三題,其最大因素負荷量分別為 .26、.29、.39; B 版本的第 2、4 題等兩題,其最大因素負荷量分別為 .37、.38,同樣皆未超過 Pett、Lackey 與 Sullivan(2003)建議因素負荷量高於 .40 的標準,但卻未刪除這些有問題的題目,故同樣存在建構效度的問題。

Panaoura 與 Philippou(2003)以 246 位四至六年級學生,參考 Sperling 等人(2002)的「少年版後設認知覺知量表」,發展適合國小學生的數學後設認知能力量表(metacognitive ability in mathematics)。一開始編製 30 道題目,透過探索性因素分析,獲得九因素共 25 題。進一步透過驗證性因素分析,最後獲得 8 題「後設認知知識」與 7 題「後設認知調整」兩分量表,另外還有 3 題(A12、A21、A25)同時橫跨後設認知知識與後設認知調整兩因素。總量表共 15 題,其信度考驗的 α 係數為 .83,顯示具良好信度。但因保留橫跨兩因素的 3 題,易造成聚斂效度的降低,而影響量表的建構效度。

林清山與張景媛(1993),以 126 位國中生,編製測量國中生的數學後設認知量表,該量表包含「目標設定」、「自我監控」、「自我評鑑」與「自我修正」四分量表。每分量表各有 8 題,採 Likert 四點量表的答題型態。量表的建構效度是以每分量表試題,單獨進行探索性因素,因素分析結果顯示各分量表皆獲得一個因素,表示該量表具適切建構效度。量表信度考驗方面,四分量表的 α 係數分別為 .85、.87、.84、.88,顯示具良好信度。林清山、張景媛編製的後設認知量表,只進行探索性因素分析,因未進一步進行驗證性因素分析,無法確定每分量表的一因素之因素結構模式,是否能獲實證資料支持。

涂金堂(1996)直接採用林清山與張景媛(1993)後設認知量表,應用於 438 位國小六年級學生。在量表的建構效度方面,同樣採單獨分量表之探索性因素分析,研究結果顯示各分量表只抽取一個因素,顯示該量表具良好建構效度。在量表信度考驗,四各分量表的 α 係數分別為 .74、.76、.78、.88,顯示具良好信度。涂金堂同樣只進行探索性因素分析,未進行驗證性因素分析,故同樣無法得知因素結構模式是否獲實證資料支持。

綜合上述對現有後設認知量表的探討可知,在量表效度考驗方面,有些量表保留因素負荷量過低的題目、有些量表保留同時橫跨兩因素的題目,如此都會導致量表建構效度的降低。同時,較多量表只採用探索性因素分析,較少採用驗證性因素分析。根據 Jackson 等人(2009),以及 Noar

(2003)等測驗統計學者的主張，量表編製除了進行探索性因素分析，應進一步進行驗證性因素分析，才能提出更具說服力的建構效度證據。

另外，男女生在後設認知能力上是否有顯著性差異，一直是許多研究關注的重點(涂金堂, 1996; 陳李綱, 1992; Leutweiler, 2009; Mok, Fan, & Pang, 2007; Zimmerman & Martinez-Pons, 1990)。在比較不同群體受試者在某個量表得分的差異情形時，若該量表的因素個數、因素負荷量、因素共變數等測量結果，不會因不同組別受試者而有所差異，則較能獲得正確的統計結果，此即所謂測量恆等性。然而許多針對男女生後設認知量表得分的差異性考驗，並未進行測量恆等性的考驗。

針對現有後設認知量表在效度考驗的限制，本研究決定以領域特定的數學科，編製一份適用中小學生的數學後設認知量表。在效度考驗方面，先透過專家審題，以確保量表具適切的內容效度。再透過探索性因素分析，先找出量表的因素結構。進一步透過驗證性因素分析，考驗由探索性因素分析獲得之因素結構模式，是否獲實證資料支持。再透過測量恆等性考驗，以評估數學後設認知量表，對不同性別的國中男女生，是否具有相同的因素個數、因素負荷量、因素共變數。最後則進行不同性別潛在平均數差異性考驗，以探就不同性別的國中男女生，在數學後設認知量表的潛在平均數是否有顯著性差異。

方法

一、研究對象

本研究的母群以全國五、六、七與八年級學生為研究對象，本研究分成預試樣本與正式樣本。預試樣本用以進行項目分析、探索性因素分析與驗證性因素分析；正式樣本進行測量恆等性考驗與不同性別的分數差異性考驗。

本研究預試樣本採叢集抽樣方法，從北、中、南、東等四個地區，各抽選 1 所國中（七年級與八年級各 1 班）與 1 所國小（五年級與六年級各 1 班）。國中每班寄發 35 份問卷，國小每班寄發 30 份問卷，共寄發 520 份問卷，扣除無效問卷，回收有效問卷為 457 份（國中 250 份、國小 207 份），有效問卷回收率為 87.88%。

正式樣本同樣採叢集抽樣方法，從北、中、南與東等四個地區，各抽取 2 所國中（每所國中的七年級與八年級各 1 班）與 2 所國小（每所國小的五年級與六年級各 1 班）。國中每班寄發 32 份問卷，國小每班寄發 28 份問卷，共寄發 960 份問卷，扣除無效問卷，回收有效問卷為 861 份（國中 432 份、國小 429 份），有效問卷回收率為 89.69%。

本研究進行抽樣時，除考量不同地區外，也考量學校規模大小，故同時抽取包含大型、中型與小型學校的學生，讓樣本更具代表性。但針對北、中、南、東四區的抽樣學校數，是採每區域抽取相同的班級數，而非採分層比例的抽樣方式，如此可能產生推論上的限制。

後設認知量表同時採用探索性因素分析與驗證性因素分析，以評估量表的建構。本研究根據 Clark 與 Watson(1995)建議探索性因素分析人數介於 100 至 200 人即可，以及 Schumacker 與 Lomax (1996) 研究發現多數進行結構方程模式人數是介於 250 至 500 人，故本研究將預試樣本 457 位受試者，透過 SPSS 的隨機取樣程序，隨機選取 200 位受試者進行探索性因素分析，257 位受試者進行驗證性因素分析。將正式樣本 861 份，進行測量恆等性考驗與不同性別的分數差異性考驗。

二、量表編製歷程

「後設認知量表」的編製，首先依據後設認知理論之相關文獻，並參考 Schraw 與 Dennison (1994)，以及 Sperling 等人(2002)所編製的「後設認知覺知量表」，由於本研究認同 Larkin(2010)主張後設認知知識中的「條件性知識」，較偏屬於後設認知技能的監控與控制，且本研究認為「監

控」分層面比「條件性知識」分層面，更強調「個體在知道什麼情況下，採取什麼活動」的動作，為避免產生區辨效度的問題，故未採用「條件性知識」分層面。本研究的後設認知量表包含後設認知知識與後設認知技能等兩部分，其中後設認知知識包含「陳述性知識」與「程序性知識」兩個分量表，後設認知技能包含「計畫」、「監控」與「評估」等三個分量表，故本研究將後設認知量表的分層面，設定為「陳述性知識」、「程序性知識」、「計畫」、「監控」、「評估」等五個分量表。

「陳述性知識」指受試者瞭解自己的技能、認知資源和能力之後設認知知識，例如第 7 題「遇到困難的數學題目時，我會花較多時間去了解題意」。「程序性知識」指受試者懂得如何逐步執行策略或步驟之後設認知知識，例如第 10 題「遇到不會解的數學題目，我會先找出題目的重要訊息」。

「計畫」指受試者懂得在學習歷程或解題歷程前，進行目標設定、規劃學習或解題步驟。例如第 20 題「學習新的數學單元時，我會設定學習目標」。「監控」指受試者懂得在學習歷程或解題歷程中，檢視自己對策略使用、計畫執行，是否有錯誤之處，例如第 26 題「準備數學考試時，我會隨時檢視自己的複習進度」。「評估」指受試者懂得在學習歷程或解題歷程後，評判自己學習效果或解題效果，是否有需要修正或調整的地方。例如第 29 題「解出數學題目的答案後，我會進行驗算以確定答案是否正確」。

後設認知量表每個分量表各編寫 8 題，總量表共 40 題。本量表採李克特 5 點量表計分方式，從非常同意（5 分）、同意（4 分）、不確定（3 分）、不同意（2 分）、非常不同意（1 分），受試者得分數愈高，表示其後設認知能力越強。

初擬試題完成後，聘請 8 位學者專家，針對題目內容適切性（適切、不適切或需修改），進行初擬試題之審題，以檢定本量表的內容效度。本研究採 Lester 與 Bishop（2000）建議，保留 80% 以上專家勾選屬於「適切」的題目，結果刪除 5 題，保留 35 題進行預試。

預試與正式樣本的施測，委請施測班級導師協助，為避免不同施測程序影響量表的心理計量特質，請協助施測的班級導師，先進行施測指導語說明，才讓受試者開始答題。

三、資料分析軟體

本研究採用 SPSS 21.0，作為進行項目分析、探索性因素分析、信度分析的統計軟體，以 AMOS 21.0 進行驗證性因素分析、測量恆等性與潛在平均數差異分析考驗。

研究結果

一、探索性因素分析

探索性因素分析進行前，為確保題目品質不致影響因素分析結果，先進行項目分析，先將品質不佳的題目刪除後，才進行探索性因素分析。

（一）項目分析

本研究「後設認知量表」的項目分析評判指標，主要以「決斷值」與「校正後題目與總分相關」這兩項（涂金堂，2012；Spector, 1992）。若這兩項指標皆屬不良指標，即刪除該題。

1. 決斷值的評判依據

以量表總分最高前 27% 為高分組，量表總分最低後 27% 為低分組，進行每道題目獨立樣本 t 考驗，若高低分組平均數之差異，顯著性 p 值未小於 .05，則為不良指標。

2. 校正後題目與總分的相關的評判依據

若每題的校正後題目與總分相關係數低於 .30，則為不良指標。

預試的 35 題「後設認知量表」，經項目分析後，所有題目在「決斷值」與「校正後題目與總分相關」這兩項指標，皆沒有任何題目出現不良指標，故保留 35 題進行探索性因素分析。

(二) 探索性因素分析

本研究編製的「後設認知量表」主張「陳述性知識」、「程序性知識」、「計畫」、「監控」、「評估」等五個分量表是彼此具關聯性，故進行探索性因素分析時，採「主軸因子」(principal axis factor)的共同因素萃取法，配合「直接斜交」的斜交轉軸(oblique rotation)方式。

在因素個數決定方法上，則綜合 Kaiser 特徵值大於 1、Cattell 的陡坡圖(scree plot)、Horn(1965)平行分析法(parallel analysis)，以及後設認知量表的分量表個數，決定最後的因素個數。

「後設認知量表」探索性因素分析結果顯示，KMO 為 .93，且 Bartlett 檢定統計值 3553.11，顯著性 p 值小於 .05，顯示取樣適當，適合進行因素分析。

在決定因素個數方面，因有 5 個特徵值大於 1，若採用 Kaiser 特徵值大於 1 的標準，則保留 5 個因素；而透過陡坡圖的判斷是 3 個因素較合適；若根據 O'Connor (2000) 提供的平行分析法之 SPSS 語法，平行分析法統計結果如表 1。由表 1 可知，有 2 個因素分析的實際特徵值，大於模擬特徵值，顯示平行分析法應萃取 2 個因素較合適；而由於本量表包含 5 個分量表，故選取 5 個因素較為合適。故綜合上述特徵值大於 1、陡坡圖、平行分析法，以及分量表個數，本研究最後決定分別選取 5 個、3 個、2 個因素。再根據這三種因素分析結果，決定較合適的因素個數。

表 1 平行分析法的因素個數決定

特徵值 序號	真實資料 的特徵值	O'Connor (2000) 模擬的特徵值	因素保留 結果
1	13.64	1.95	保留
2	2.76	1.82	保留
3	1.56	1.73	不保留
4	1.26	1.65	不保留
5	1.13	1.58	不保留

採萃取 5 個因素的統計結果顯示，第 4 個因素只有 2 題，第 5 個因素只有 1 題，由於 1 道題目無法計算分量表的信度，故 5 個因素的統計結果並不適切。而採用 3 個因素的統計結果顯示，「陳述性知識」與「程序性知識」兩分量表題目聚集成同一因素，「計畫」與「監控」兩分量表題目也聚集成同一因素，而「評估」則單獨自成一因素。若採 2 個因素的統計結果顯示，「陳述性知識」與「程序性知識」兩分量表題目聚集成同一因素，而「計畫」、「監控」與「評估」三分量表題目也聚集成同一因素。

由於選取 3 個因素與選取 2 個因素的因素分析結果，皆能符合後設認知的理論架構，但因 3 個因素能更細分認知監控與評估的部分，故本研究決定採用 3 個因素之因素結構。將「陳述性知識」與「程序性知識」兩分量表合併改稱「後設認知知識」分量表，「計畫」與「監控」兩分量表合併改稱「計畫與監控」分量表，而「評估」分量表維持原有名稱。

確定 3 因素個數後，採主軸因子萃取法配合直接斜交轉軸法，經過多次因素分析後，將因素負荷量小於 .40 (Pett, Lackey, & Sullivan, 2003; Stevens, 2002)、出現同時橫跨兩個以上因素、不符合原有因素向度的題目刪除。結果刪除 22 題，保留 13 題，如表 2 所示。

由表 2 可知，因素一屬於「後設認知知識」分量表，特徵值 5.96，可解釋變異量為 45.82%，包含第 7、8、9、10、11 等 5 題；因素二屬「計畫與監控」分量表，特徵值 1.77，可解釋變異量為 13.58%，包含第 18、19、20、21、26 等 5 題；因素三屬「評估」分量表，特徵值為 0.98，可解釋變異量為 7.53%，包含第 29、30、34 等 3 題。而三個因素總解釋變異量為 66.93%，顯示後設認知量表的「後設認知知識」、「計畫與監控」與「評估」等三因素，可以適切解釋後設認知量表 13 題之變異量。

表 2 後設認知量表探索性因素分析之統計結果 (N = 200)

題目	後設認知知識	計畫與監控	評估
	樣式係數 (結構係數)	樣式係數 (結構係數)	樣式係數 (結構係數)
7. 遇到困難的數學題目時，我會花較多時間去了解題意。	.63 (.71)	-.01 (.36)	.15 (.51)
8. 解數學題目時，我會先確定是否了解題目的意思。	.73 (.72)	-.03 (.32)	.01 (.43)
9. 不了解數學題目的意思時，我會多讀幾遍來了解題意。	.72 (.76)	.13 (.44)	-.04 (.45)
10. 遇到不會解的數學題目，我會先找出題目的重要訊息。	.71 (.80)	-.01 (.40)	.17 (.58)
11. 解數學題目時，我會確了解題意後，才開始進行解題的動作。	.82 (.79)	-.01 (.35)	-.05 (.43)
18. 準備數學考試時，我會先擬定複習計畫。	.08 (.44)	.82 (.84)	-.04 (.44)
19. 我會規劃寫數學作業的時間。	-.05 (.32)	.72 (.73)	.07 (.42)
20. 學習新的數學單元時，我會設定學習目標。	.04 (.39)	.82 (.82)	-.04 (.41)
21. 解數學問題時，我會整合與解題相關的訊息，以規劃解題方向。	.40 (.61)	.47 (.64)	-.01 (.46)
26. 準備數學考試時，我會隨時檢視自己的複習進度。	-.03 (.37)	.72 (.77)	.12 (.48)
29. 解出數學題目的答案後，我會進行驗算以確定答案是否正確。	.26 (.58)	-.07 (.36)	.61 (.72)
30. 解數學題目後，我會再想想是否有更簡單的解題方法。	.01 (.39)	.04 (.37)	.61 (.64)
34. 解數學題目後，我會思考在解題歷程中運用了那些數學知識。	-.03 (.38)	.28 (.51)	.48 (.61)
未轉軸的特徵值	5.96	1.77	0.98
未轉軸的解釋變異量	45.82%	13.58%	7.53%
轉軸平方和負荷量	4.46	4.15	3.44

註：採用主軸因子法配合直接斜交轉軸法，粗體的數據代表樣式係數或結構係數高於 .40。

(三) 信度分析

將探索性因素分析獲得的 13 題，進行 Cronbach's α 係數考驗，以判斷量表題目一致性情形。研究結果顯示，後設認知總量表 α 係數 .90，「後設認知知識」、「計畫與監控」與「評估」等三分量表的 α 係數，分別為 .87、.87、.69，除了「評估」分量表因題數少，而影響其 α 係數未符合 Nunnally 與 Bernstein (1994) 建議應高於 .70 的標準，但仍接近 .70 的標準，其它兩分量表的 α 係數皆符合 Nunnally 與 Bernstein 的建議標準，顯示後設認知總量表與各分量表，具良好信度。

二、驗證性因素分析

本研究以預試第二部分的 257 位樣本，根據探索性因素分析獲得的 13 題三個相關因素之因素結構，進行驗證性因素分析。遺漏值處理方式，採統計軟體 SPSS 21.0 的「序列平均數」遺漏值置換方式處理。

首先，在常態分配檢定方面，Kline (2005) 建議若偏態指數 (skew) 絕對值小於 3，峰度指數 (kurtosis) 絕對值小於 8，則沒有嚴重的常態分配問題。本研究後設認知量表 13 題之偏態指數介於 -1.13 至 -0.10，峰度指數介於 -0.58 至 1.70，顯示這 13 題並未違反常態分配基本假定。

其次，在因素結構模式之適配性方面，Bagozzi 與 Yi (1988) 建議評估驗證性因素分析模式之適切性，應同時考量初步適配 (preliminary fit criteria)、整體模式適配 (overall model fit)、模式內在結構適配 (fit of internal structural of model) 等三個部分。

1. 初步適配情形

統計結果顯示所有變異數皆顯著不為 0，且沒有任何負值的變異數；因素負荷量介於 .67 至 .81 之間，所有題目之因素負荷量皆高於 .50，且沒有任何因素負荷量高於 .95 之情況 (Bagozzi & Yi, 1988)；也沒有高過 1 的積差相關係數，顯示初步適配情形良好。

2. 整體適配情形

在模式整體適配部分，根據 Jackson 等人 (2009) 研究結果發現，驗證性因素分析整體模式適配性評判指標，以 χ^2 卡方考驗 (含卡方值與 p 值)、CFI、RMSEA、TLI 與 GFI 等五項，是最常被採用的評判指標。故本研究以這五項適配度指標，以評判模式整體適配情形。驗證性因素分析統計結果顯示， $\chi^2 = 127.303$ ， $p < .001$ 、CFI = .960、RMSEA = .06、TLI = .949、GFI = .932。由五項整體適配度指標顯示， χ^2 與 RMSEA 這兩項指標屬不良適配情形，而 CFI、TLI 與 GFI 等三項指標皆屬良好適配，故就模式整體適配情形而言，13 題後設認知量表驗證性因素分析模式尚屬良好適配。

由於整體模式適配未達良好適配情形，乃透過 AMOS 提供之修正指標 (modification indices)：若刪除第 a21 題能降低卡方考驗的數值，再配合考量後設認知的理論文獻，最後決定刪除第 a21 題，最後變成 12 題三個相關因素模式。這 12 題之積差相關係數矩陣，如表 3 所示。由表 3 可知，12 題彼此兩兩之積差相關係數介於 .30 至 .65 之間，顯示所有題目皆具中度正相關。

表 3 後設認知量表 12 題之積差相關係數矩陣

	1	2	3	4	5	6	7	8	9	10	11	12
1. a7	—											
2. a8	.51***	—										
3. a9	.61***	.62***	—									
4. a10	.58***	.48***	.58***	—								
5. a11	.49***	.58***	.60***	.56***	—							
6. a18	.41***	.33***	.37***	.38***	.40***	—						
7. a19	.34***	.34***	.32***	.30***	.43***	.60***	—					
8. a20	.44***	.37***	.38***	.40***	.42***	.57***	.59***	—				
9. a26	.40***	.40***	.43***	.34***	.42***	.64***	.65***	.55***	—			
10. a29	.40***	.34***	.43***	.37***	.41***	.37***	.41***	.53***	.51***	—		
11. a30	.39***	.39***	.38***	.33***	.40***	.45***	.36***	.42***	.46***	.43***	—	
12. a34	.39***	.39***	.39***	.38***	.38***	.40***	.45***	.48***	.46***	.51***	.52***	—
平均數	3.92	3.98	3.98	3.90	3.92	3.19	3.14	3.13	3.40	3.47	3.46	3.47
標準差	0.95	0.85	0.95	0.94	0.97	1.12	1.11	1.07	1.10	1.09	1.15	1.11

本研究編製的後設認知量表所進行的驗證性因素分析，最後獲得 12 題三個相關因素之因素結構模式圖，如圖 1 所示。

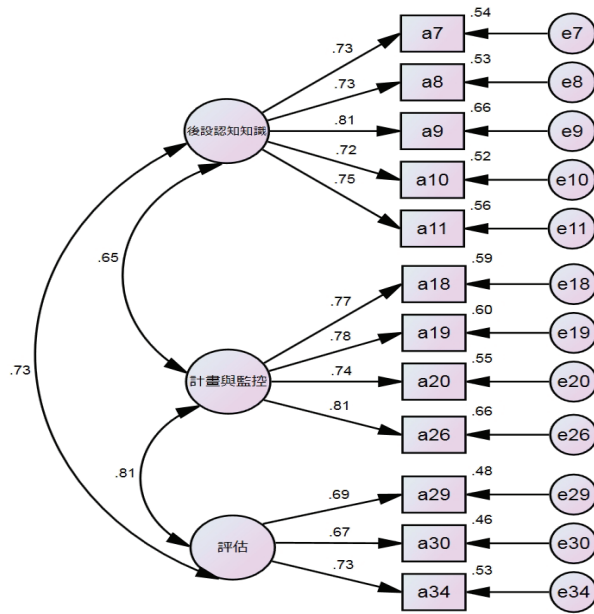


圖 1 12 題後設認知量表的驗證性因素分析結果

12 題後設認知量表之五項整體適配度摘要結果，如表 4 所示。由表 4 可知，除易受到樣本人數影響的 χ^2 這項指標屬不良適配情形（余民寧，2006），其它 CFI、RMSEA、TLI 與 GFI 等四項指標皆屬良好適配，故就整體適配情形而言，12 題後設認知量表驗證性因素分析模式屬於良好適配。

表 4 12 題後設認知量表驗證性因素分析的模式整體適配指標之適配情形

模式	$\chi^2 (p)$	CFI	RMSEA	TLI	GFI
三個相關因素模式	79.306 ($p = .007$)	.980	.047	.975	.953
評判標準	$p > .05$	$> .90$	$< .05$	$> .90$	$> .90$

註：評判標準採用余民寧（2006）所提的建議判斷標準

3. 模式內在結構適配情形

就模式內在結構適配情形方面，Hair、Black、Babin 與 Anderson（2010）主張可分別透過題目信度（item reliability）、平均變異數抽取量（average of variance extracted）與組合信度（composite reliability）等三項指標判斷。Hair 等人建議題目信度至少要高於 .25、平均變異數抽取量高於 .50、組合信度高於 .70，即符合模式內在結構適配的適切標準。

後設認知量表 12 題三個相關因素驗證性因素分析模式之題目信度、組合信度與平均變異數抽取量等三項指標的統計結果，如表 5 所示。

表 5 後設認知量表 12 題之題目信度、組合信度與平均變異數抽取量摘要表

因素	題目	因素負荷量	題目信度	組合信度	平均變異數抽取量
後設認 知知識	7	0.73	0.54	.87	.56
	8	0.73	0.53		
	9	0.81	0.66		
	10	0.73	0.53		
計畫與 監控	11	0.75	0.56	.86	.60
	18	0.77	0.59		
	19	0.78	0.60		
	20	0.74	0.55		
評估	26	0.81	0.66	.74	.49
	29	0.73	0.48		
	30	0.68	0.46		

由表 5 可知，根據 Hair 等人（2010）建議標準，後設認知量表的 12 題信度，皆符合題目信度高於 .25 的標準。後設認知量表三分量表（「後設認知知識」、「計畫與監控」、「評估」），其組合信度皆高於 .70 標準。而平均變異數抽取量方面，除「評估」分量表為 .49 接近 .50 的標準，其它兩分量表之平均變異數抽取量皆高於 .50，故顯示 12 題三個相關因素之驗證性因素分析模式具良好內在結構適配情形。

在後設認知量表的區辨效度考驗方面，除「評估」分量表之平均變異抽取量未高於「後設認知知識」與「評估」兩者積差相關係數之平方，以及「計畫與監控」與「評估」之平均變異抽取量未高於兩者積差相關係數之平方外，其它分量表的平均變異抽取量，皆高於兩分量表的積差相關係數平方，顯示本量表具有尚佳的區辨效度。

綜合上述初步適配、整體適配與內在結構適配等三項驗證性因素分析模式適配性的評估，可知 12 題三個相關因素的後設認知量表，其驗證性因素分析模式獲得實證資料支持。

三、複核效度考驗分析統計結果

經過以 257 位預試樣本所建構的後設認知量表 12 題三個相關因素的驗證性因素分析模式，接續便根據此模式，以 861 位正式樣本進行複核效度考驗，以考驗 12 題的後設認知量表因素模式是否適用於不同的樣本。

進行複核效度考驗時，應先確認預試樣本與正式樣本因素結構模式是否獲實證資料支持。表 6 為預試樣本、正式樣本後設認知量表驗證性因素分析整體適配指標，除 χ^2 屬不良適配，其他 CFI、RMSEA、TLI 與 GFI 等四項皆屬良好適配，顯示不論是預試樣本或正式樣本，後設認知量表皆具良好適配。

表 6 12 題後設認知量表驗證性因素分析整體適配指標摘要表

樣本類型	樣本數	$\chi^2 (p)$	CFI	RMSEA	TLI	GFI
預試樣本	257	79.306 (.007)	.980	.047	.975	.953
正式樣本	861	163.210 (<.001)	.978	.051	.971	.970
評判標準		$p > .05$	> .90	< .05	> .90	> .90

註：評判標準採余民寧（2006）所提建議標準

由於後設認知量表的三個相關因素之因素結構，不論是對預試樣本或正式樣本受試者，皆顯示具適配情形。接續進行複核效度考驗，統計結果如表 7。

表 7 12 題後設認知量表之複核效度的模式比較結果 (N = 1,118)

模式類型	模式比較	χ^2 (p)	df	CFI	$\Delta\chi^2$	Δdf	p	ΔCFI
模式 A (相同型態模式) (沒有設定任何參數相等)	—	242.581 ($< .001$)	102	.978	—	—	—	—
模式 B (測量模式) (設定因素負荷量相等)	B 比 A	250.547 ($< .001$)	111	.979	7.966	9	.538	.001
模式 C (結構模式) (設定因素負荷量相等， 因素間的共變數相等)	C 比 B	255.274 ($< .001$)	117	.979	4.727	6	.579	$< .001$

註： $\Delta\chi^2$ 表兩個模式的卡方值之差； Δdf 表兩個模式的自由度之差； ΔCFI 表兩個模式的 CFI 之差

由表 7 可知，測量模式與相同型態模式相比較，兩者的卡方值相差 7.966，自由度相差 9，顯著性為 .538；而兩者的 CFI 相差 .001，小於 Cheung 與 Rensvold (2002) 建議低於 .01 差異標準，顯示測量模式獲得實證資料支持，亦即預試樣本與正式樣本具有相同的因素負荷量。

結構模式與測量模式相比較，兩者的卡方值相差 4.727，自由度相差 6，顯著性為 .579，由於卡方考驗的顯著性低於 .05，顯示結構模式未獲得實證資料支持。但由兩者 CFI 相差小於 .001，小於 Cheung 與 Rensvold (2002) 建議低於 .01 差異標準，顯示結構模式獲得實證資料支持。由於卡方考驗容易受到樣本人數的影響，而本研究屬於大樣本，採用 ΔCFI 指標會較合適，故本研究的結構模式獲得實證支持。

由上述複核效度考驗結果可知，不論預試樣本與正式樣本，後設認知量表皆能獲得相同的三個因素，並且每個因素與每道題目的因素負荷量皆相同，且因素之間具有相同的共變數。綜合上述的討論，可知 12 題三個相關因素之後設認知量表，具有良好且穩定的因素結構。

四、測量恆等性分析

同一份量表對不同群體受試者，是否出現不同的測量模式，是量表編製很重要的檢驗步驟，即所謂測量恆等性考驗 (measurement equivalence)，也稱為測量不變性 (measurement invariance)。早期對量表測量恆等性考驗，較常採探索性因素分析，比較不同組別受試者的因素負荷量是否相似。近年來，隨著 SEM 統計分析技術的快速發展，較偏向採多組別驗證性因素分析 (multiple-group confirmatory factor analysis) 的考驗模式，探討不同組別受試者，在因素結構模式中的因素個數、因素負荷量、迴歸分析截距、因素變異數、因素共變數、殘差變異數等是否相同 (Nimon & Reio, 2011)。

以多組別驗證性因素分析進行測量恆等性考驗時，較常採「相同型態恆等性」(configural invariance)、「量尺恆等性」(metric invariance) 與「純數恆等性」(scalar invariance) 等三種考驗方式 (Gregorich, 2006; Schmitt & Kuljanin, 2008)。另外，「因素的共變數恆等性」(invariance of factor covariances)，也是較常被進行的測量恆等性考驗 (Steinmetz, Schmidt, Tina-Booh, Wiczorek, & Schwartz, 2009)。故本研究即針對不同性別的受試者，進行上述四種測量恆等性之考驗。

「相同型態恆等性」考驗是探究不同組別受試者在同一份量表上，其每個因素所具有的題數是否相同。由於此種考驗只關注每個因素的題數是否相同，並未指定不同組別對應的估計參數 (例如因素負荷量) 相同，故常被用來作為基準模式 (baseline model)，用以判斷其他考驗方法是否獲得支持的參照架構。

「量尺恆等性」考驗也被稱為「弱因素恆等性」(weak factor invariance)，主要是探究不同組別受試者在同一份量表上，每個因素所對應的題目之因素負荷量是否相同。假若每個因素對不同組別受試者具相同意義，則因素與題目之間應具相同的因素負荷量。

「純數恆等性」考驗也被稱為「強因素恆等性」(strong factor invariance)，主要是探究不同組別受試者在同一份量表上，每個因素所對應的題目之迴歸分析截距是否相同。假若不同組別由共

同因素預測題目的迴歸分析截距相同，則表示具相同潛在分數的不同組別受試者，其題目得分也會相同。

「因素的共變數恆等性」考驗主要是探究不同組別受試者在同一份量表上，其各個因素之間的共變數是否相同。假若不同組別的受試者具有相同的因素共變數恆等性，則顯示不同組別的受試者，在各因素之間具有相同的關聯性。

檢驗上述四種測量恆等性是否獲得支持的方法，較常採用兩模式的卡方考驗值之差異 ($\Delta\chi^2$)，配合兩模式的自由度之差異 (Δdf)，若顯著性 p 值高於 .05，則顯示測量恆等性獲得實證資料支持。但 Cheung 與 Rensvold (2002) 研究結果顯示，採用兩模式的卡方考驗值差異方法，因卡方考驗易受樣本人數影響，故容易影響此種方法的準確性。Cheung 與 Rensvold 建議採用兩模式的 CFI 值之差異 (ΔCFI)，若兩模式的 ΔCFI 小於 .01，則顯示測量恆等性獲得支持。

本研究編製的後設認知量表，以國中小男女生為適用對象，故將針對男女生，進行「相同型態」、「量尺」、「純數」與「因素的共變數恆等性」等四種測量恆等性考驗，以瞭解「後設認知量表」對不同性別的學生是否具測量恆等性。

進行測量恆等性考驗時，應先確認不同樣本之因素結構模式是否獲得實證資料支持。表 8 為所有受試者與不同男女生的後設認知量表驗證性因素分析整體適配指標，由表 6 可知，除 χ^2 這項指標屬於不良適配情形，其他 CFI、RMSEA、TLI 與 GFI 等四項指標皆屬良好適配，顯示後設認知量表的三個相關因素之因素結構，不論是對全體受試者或男女生組別學生，皆顯示具適配情形。

表 8 12 題後設認知量表驗證性因素分析整體適配指標之適配情形

樣本類型	樣本數	$\chi^2 (p)$	CFI	RMSEA	TLI	GFI
所有受試者	861	163.210 (< .001)	.978	.051	.971	.970
男生	431	148.677 (< .001)	.959	.067	.947	.948
女生	430	84.266 (< .001)	.988	.039	.985	.969
評判標準		$p > .05$	> .90	< .05	> .90	> .90

註：評判標準採用余民寧 (2006) 所提的建議判斷標準

由於表 8 不同性別受試者在後設認知量表上，具相同因素結構個數，故接續以「性別」進行測量恆等性考驗，考驗結果如表 9 所示。由表 9 可知，「量尺恆等性」與「相同型態恆等性」兩模式相比較，兩者的卡方值相差 19.938，自由度相差 9，顯著性 $p = .018$ ，由於卡方考驗顯著性 p 值低於 .05，顯示「量尺恆等性」未獲得實證資料支持。但兩者的 CFI 相差 .002，小於 Cheung 與 Rensvold (2002) 建議低於 .01 差異標準，則顯示「量尺恆等性」獲得實證資料支持。由於卡方考驗容易受到樣本人數的影響，而本研究屬於大樣本，採用 ΔCFI 指標會較合適，故本研究進行不同性別的「量尺恆等性」獲得實證支持。

針對「純數恆等性」與「量尺恆等性」兩模式相比較，兩者卡方值相差 39.406，自由度相差 12，顯著性 $p < .001$ ，顯示「純數恆等性」未獲實證資料支持。但兩者的 CFI 相差 .006，小於 Cheung 與 Rensvold (2002) 建議低於 .01 差異標準，則顯示「純數恆等性」獲實證資料支持。由於卡方考驗容易受到樣本人數的影響，而本研究屬於大樣本，採用 ΔCFI 指標會較合適，故本研究進行不同性別的「純數恆等性」獲得實證支持。

針對「因素的共變數恆等性」與「純數恆等性」兩模式相比較，兩者卡方值相差 7.997，自由度相差 6，顯著性 $p = .238$ ，顯示「因素的共變數恆等性」獲實證資料支持。而兩者的 CFI 相差小於 .001，小於 Cheung 與 Rensvold (2002) 建議低於 .01 差異標準，同樣顯示「因素的共變數恆等性」獲實證資料支持，故本研究進行不同性別的「因素的共變數恆等性」獲得實證支持。

表 9 不同性別受試者在後設認知量表之恆等性考驗結果

模式	模式比較	χ^2 (<i>p</i>)	<i>df</i>	CFI	$\Delta\chi^2$	Δdf	<i>p</i>	ΔCFI
1.相同型態恆等性	-	232.943 (<i><.001</i>)	102	.975	-	-	-	-
2.量尺恆等性	2 比 1	252.881 (<i><.001</i>)	111	.973	19.938	9	.018	.002
3.純數恆等性	3 比 2	292.287 (<i><.001</i>)	123	.967	39.406	12	.000	.006
4.因素的共變數恆等性	4 比 3	300.284 (<i><.001</i>)	129	.967	7.997	6	.238	<i><.001</i>

由於本研究的「量尺恆等性」、「純數恆等性」與「因素的共變數恆等性」皆獲實證資料支持，顯示後設認知量表對不同性別的男女生，不僅同樣具有三個因素個數，且每個共同因素所對應的題目之因素負荷量皆相同，且每個共同因素所對應的題目之迴歸分析截距也是相同的，且每個因素的共變數也相同，代表後設認知量表適用於評量不同性別的男女生。

五、不同性別之潛在平均數差異分析

進行不同組別之潛在變項平均數差異考驗前，得先確定不同組別受試者在該變項具有「相同型態」、「量尺」與「純數」等三種測量恆等性 (Sass, 2011)。根據前面對不同性別在後設認知量表三個因素的測量恆等性考驗，顯示具有「相同型態」、「量尺」與「純數」等三種測量恆等性，故接續進行不同性別在後設認知量表三個分量表的潛在分數之差異分析。男女生潛在平均數的差異考驗統計結果， $\chi^2 = 290.057$ ，*p* < .001、CFI = .967、RMSEA = .041、TLI = .964，顯示具有良好的整體適配情形。男女生在後設認知三個分量表的潛在平均數差異摘要表，如表 10 所示。由表 10 可知，男女生在「後設認知知識」、「計畫與監控」、「評估」等三個分量表的潛在平均數皆未有顯著性差異。

表 10 不同性別國中生在後設認知三個量表潛在平均數差異性考驗摘要表

潛在變項	男生潛在平均數 <i>M</i> (<i>n</i> = 431)	女生潛在平均數 <i>M</i> (<i>n</i> = 430)	<i>p</i>
後設認知知識	0.00	-0.01	.89
計畫與監控	0.00	0.06	.44
評估	0.00	-0.03	.65

註：以男生為參照組，故男生的潛在平均數設定為 0

六、信度分析

經過探索性因素分析與驗證性因素分析後，確定後設認知量表的三個相關因素之因素結構具良好建構效度後，接續進行 α 係數考驗，以評判後設認知量表的信度。將正式施測 861 位受試者填答資料，進行信度分析，統計結果如表 11 所示。

表 11 後設認知量表正式施測 Cronbach 的 α 係數考驗結果 ($N = 861$)

分量表與總量表	題號 (題數)	α 係數	95% CI	
			LL	UL
後設認知知識分量表	1-5 (5 題)	.87	.86	.89
計畫與監控分量表	6-9 (4 題)	.87	.85	.88
評估分量表	10-12 (3 題)	.72	.69	.75
後設認知總量表	1-12 (12 題)	.90	.89	.91

由表 11 可以看出正式施測樣本的信度分析結果，不論總量表或各分量表的 α 係數，皆高於 Nunnally 與 Bernstein (1994) 建議的 .70，顯示後設認知量表具有良好信度。

結論與建議

一、結論

(一) 後設認知量表的因素結構

當前學者專家對後設認知內涵的見解，大多主張後設認知包含後設認知知識與後設認知技能兩部分的內涵。對後設認知知識的內涵，Flavell (1987) 主張可細分成個人變項、工作變項、策略變項等三項，而 Schraw 與 Dennison (1994) 則主張包含陳述性知識、程序性知識、條件性知識等三種。至於後設認知技能的內涵，則大多數學者主張包含計畫、監控、評估等三類。本研究採用 Larkin (2010) 主張，認為後設認知知識的條件性知識，較偏屬於後設認知技能的監控與控制，而「監控」分層面比「條件性知識」分層面，更能展現後設監控的部分，故本研究所編製的「後設認知量表」，未納入「條件性知識」分層表，而包含「陳述性知識」、「程序性知識」、「計畫」、「監控」、「評估」等五個分量表。

現有的後設認知量表 (Panaoura & Philippou, 2003; Schraw & Dennison, 1994; Sperling et al., 2002)，常出現有些題目同時橫跨兩個因素，如此容易產生降低聚斂效度的問題 (Hair et al., 2010)。未避免如此的情況產生，本研究進行的探索性因素分析，將橫跨兩個因素的題目刪除。本研究透過主軸因子配合直接斜交轉軸的探索性因素，萃取出「後設認知知識」、「計畫與監控」、「評估」等三個因素，共 13 道題目。其中，「陳述性知識」與「程序性知識」兩分量表合併為「後設認知知識」一個因素，而「計畫」與「監控」兩分量表合併為「計畫與監控」一個因素。此統計結果，和 Schraw 與 Dennison (1994) 將陳述性知識、程序性知識、條件性知識等三個分量表，合併為後設認知知識，而將計畫、訊息處理策略、監控、除錯策略、學習評估等五個分量，合併成後設認知技能的統計結果相似。

除了透過探索性因素分析外，本研究更進一步採用驗證性因素分析，以檢證由探索性因素分析所獲得的因素結構，是否能獲得其它樣本資料的支持。驗證性因素分析結果，整體適配指標： $\chi^2 = 79.306$ ， $p = .007$ ；CFI = .980；RMSEA = .047；TLI = .975；GFI = .953，顯示 12 題三個相關因素的因素結構，獲得實證資料的支持，顯示本研究編製後設認知量表具有適切的建構效度。

雖然大多數學者專家認同後設認知理論的後設認知知識與後設認知技能兩種內涵，皆可再細分成幾項 (例如後設認知知識又細分成陳述性知識、程序性知識、條件性知識；後設認知技能又細分成計畫、監控、評估等三項)。但根據這些細項所編製的後設認知量表 (例如六個分量表)，在透過探索性或驗證性因素分析後，卻常無法獲得相對應的因素個數 (六個因素)，而是較常獲得後設認知知識與後設認知技能兩個因素 (Panaoura & Philippou, 2003; Schraw & Dennison, 1994; Sperling et al., 2002)。此結果值得再進一步探究，何以根據理論所細分的六個分量表，透過因素分析統計方法，卻常只獲得兩個因素。

(二) 後設認知量表的信度考驗

本研究的後設認知量表，分別針對預試樣本與正式樣本，進行 Cronbach 的 α 係數考驗。「後設認知知識」、「計畫與監控」、「評估」等三分量表，在 200 位預試樣本的 α 係數，分別是 .87、.87、.69；在 861 位正式樣本的 α 係數，分別為 .87、.87、與 .72。由本研究預試樣本與正式樣本的信度考驗結果可知，本研究編製的後設認知量表具良好信度。此結果和林清山與張景媛（1993）、涂金堂（1996）、Schraw 與 Dennison（1994）、Panaoura 與 Philippou（2003）等人的研究結果相似，都顯示現有後設認知量表的 α 係數，大多符合 Nunnally 與 Bernstein（1994）建議信度應高於 .70 的標準。

(三) 後設認知量表的測量恆等性

一份量表的編製，常假定同時適用不同群體（例如男女生），但不同群體在該量表的因素結構是否相同，則須藉由測量恆等性考驗，才能加以確認。故測量恆等性的考驗，是目前量表編製者常需檢定的重要統計方法（Gregorich, 2006; Schmitt & Kuljanin, 2008），然而現成的後設認知量表，則大多缺乏此項考驗。

本研究針對不同中小學男女生，採用多組別驗證性因素分析，以檢視不同群體的因素結構是否相同。針對不同性別的測量恆等性考驗，研究結果顯示，在「量尺恆等性」、「純數恆等性」與「因素的共變數恆等性」這三項考驗，皆具有測量恆等性，顯示男女生在後設認知量表上的得分，具有相同的因素個數、相同的因素負荷量、相同的題目平均數，以及相同的因素共變數，顯示本研究編製的後設認知量表，對中小學的男女生，具有同樣的測量結果。

(四) 男女生後設認知發展情形

本研究針對中小學男女生的後設認知發展情形，進行潛在平均數差異考驗。統計結果顯示，男女生在後設認知量表三個分量表的潛在分數沒有顯著性差異。此結果與陳李綢（1992）的研究結果相似，陳李綢以 200 名國小五年級男女生為研究對象，研究結果發現男女生在後設認知能力的三個分量表（自我預測、自我評估、解題能力）上，皆無顯著性差異。但與 Leutwyler（2009）、O'Neil 與 Brown（1998）、涂金堂（1996）等人的研究結果相異，本研究認為男女生後設認知的差異考驗，可能因不同研究者發展的後設認知量表之分量表內涵不同，而導致不同的研究結果。

二、建議

(一) 採領域特定性編製適合國中小各學科領域的後設認知量表

現有後設認知量表，有些是採領域普遍方式編製，此種編製方式假定該量表適用於探究各學科領域的後設認知。然而 Veenman 與 Spaans（2005），以及 Kelemen 等人（2000）的研究結果顯示，後設認知應屬領域特定性，若採領域特定方式編製後設認知量表，將更能精準測量受試者的後設認知。本研究即採領域特定性的數學領域，編製適合測量中小學生數學後設認知量表。建議未來的研究，可繼續編製適合不同學習領域（語文領域、自然科學、社會科學、藝術領域）的後設認知量表，以協助各級學校教師透過後設認知量表的施測，以瞭解學生在各學習領域的後設認知能力發展情形。

(二) 呈現更多向度的建構效度證據

量表的效度類型大致可分成內容效度、效標關聯效度與建構效度等三類。本研究編製的後設認知量表，除了聘請專家審題，確保適切內容效度，同時採用探索性因素分析、驗證性因素分析、測量恆等等統計方式，獲得良好建構效度。由於國內缺乏適合同時評量中小學生數學領域的後設認知量表，故本研究並未進行效標關聯效度的考驗。建議未來若也是編製適合中小學生數學領域的後設認知量表之研究，可以考慮以本研究所發展的後設認知量表，進行效標關聯效度的考驗，以呈現更多元化的效度證據。

(三) 探討後設認知與學習相關因素的關聯性

學者對後設認知的研究結果顯示後設認知是影響學習成果的重要因素 (Wang et al., 1990)，本研究著重編製具信效度的後設認知量表，並未探究後設認知與學習相關因素關聯性。建議未來對後設認知的研究，可考慮探究後設認知與學習取向、學習態度、學習技巧、知識信念、學習信心、學業成就...等學習相關因素的關聯性。

(四) 探究不同年級受試者在後設認知量表得分的測量恆等性

一份量表對不同群體受試者，是否出現不同的測量模式，是量表編製很重要的測量恆等性考驗。本研究只針對不同性別的受試者，進行後設認知量表的測量恆等性考驗，並未對不同年級的群體進行測量恆等性考驗。建議未來的量表編製者，可針對不同年級的受試者，進行測量恆等性的考驗。

參考文獻

- 于富雲、陳玉欣 (2007)：不同知識表徵建構的學習策略對自然科學習成效之影響。*科學教育學刊*，**15**，99-118。[Yu, F. Y., & Chen, Y. S. (2007). Effects of learning strategies of different knowledge representation on science learning. *Chinese Journal of Science Education*, 15, 99-118.]
- 余民寧 (2006)：潛變項模式：SIMPLIS 的應用。台北：高等教育。[Yu, M. N. (2006). *Latent variable models: The application of SIMPLIS*. Taipei, Taiwan: Higher Education.]
- 涂金堂 (1996)：國小學生後設認知與數學解題表現之相關研究。*國教學報*，**8**，133-164。[Tu, C. T. (1996). A study on metacognition and mathematical problem solving of the elementary school students. *Journal of elementary education*, 8, 133-164.]
- 涂金堂 (2012)：量表編製與 SPSS。台北：五南。[Tu, C. T. (2012). *Scale development with SPSS*. Taipei, Taiwan: Wu Nan.]
- 林清山、張景媛 (1993)：國中生後設認知、動機信念與數學解題策略之關係研究。*教育心理學報*，**26**，53-74。[Lin, C. S., & Chang, C. Y. (1993). A study of the relationship between junior high school students' metacognition, and mathematics problem-solving strategies. *Bulletin of Educational Psychology*, 26, 53-74]
- 陳李綱 (1992)：國小男女生後設認知能力與數學作業表現的關係研究。*教育心理學報*，**25**，67-90。[Chen, L. C. (1992). The relationship between elementary student metacognition and student mathematics performance. *Bulletin of Educational Psychology*, 25, 67-90]
- 蘇宜芬、林清山 (1992)：後設認知訓練課程對國小低閱讀能力學生的閱讀理解能力與後設認知能力之影響。*教育心理學報*，**25**，245-267。[Su, Y. F., & Lin, C. S. (1992). *Bulletin of Educational Psychology*, 25, 245-267.]
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74-94.

- Bathgate, M. E., Sims-Knight, J. E., & Schunn, C. D. (2012). Thoughts on thinking: Engaging novice music students in metacognition. *Applied Cognitive Psychology, 26*, 403-409.
- Benton, C. W. (2013). Promoting metacognition in music classes. *Music Educators Journal, 100*, 52-59.
- Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology, 8*, 583-595.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.
- Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 94*, 131-142.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.
- Desoete, A., & Veenman, M. (2006). Metacognition in mathematics: Critical issues in nature, theory, assessment and treatment. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics education* (pp. 1-10). New York, NY: Nova Science.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Lawrence Erlbaum.
- Flavell, J. H. (1987). Speculation about the nature and development of metacognition. In F. E. Weiner & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Erlbaum.
- Garofalo, J., & Lester, F. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*, 163-76.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*, 78-94.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis: A global perspective*. Saddle River, NJ: Pearson Prentice Hall.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*, 6-23.
- Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning, 7*, 133-149.

- Kaberman, Z., & Dori, Y. J. (2009). Metacognition in chemical education: Question posing in the case-based computerized learning environment. *Instruction Science, 37*, 403-436.
- Kapa, E. (2001). A metacognitive support during the process of problem solving in a computerized environment. *Educational Studies in Mathematics, 47*, 317-336.
- Kelemen, W. L., Frost, P. J., & Weaver, C. A. I. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*, 92-107.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.
- Kramarski, B. (2004). Making sense of graphs: Does metacognitive instruction make a difference on students' mathematical conceptions and alternative conceptions? *Learning and Instruction, 14*, 593-619.
- Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*, 281-310.
- Larkin, S. (2010). *Metacognition in young children*. New York, NY: Routledge.
- Lester, F. K. (1982). Building bridges between psychological and mathematics education research on problem solving. In F. K. Lester & J. Garofalo (Eds.), *Mathematical problem solving: Issues in research*. Philadelphia, PA: The Franklin Press.
- Lester, P., & Bishop, L. K. (2000). *Handbook of tests and measurement in education and the social sciences*. Lanham, MD: The Scarecrow Press.
- Leutwyler, B. (2009). Metacognitive learning strategies: Differential development patterns in high school. *Metacognition and Learning, 4*, 111-123.
- Ministry of Education, Singapore. (2000). *Curriculum planning & development division, Ministry of Education 2001-Mathematics syllabus (primary)*. Singapore, Singapore: Ministry of Education.
- Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*, 249-259.
- Mok, Y. F., Fan, R. M., & Pang, N. S. (2007). Developmental patterns of school students' motivational and cognitive-metacognitive competencies. *Educational Studies, 33*, 81-89.
- Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multi method analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology, 77*, 177-195.
- Nimon, K., & Reio, T., J. (2011). Measurement invariance: A foundational principle for quantitative theory building. *Human Resource Development Review, 10*, 198-214.

- Noar, S. M. (2003). The role of structural equation modeling in scale development. *Structural Equation Modeling, 10*, 622-647.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers, 32*, 396-402.
- O'Neil, H. F., & Brown, R. S. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied Measurement in Education, 11*, 331-351.
- Panaoura, A., & Philippou, G. (2003). The construct validity of an inventory for the measurement of young pupils' metacognitive abilities in mathematics. In N. A. Pateman, B. J. Doherty, & J. T. Zilliox (Eds.), *Proceedings of the 27th conference of the international group for the psychology of mathematics education* (Vol. 3, pp. 437-444), Honolulu, HI.
- Pennequin, V., Sorel, O., Nanty, I., & Fontaine, R. (2010). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking and Reasoning, 16*, 198-220.
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis*. Thousand Oaks, CA: Sage.
- Poitras, E., Lajoie, S., & Hong, Y. J. (2012). The design of technology-rich learning environments as metacognitive tools in history education. *Instructional Science, 40*, 1033-1061.
- Pressley, M., & Gaskins, I. W. (2006). Metacognitively competent reading comprehension is constructively responsive reading: How can such reading be developed in students? *Metacognition and Learning, 1*, 99-113.
- Rickey, D., & Stacy, A. M. (2000). The role of metacognition in learning chemistry. *Journal of Chemical Education, 77*, 915-919.
- Rosenzweig, C., Krawec, J., & Montague, M. (2011). Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: Think-aloud analysis. *Journal of Learning Disabilities, 44*, 508-20.
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*, 347-363.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and limitations. *Human Resource Management Review, 18*, 210-222.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*, 351-371.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.
- Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 1074-1083.
- Silver, E. A. (1985). Research on teaching mathematical problem solving: Some underrepresented themes and needed directions. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Spector, P. (1992). *Summated rating scale construction*. Thousand Oaks, CA: Sage.
- Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*, 51-79.
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality and Quantity, 43*, 599-616.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*, 159-176.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3-14.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research, 84*, 30-43.
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to SE and strategy use. *Journal of Educational Psychology, 82*, 51-59.

收稿日期：2014年09月05日

一稿修訂日期：2014年11月21日

接受刊登日期：2014年12月04日

Bulletin of Educational Psychology, 2015, 47(1), 109-131
National Taiwan Normal University, Taipei, Taiwan, R.O.C.

An Evaluation of the Reliability and Validity of the Mathematical Metacognition Scale

Chin-Tang Tu

Center for Teacher Education

National Kaohsiung Normal University

The main purpose of this study was to develop the Mathematical Metacognition Scale (MMS) on 5th – 8th graders in elementary and junior high schools. The present study included two phases: pretest and test. By adopting purposeful sampling and cluster sampling in both phases, 457 and 861 students were selected for the pretest and the test, respectively. Exploratory factor analysis and confirmatory factor analysis were conducted with the pretest data. The confirmatory factor analysis resulted in three factors including metacognitive knowledge, planning and monitoring, and evaluation. The Cronbach's alphas of the MMS subscales were .87, .87, .72, and .90 for the total score, which indicated good reliability. Measurement invariance was conducted to the test and indicated gender invariance. Therefore, it showed that the MMS has good construct validity.

**KEY WORDS: confirmatory factor analysis, exploratory factor analysis,
mathematical metacognition, measurement invariance**