

國小單步驟加減法文字題解題檢核評量之編製發展與實測分析研究*

許家驊

國立嘉義大學
教育學系

本研究旨在透過成份及內容分析編製發展單步驟加減法文字題檢核評量，檢核能力成份可分成反思及錯誤偵測二向度，反思向度由檢查行動及檢核技巧方式分層面組成，錯誤偵測向度由偵測答案計算、運算程序、題目語意脈絡、數量訊息擷取錯誤、正確無誤分層面組成，單步驟加減法文字題學習內容則取自國小各版本低年級數學教科書。之後研究者使用成份及內容分析解析單步驟加減法文字題解題中涉及之檢核能力運作與題型語意基模，接著使用所有分析結果編製發展單步驟加減法文字題檢核評量。考量不同學校學習內容及經驗同質性，採合目標與叢集取樣抽取嘉義縣市二年級學生 864 人，搭配解題經驗前導誘發評量實施單步驟加減法文字題檢核評量後，再隨機分成兩群進行信效度交叉驗證分析。試探性及驗證性因素分析發現前述二向度各分層面因子結構，且因素萃取、抽取變異比及模式適配度良好，並具良好效標關聯效度（以學生課堂檢核表現教師評定分數為效標）。 α 係數及組合信度、不同向度單題、分層面及分向度極端組 t 考驗亦均良好。而不同解題表現學生之解題檢核（反思及錯誤偵測）實測表現具顯著差異。

關鍵詞：文字題解題、單步驟文字題、學習評量、檢核

* 1. 本篇論文通訊方式：許家驊，通訊方式：jhs@mail.ncyu.edu.tw。
2. 作者非常感謝所有編審委員及編輯團隊為使本文更臻完善所付出之辛勞，謹此誌謝。

依據諸多學者看法，解題檢核 (checking) 是解題歷程及成功表現不可或缺的能力之一，例如 Abbas (2016)、Van Boxtel (2016)、Peltier 與 Vannest (2018) 認為解題檢核能力能提升解題回顧反思 (reviewing and reflection) 的效率，Morin、Watson、Hester 與 Raver (2017)、Peltier 與 Vannest (2016a) 認為解題檢核能力能有助於發現解題錯誤、從而增進自我調整及成功表現的機率，Jitendra 等人 (2013)、Peltier 與 Vannest (2016b) 發現解題檢核能力越強的個體，其解題表現越佳。

自前述可知解題檢核能力的重要性，特別是國小低年級學生，解題檢核能力往往是影響其解題表現之重要因素之一 (Abbas, 2016)，但其內涵意義及如何有效進行評估，卻少有學者探討，因此本研究希望探討解題歷程及檢核能力內涵，並配合單步驟 (single-step) 解題學習內容分析，據之設計編製適用於國小低年級學生單步驟解題檢核之品質良好及有效評估工具。其中涉及解題歷程與檢核能力內涵、解題學習內容、評估方法三部份說明，最後再作歸納。

一、解題歷程與檢核能力內涵

依 Kong 與 Orosco (2016)、Peltier 與 Vannest (2016b)、Peltier 與 Vannest (2018) 的看法，文字題解題較一般計算題為難，原因在於文字題解題需經一連串解題心理表徵及運作執行歷程始能完成，而非僅是數量及運算規則 (algorithm) 之執行運用。次依 Özcan (2016)、Peltier 與 Vannest (2016a)、Peltier 與 Vannest (2018) 的觀點，文字題解題歷程包含讀題 (reading)、題意理解表徵 (comprehension and representation)、計畫及方法策略選擇 (planning and strategy selection)、執行解題 (execution)、解題回顧，其中解題回顧階段即涉及解題檢核之運作。

再就運作能力性質而言，Kong 與 Orosco (2016)、Morin 等人 (2017)、Özcan (2016)、Peltier 與 Vannest (2018) 認為前述解題歷程，前三個階段以認知 (cognition) 能力為主、後設認知 (metacognition) 能力為輔，但最後階段則以後設認知能力為主、認知能力為輔。此不僅符合解題歷程四種理論模式中之交互運作聯合模式本質 (將認知與後設認知區分但一起說明解題歷程)，且與解題策略教學研究發現同時進行二種解題能力策略教學，更能有效提升個體解題表現之結果吻合 (Kong & Orosco, 2016; Morin et al., 2017; Özcan, 2016; Peltier & Vannest, 2016a; Peltier & Vannest, 2016b; Peltier & Vannest, 2018)。

因此可發現解題檢核與後設認知能力間具有異同關係，依 Kong 與 Orosco (2016)、Morin 等人 (2017)、Özcan (2016) 認為後設認知可分知識 (knowledge)、執行控管 (executive control)、信念 (belief) 三層面觀點，解題檢核能力同屬後設認知執行控管能力之一，但後設認知知識及信念仍可能影響其運作。

歸納學者觀點，解題檢核能力至少可分為二個向度，一為 Abbas (2016)、Jitendra 等人 (2013)、Morin 等人 (2017) 所提反思回顧檢查運作能力 (簡稱反思向度)，二為 Bunt、Terry 與 Lank (2013)、Pratt 與 Urbanowski (2016) 所提錯誤覺察 (error awareness) 發現及偵測 (detection) 能力 (簡稱錯誤偵測向度)，前者與是否具有檢查行動、如何檢查有關，例如在解題後，個體是否會主動再次檢查解題結果？檢查那些地方，是答案或運算式或題目語意？其順序及組合為何？後者則與是否能偵測出各類解題時可能出現的錯誤有關，例如是否能偵測出答案或運算式或題目語意或數量對應安排有錯誤，或是均無問題？

歸納上述向度及內涵，反思向度可分為檢查行動及方式，前者評估重點在於了解個體有無解題驗算檢查之運用習慣、而後者則在於了解個體單一或多元解題驗算檢查方法之運用能力 (Peltier & Vannest, 2016a; Peltier & Vannest, 2016b; Van Boxtel, 2016)。

錯誤偵測向度可分為偵測答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤、數量訊息擷取錯誤、正確無誤，各者評估重點分別在於評估答案及運算程序正確性、題目脈絡訊息的合理性、題目與算式數量訊息的一致性、作答心向及一般解題評量經驗應用性 (Yin et al., 2016; Zhang, 2017)。

上述均可搭配接續介紹之國小低年級學生各類型單步驟解題學習內容進行實務設計。

二、國小低年級學生單步驟解題學習內容(九年一貫 97 數學領域微調課綱第一學習階段)

綜合吳昭容(1990)、Furnell 與 Speer(2013)、Morin 等人(2017)、Peltier 與 Vannest(2016a)、Peltier 與 Vannest(2016b)、Peltier 與 Vannest(2018)、Powell(2011)、Verschaffel、Greer 與 DeCorte(2007)加減法數學文字題語意基模(semantic schema)觀點,分析各主要版本(康軒、南一、翰林)出版之一至二年級數學課本及教學手冊第一至四冊相關單元內容,之後再針對所涉運算量數訊息進行解析,可得幾類型單步驟加減法解題學習活動。

在康軒版楊瑞智主編(2016,2017)、南一版李源順主編(2016,2017)、翰林版林長壽主編(2016,2017)之一二年級上下學期單步驟文字題共同題型基模方面,共含合併型—求整體量,改變型結果量未知—增加及減少,比較型—比多三型四類。在共同運算量數類型方面,共含個位數對個位數進位加、不進位加、不借位減法,二位數對個位數不進位加、借位減法,二位數對二位數不借位減法四類,涉及量數大小除翰林版在 70 以內,餘康軒、南一版均在 50 以內。其題型及示例如表 1 所示:

表 1 國小低年級各主要版本單步驟加減法文字題題型示例

1.合併型—求整體量:小吃店上午賣了 4 碗湯,下午賣了 8 碗湯,一共賣出幾碗湯?
2.改變型—結果量未知—增加:大明有 20 元,媽媽又給他 4 元,現在大明有多少元?
3.改變型—結果量未知—減少:大明有 16 元,他買了一個 4 元的書套,現在大明有多少元?
4.比較型—比多:妹妹有 5 個芭樂,弟弟有 9 個芭樂,弟弟比妹妹多幾個芭樂?

三、解題檢核評估方法

依前述學者所提解題檢核能力二個向度及內涵觀之,反思向度能力可使用反思性(reflective)問題編製評量進行評估(Abbas, 2016; Jitendra et al., 2013; Morin et al., 2017),錯誤偵測向度能力則可借用錯誤偵測(error detection)設計編製評量進行評估(Bunt et al., 2013; Pratt & Urbanowski, 2016)。

具體言之,亦即可參照前述表 1 題型內容,對應 Peltier 與 Vannest(2016a)、Peltier 與 Vannest(2016b)、Van Boxtel(2016)反思向度內涵編製解題檢查(含驗算)行動與方式(action and method)之反思性問題評量,前者試探個體於解題後是否會再檢查或驗算,後者試探個體如何進行檢查或驗算,二者之重點均在於題目、算法、答案之檢查或驗算。

對應 Yin 等人(2016)、Zhang(2017)錯誤偵測向度內涵編製計算答案(answer)錯誤、運算程序(operation)錯誤、題目語意脈絡(semantic context)錯誤、數量訊息擷取(quantity displace)錯誤、正確無誤(correct and no error)之錯誤偵測評量,前者可搭配改變型—減少—減法問題,次者可搭配改變型—增加—加法問題,三者可搭配比較型—比多—減法問題,四者可搭配合併型—整體—加法問題,最後可搭配合併型—整體—加法問題。

同時綜合上述二向度評量評估個體解題檢核能力,其設計概念與內容詳如後續表 2 及表 3 所示。

四、研究目的與問題整合歸納

整體而言，可發現數學文字題解題檢核能力對國小低年級學生解題表現具有重要關鍵影響，但缺乏適用評估工具設計，因此聚焦於文字題解題特定領域之國小低年級學生解題檢核能力評估工具仍待開發。依教育部（2008）九年一貫 97 數學領域微調課綱，第一學習階段低年級代數部份「A-1-03 能理解加減互逆，並運用於驗算與解題。」能力指標編製之各版本課本中，有關驗算檢查教學單元均安排在二年級上下學期，考量驗算檢查即為解題檢核能力之一，故本研究將針對國小二年級學生進行。

故本研究希冀參照數學文字題解題檢核（簡稱解題檢核）能力內涵進行歸納分析，以前述表 1 及後續表 2、3 為編製架構，使用學習評量形式設計，據之編製良好適用之國小二年級學生單步驟加減法文字題解題檢核評量，並搭配單步驟加減法文字題解題評量實施，再針對不同解題表現個體之解題檢核實測表現進行分析，以了解其應用功能。

但由於一般試探問題大致採無當下實作狀況下，請個體自行回溯過去並自陳報告（self-report）的方式實施，考量國小低年級學生可能較無此種答題經驗起見，故採先實施單步驟文字題解題評量，再搭配實施檢核評量的方式進行，一方面可喚起（activate）其過去已學單步驟文字題題型之解題經驗（解題經驗前導誘發），另一方面亦可蒐集其單步驟解題表現（解題先備能力蒐集）。依前述目的，本研究具體問題可細列如下：

- （一）單步驟加減法文字題解題檢核評量在解題檢核表現測量上之效度為何？
- （二）單步驟加減法文字題解題檢核評量在解題檢核表現測量上之信度為何？
- （三）單步驟加減法文字題解題檢核評量在解題檢核表現測量上之鑑別力為何？
- （四）不同解題表現個體之單步驟加減法文字題解題檢核評量實測結果為何？

研究方法

一、研究對象

因九年一貫課程採一綱多本方式實施，為便於統一教學版本、協調單元教學進度、顧慮研究對象在校解題學習經驗同質性，故採合目標與叢集取樣（purposeful and cluster sampling）於使用適當版本之嘉義縣市國小普通班二年級學生中，以學校為叢集單位選取所需樣本。而依 Crocker 與 Aligina（1986）所提「每項目至少需五人」之標準計算，解題檢核評量預定編製三十個計分項目（反思向度 10 題、錯誤偵測向度 20 題），每項目乘上五人，至少約需 150 人以上。最後實際取樣人數，嘉義縣抽取五校二十五班（康軒版三校十五班 273 人、翰林及南一版共二校十班 192 人）合計 465 人，嘉義市抽取四校十八班（康軒版二校九班 201 人、翰林及南一版共二校九班 198 人）合計 399 人，總計 864 人（因無遺漏值，故均為可用有效樣本），符合前述項目編製分析之人數標準。

為利後續部分品質之交叉驗證（cross validation），故將前述各校內班級均以班為單位隨機分派為兩群後（432 人及 432 人），並針對此二校際樣本群之單步驟加減法文字題解題先備能力表現（使用評量如後續研究工具三所列）進行 t 考驗，以確認其同質性，結果發現二者間並無顯著差異（單步驟 $M = 11.975$ 、 11.583 ， $t = .692$ ， $p = .251$ ； $p > .05$ ）。代表此二校際樣本群之單步驟加減法文字題解題先備能力表現符合同質群假定，適合作為後續部分品質指標（信效度）交叉驗證之用，除驗證性因素及組合信度分析以第二校際樣本群（ $N_2 = 432$ ）為之，其餘分析均以第一校際樣本群（ $N_1 = 432$ ）為標的群體進行。

二、研究架構

將分兩階段進行，一為評量發展、二為功能驗證，前者將歷經整合解題檢核內涵與實徵研究之文獻探析、解題檢核能力分析與學習內容分析、編製評量三個次階段，後者則含評量品質分析、實測分析二個次階段，此外為達解題經驗前導誘發、解題表現及效標蒐集目的，將搭配解題評量實施，其架構及內容如圖 1。

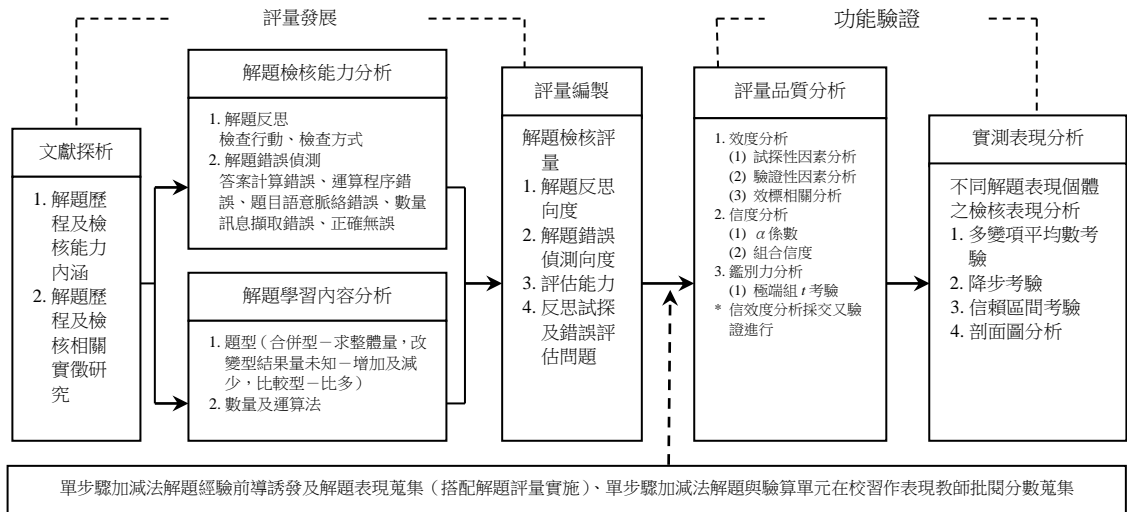


圖 1 單步驟加減法文字題解題檢核評量發展及功能驗證研究設計架構

三、研究工具

以下第一及二項為主要工具，第三項為輔助工具（解題經驗前導誘發及解題表現蒐集）。

(一) 單步驟加減法文字題解題檢核評量（簡稱解題檢核評量）

係整合緒論所提解題檢核之解析向度及主張為編製基礎，再以解題反思及錯誤偵測向度來設計項目內容。反思向度分為檢查行動、檢查方式二分層面編製，錯誤偵測向度分為偵測答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤、數量訊息擷取錯誤、正確無誤五分層面編製，其編製說明分述如後。

1. 反思向度評量

依緒論所提之解題歷程檢核能力分析，編製反思試探問題，其例如表 2 所示。採整合後續第三項解題評量自行解題後再對試探問題自行選答方式進行，之後對二試探問題涉及之檢查行動與方式進行分類給分，有檢查行動、使用任一種表列方式或自述方式者為 1 分，餘為 0 分，共五大題（每大題含解題 1 題、檢查行動及方式各 1 題）。

2. 錯誤偵測向度評量

依緒論所提之解題錯誤偵測類型分析，再搭配緒論所提學習內容範圍（表 1），各題型輪流使用編製錯誤偵測評量，其例如表 3 所示。

表 2 單步驟加減法文字題解題反思向度評量內容示例 (搭配解題評量)

*題型：合併型－整體－加		
請您先自行解題再回答試探問題		
文具店下午賣了 5 個擦子，晚上又賣了 7 個擦子，一共賣出幾個擦子？		
內容	試探問題	選項
檢查行動	() 寫完題目後您會再檢查或驗算嗎？ 請從右邊的答案中選一個答案，把號碼填在括號中	(1) 會 (2) 不會
檢查方式	若您會再檢查或驗算，請再回答下一題，若您不會再檢查或驗算，請不用回答下一題 () 您是怎麼檢查或驗算的？ 請從右邊的答案中選一個答案，把號碼填在括號中	(1) 只有唸唸題目 (2) 只有想想算法對不對？ (3) 只有算算答案對不對？ (4) 再唸唸題目、想想算法對不對？ (5) 再唸唸題目、算算答案對不對？ (6) 再想想算法對不對？然後算算答案看對不對？ (7) 再唸唸題目、想想算法對不對？然後再算算答案看對不對？ (8) 其他 (請寫出您是怎麼做的)

表 3 單步驟加減法文字題解題錯誤偵測向度評量內容示例

錯誤種類	題型	評估問題
下述各題上方均附右方說明		請您仔細的看看這一題的題目、算法及答案，若您覺得都沒問題的話，請在題目最後面打勾，若您覺得有奇怪地方的話，請圈起來。
答案計算錯誤	改變型－減少－減	妹妹有 9 顆柳丁，把 5 顆裝到籃子裡，還剩下幾顆？ $9-5=3$
運算程序錯誤	改變型－增加－加	玉明第一次打球得 5 分，第二次打球得 2 分，兩次共得幾分？ $5-2=3$
題目語意脈絡錯誤	比較型－比多－減	大明投球得 3 分，小玉投球得 5 分，大明比小玉多幾分？ $5-3=2$
數量訊息擷取錯誤	合併型－整體－加	美美第一次拍球拍了 4 下，第二次拍球拍了 1 下，兩次共拍了多少下？ $4+2=6$
正確無誤	合併型－整體－加	小明早上買了 5 個糖果，下午買了 2 個糖果，今天共買了多少個糖果？ $5+2=7$

除具有各類型錯誤外，並加入正確無誤題以避免作答心向，共二十題並經隨機處理，採自行挑選錯誤後分類給分，反應正確 1 分，反應錯誤或無反應 0 分，因與一般評量性質不同，故附有練習評量 (第二項工具)。

二向度評量編修後，採四位數學教育專家審閱修正、解題學習課程本位、解題檢核評量結構基模同構性 (isomorphic) 原則，亦即同時反映解題學習課程共同內容結構、解題檢核之反思及錯誤偵測二向度評量共同內容結構形式來支持內容效度 (結果如實施程序一所述)，預試後進行因素分析作為構念關聯 (construct-related) 或構念效度指標，再以個體於單步驟解題與驗算單元教學後之單步驟解題與驗算單元課堂習作教師批閱分數為效標，進行相關分析作為效標關聯 (criteria-related) 效度指標，並採內部一致性 α 係數及組合 (composite) 信度作為信度指標，以極端組 t 考驗及單題總分相關為鑑別力指標，如後續結果與討論所示。

另考量作答性質，二向度評量將分開實施，且為喚起國小學生解題經驗與回憶（解題經驗前導誘發），故在填答每部分評量前，將分別先讓個體進行錯誤偵測向度評量練習及部分單步驟解題後，再回答各評量項目。

（二）單步驟加減法文字題解題錯誤偵測練習評量與範例教學說明（簡稱練習評量）

因前項工具填答形式不同，故另行設計練習評量（如表 3）及說明，作為正式實施之練習。本項將配合文字及口述說明簡報實施（Microsoft powerpoint 製作，單槍投影機播放），含作答說明、單題注音文字視訊閃示及音訊口述說明（讀題引導及正確答案回饋），僅為了解個體答題狀況之用，結果不予計分。流程如圖 2。

<p>※以下搭配表 3 示例進行練習與範例教學說明</p> <p>一、請檢查算式的答案有沒有問題？【1. 把您的算式算算答案，看是不是跟題目給的算式答案一樣？（答案有沒有錯？）（檢查算式的答案）】【教作至評量正確為止，答案錯誤偵測評量教學終止】</p> <p>二、請檢查算式的算法有沒有問題？【1. 從我們知道的數字和題目要我們做的事，要怎麼來算這一題？ 2. 可不可以列出算式？ 3. 您算式中的算法跟題目給的算式算法一樣嗎？（算式中的算法對不對？）（檢查運算）】【教作至評量正確為止，運算錯誤偵測評量教學終止】</p> <p>三、請檢查題目本身有沒有問題？【1. 從題目中，我們知道的有那些數字？ 2. 這些數字告訴我們什麼？ 3. 這些數字的大小對不對？（檢查量數關係） 4. 題目要我們做什麼？ 5. 題目要我們做的事，可不可以做？（檢查問句）】【教作至評量正確為止，題目語義錯誤偵測評量教學終止】</p> <p>四、請檢查算式中的數字有沒有問題？1. 您算式中的數字跟題目給的算式數字一樣嗎？（算式中的數字對不對？）（檢查數量擷取）】【教作至評量正確為止，數量訊息擷取錯誤偵測評量教學終止】</p> <p>五、都沒有問題嗎？【1. 有問題則重新教學 2. 無問題進行下一題】</p>
--

圖 2 解題錯誤偵測向度評量範例教學及說明流程

（三）單步驟加減法文字題解題評量（簡稱解題評量）

依緒論所提學習內容範圍，分析歸納所涉文字題相關題型與計算難度而編製。題型內容同第一項工具共五題且各型輪流使用並經隨機處理。其編製內容細目、類型示例如表 1 所示，採整合前述第一項解題反思評量自行解題進行，可對解題運算程序及答案部分給分。

採第一項工具所述方式支持內容效度（結果如實施程序一所述）。經以使用前述各版課本之 112 位非取樣學校二年級學生預試，試探性因素分析，因素結構與編製架構相符，斜交轉軸特徵值均大於 1，估計解釋變異比均具相當比例（共 71.85%）。另外具有良好效標關聯效度，其係數為 .877（以個體於單步驟解題與驗算單元教學後之單步驟解題與驗算單元課堂習作教師批閱分數為效標）。單題及整體決斷值 t 考驗均達 0.05 顯著水準，顯示具良好鑑別力。內部一致性 α 係數 .896，顯示具良好信度。

四、實施程序

依 Millman 與 Greene（1993）所提前置預試（preliminary tryout）及測試項目組合即施行工具（experimental items as the operational instrument）程序進行，以下試作及正式施測即屬前述程序，於學校單步驟解題與驗算單元教學後進行。

（一）解題及檢核評量編製、施測準備

解題評量部分以表 1 內容為編製架構，選擇題型編擬問題，檢核評量部分以表 1 至表 3 內容為編製架構，編擬評量項目。再由四位數學教育專家（兩位大學教授、兩位小學數學老師）進行審閱。接著綜合審視結果，進行解題評量部分「計算類型由二位對二位不借位減法調為二位對一位借位減法，加大題與題間之作答空格」、檢核評量（錯誤偵測向度）部分「題目文字敘述簡化口語化，計算數量勿過大過難，至多以二位對一位不進位加法及不借位減法為限」之修正調整及編輯。再依正式錯誤偵測評量內容編製練習評量，並針對練習及施測說明製作標準化視訊與音訊錄音簡報，之後依填答程序，設計說明指導語。最後組合解題及檢核評量並進行注音標示。

（二）解題及檢核評量試作

於非取樣學校中（符合樣本群使用版本），挑選三十位二年級學生先進行一次約 40 分鐘錯誤偵測評量範例練習，請受測者隨施測者讀誦一遍說明語，再由施測者輔助受測者練習填答錯誤偵測範例（如表 3 題型及圖 2 範例教學說明流程），之後於不同時間分五次交互實施解題及檢核評量試作，每次程序說明如後。

1. 共同填答說明

先進行約 5 分鐘填答說明，由施測者對受測者精簡口頭說明，再由受測者提問、施測者回答。

2. 進行解題及檢核評量填答

請試作者分別對部分單步驟解題評量、反思及錯誤偵測向度檢核評量填答約 35 分鐘。為喚起個體解題經驗與回憶（解題經驗前導誘發），同時蒐集其解題表現，故在填答每部分檢核評量前一時段，先進行部分單步驟解題評量，後一時段再填答檢核評量，每次之單數段為解題及反思整合評量、偶數段為錯誤偵測評量，五次後，解題及檢核評量均實施完畢（每次解題及反思整合 1 題、錯誤偵測 4 題）。

另採分成 6 小組（每組 5 人，1 組 1 位施測者）同時試作，以利施測者詢問受測者填答理解狀況及相關問題，並判斷其回應與表 1 至表 3 編製細目界定內容之一致性（正式實施無此程序）。結果發現受測者均能理解評量內容，並能於預定時間內回答完畢，且無相關問題，而互動回應內容亦與編製界定內容方向一致，狀況良好，無須再行調整。

（三）解題及檢核評量正式施測

依前述程序對所有研究對象，先進行一次錯誤偵測評量範例練習，後於不同時間分五次交互實施解題及檢核評量正式施測，五次後，解題及檢核評量均實施完畢。因研究對象較多且各校行事安排不同，正式施測乃委託熟悉班級學生狀況之班級導師於班級晨光及導師時間協助施測，且因各校升旗開會日之行事安排均不同，故採同期間（儘可能於同一週）但非同時間的方式施行。此外於正式施測期程開始前，研究者及助理會先至各校與協助導師座談及說明實施注意事項（施測袋上亦附有此注意事項說明文件），並隨時提供諮詢。

（四）評量功能驗證

整合資料進行品質（含信效度交叉驗證）及實測結果分析。

五、資料處理與分析

針對研究問題一，進行試探性及驗證性因素分析（exploratory and confirmatory factor analysis, EFA and CFA）、效標相關分析。針對研究問題二，進行傳統信度及組合信度（composite reliability）分析。針對研究問題三，進行極端組 t 考驗。針對研究問題四，進行獨立樣本單因子多變項平均數考驗（MANOVA）及 η^2 效果值、Roy-Bargman 降步（step-down） F 考驗及同時信賴區間（simultaneous confidence interval）考驗。以上除驗證性因素分析係以 Amos 進行外，餘均採 SPSS for Windows 為之。

結果與討論

一、構念效度分析

(一) 解題檢核評量各向度項目一階試探性因素分析

本研究雖有理論基礎及文獻依據，但為利於後續驗證性因素分析之進行，故先採試探性因素分析，探索資料組型，同時了解其與理論架構間之關係。依林清山（1991）、Hair、Black、Babin 與 Anderson（2010）所言，採主軸因素（principal axis factor, PAF）搭配斜交轉軸（oblimin rotation）分向度進行試探性共同因素分析。所有分析內容詳如表 4 及 5。

表 4 解題檢核評量反思向度項目斜交轉軸試探性因素分析 (N1 = 432)

項目	組型矩陣		結構矩陣	
	因子 1	因子 2	因子 1	因子 2
1	.723	-.055	.695	.316
2	.786	.015	.793	.418
3	.794	.083	.837	.491
4	.741	.122	.803	.502
5	.526	-.196	.425	.074
6	-.027	.632	.297	.618
7	.020	.495	.274	.506
8	-.080	.634	.245	.593
9	-.030	.449	.200	.433
10	-.026	.754	.361	.741
歸屬後分層面	檢查行動	檢查方式	檢查行動	檢查方式

註：方框表示因子歸屬。

上述分析符合林清山（1991）所提因素相關不為零，陳正昌、程炳林、陳新豐與劉子鍵（2012）所提因素相關接近 .2 至 .3 間應採用斜交轉軸之假定。

表 5 解題檢核評量錯誤偵測向度項目斜交轉軸試探性因素分析 (N1 = 432)

項目	組型矩陣					結構矩陣				
	因子 1	因子 2	因子 3	因子 4	因子 5	因子 1	因子 2	因子 3	因子 4	因子 5
1	.826	-.086	.059	.053	-.057	.868	-.285	.429	.002	.029
2	.972	-.040	-.186	.082	.022	.903	-.212	.242	.089	.131
3	.755	-.019	.230	-.109	.116	.871	-.255	.573	-.078	.154
4	.678	.052	.391	.001	-.011	.835	-.171	.680	.005	.082
5	-.122	.482	.200	.189	-.169	-.166	.522	.061	.302	-.006
6	.013	.947	.008	-.137	.005	-.213	.891	-.119	.230	.142
7	-.134	.811	.008	.069	-.025	-.329	.864	-.173	.371	.148
8	.041	.730	.027	.082	.014	-.122	.751	-.064	.368	.194

9	.514	.020	.545	.060	-.105	.738	-.182	.760	.008	-.006
10	.030	.339	.343	.054	.100	-.192	.378	.423	.238	.177
11	.155	.010	.863	-.008	.000	.531	-.157	.930	-.040	.049
12	.088	.119	.899	-.059	.006	.453	-.056	.922	-.049	.052
13	-.124	-.017	.261	.518	.147	.012	.204	.193	.554	.337
14	.245	-.004	-.120	.771	.043	.203	.260	-.044	.793	.351
15	-.082	.097	.018	.613	.003	-.093	.351	-.058	.650	.245
16	.101	.033	-.309	.752	-.131	-.050	.318	-.308	.730	.155
17	-.191	-.013	.260	.360	.363	-.036	.206	.177	.481	.486
18	.248	.307	-.276	-.127	.381	.089	.314	-.192	.148	.407
19	.022	-.086	-.086	.005	.701	.073	.062	-.036	.239	.685
20	-.038	-.051	.051	.070	.852	.080	.147	.074	.368	.866
歸屬後分層面	答案 計算 錯誤 偵測	運算 程序 錯誤 偵測	題目語 意脈絡 錯誤偵 測	數量訊 息擷取 錯誤偵 測	正確 無誤	答案 計算 錯誤 偵測	運算 程序 錯誤 偵測	題目語 意脈絡 錯誤偵 測	數量訊 息擷取 錯誤偵 測	正確 無誤

註：同表 4 註之說明。

二向度前置分析部分，KMO 值依 Kaiser (1974) 規準，因素分析之取樣適切性良好且接近極佳 (.830 及 .868)、屬中等以上範圍，而 Bartlett 近似卡方值相當大且達 .05 顯著水準 (1448.431, $df=45$, $p=.000$, $p<.05$; 6382.321, $df=190$, $p=.000$, $p<.05$)，代表相關矩陣並非單元矩陣，且各相關係數彼此不同並大於 0，適合進行因素分析。

系統分別成功萃取反思向度二個分層面、錯誤偵測向度五個分層面，由於二向度各分層面間具有不為零之中度平均正相關 (.514 及 .197)，因此僅能報告各向度各分層面轉軸後平方總和 (3.303, 2.500; 4.362, 3.337, 3.733, 2.829, 2.107)，以及斜交轉軸程序中產生之未轉軸解釋變異百分比 (33.378, 10.801; 26.585, 17.602, 8.551, 5.804, 4.790)、總累積解釋變異百分比 (44.179; 63.332)。前述平方總和亦相當於使用正交轉軸後所得之因子特徵值 (eigen value)，依 Kaiser (1960) 建議大於 1 的特徵值保留規準、陡坡圖 (scree plot) 及平行分析 (parallel analysis) 趨勢觀之，均具相當影響，故均予以保留。

自表 4 及 5 結構矩陣之方框數據看來，這些加權係數及負荷量均在邱皓政 (2001)、Hair 等人 (2010) 所提標準 .3 以上，代表每項目與因素間均具相當程度關聯，再自其在因素上的數值大小來判斷其因素歸屬，再依表 2 及表 3 理論編製架構及各項內容，進行各項目適合因子歸屬及命名。

結果發現解題反思向度第 1 至 5 項目適合歸屬因素一，均屬檢查行動性質，故命名為檢查行動，第 6 至 10 項目適合歸屬因素二，均屬檢查方式性質，故命名為檢查方式。解題錯誤偵測向度第 1 至 4 項目適合歸屬因素一，均屬答案計算錯誤偵測性質，故命名為答案計算錯誤偵測，第 5 至 8 項目適合歸屬因素二，均屬運算程序錯誤偵測性質，故命名為運算程序錯誤偵測，第 9 至 12 項目適合歸屬因素三，均屬題目語意脈絡錯誤偵測性質，故命名為題目語意脈絡錯誤偵測，第 13 至 16 項目適合歸屬因素四，均屬數量訊息擷取錯誤偵測性質，故命名為數量訊息擷取錯誤偵測，第 17 至 20 項目適合歸屬因素五，均屬正確無誤性質，故命名為正確無誤。如此結果不僅與前述文獻所提之解題反思及錯誤偵測向度歸屬相符，且進一步發現其間隱含之彼此群聚隸屬共同結構及向度。

經前述因素分析後之各描述統計及相關，解題反思二分層面 M 界於 3.799 至 4.042 間， SD 界於 1.497 至 1.656 間， r 值 .471，解題錯誤偵測五分層面 M 界於 2.403 至 3.660 間， SD 界於 .905 至 1.670 間， r 界於 .290 至 .764 間 (平均 .504)，各分層面彼此與各分層面總分間相關界於 .042 至 .951 間 (平均 .430)，91.67% 達 $p<.05$ 及 $p<.001$ ，代表各分層面內及彼此間大部份具良好關聯、內部一致性良好，如表 6、7 所示。表 7 中「正確無誤」與「檢查行動、方式及解題反思」間幾近零相關，其因可能為個體進行正確無誤題型作業與平常未經錯誤設計作業之解題經驗接近，

故未出現強烈覺察需要解題反思檢查之動機信念，與緒論所提解題檢核能力含反思向度運作可能受後設認知信念影響之前提相符。

表 6 解題檢核評量不同向度分層面描述統計 ($N1 = 432$)

統計量	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤	解題錯 誤偵測
<i>M</i>	4.042	3.799	7.840	3.190	2.796	2.403	3.160	3.660	15.208
<i>SD</i>	1.497	1.656	2.937	1.107	1.574	1.670	1.247	.905	5.311

表 7 解題檢核評量不同向度分層面相關 ($N1 = 432$)

	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤
檢查方式	.471**							
解題反思	.951**	.720**						
答案計算錯誤偵測	.371**	.195**	.360**					
運算程序錯誤偵測	.298**	.175**	.295**	.675**				
題目語意脈絡錯誤偵測	.421**	.165**	.389**	.764**	.644**			
數量訊息擷取錯誤偵測	.224**	.186**	.241**	.535**	.566**	.496**		
正確無誤	.042	.071	.058	.291**	.306**	.265**	.497**	
解題錯誤偵測	.378**	.208**	.370**	.875**	.854**	.859**	.760**	.496**

註：**概率值均為 $p = .000$ ， $p < .05$ 。

(二) 解題檢核評量各向度分層面高階試探性因素分析

自表 4 至 5、7 看來，各向度各分層面的共同向度性均為良好，且符合本評量原初之編製理論架構，加上為避免產生 Byrne (2001) 所提虛假 (bogus) 因素，以及陳蜜桃、李新民與黃秀霜 (2007) 所提擬似 (spurious) 相關問題，並降低因素分析參數估計所需樣本數之門檻，以提高分析信效度起見，此試探性因素分析將考量 Bandalos 與 Finney (2001) 「向度性」原則，審慎以同性質之分層面題包 (item parcel) 為單位進行反思及錯誤偵測向度分層面高階試探性因素分析，以了解其是否具有共同高階因子？

由於假定潛在因素間應具有一定程度之相關存在，故以反思及錯誤偵測向度各單題組合之分層面為單位，使用主軸法搭配斜交轉軸進行高階試探性共同因素分析，結果如表 8。

表 8 二向度分層面主軸因素分析及斜交轉軸

向度分層面	組型矩陣		結構矩陣	
	因素 1	因素 2	因素 1	因素 2
檢查行動	.986	-.092	.942	.377
檢查方式	.479	-.011	.474	.217

答案計算錯誤偵測	.097	.789	.473	.835
運算程序錯誤偵測	.012	.786	.386	.792
題目語意脈絡錯誤偵測	.152	.723	.496	.795
數量訊息擷取錯誤偵測	-.081	.753	.277	.714
正確無誤	-.177	.524	.073	.440
歸屬後因素命名	解題反思	解題錯誤偵測	解題反思	解題錯誤偵測

註：同表 4 註之說明。

本項分析 KMO 值依 Kaiser (1974) 規準，代表因素分析之取樣適切性尚屬良好 (.790)，而近似卡方值相當大且達 .000 顯著水準 (1188.932, $df = 21$, $p = .000$, $p < .05$)，代表因素分析使用之相關矩陣並非單元矩陣，且各相關係數彼此不同並大於 0，適合進行因素分析。

系統分別成功萃取反思、錯誤偵測向度二個因子，由於二向度因子間具有不為零之中度平均正相關 (.476)，因此僅能報告各向度因子轉軸後平方總和 (2.849; 1.813)，以及斜交轉軸程序中產生之未轉軸解釋變異百分比 (42.398, 12.363)、總累積解釋變異百分比 (54.762)。前述平方總和亦相當於使用正交轉軸後所得之因子特徵值，依 Kaiser (1960) 建議大於 1 的特徵值保留規準、陡坡圖及平行分析趨勢觀之，均具相當影響，故均予以保留。

自三矩陣方框數據看來，這些加權係數及負荷量均在邱皓政 (2001)、Hair 等人 (2010) 所提標準 .3 以上，代表反思、錯誤偵測向度每個分層面與各向度間均具相當程度之相對重要性及關聯，再自其在各因素的數值大小來判斷其因素歸屬，可發現在反思向度，檢查行動、檢查方式適合歸屬因素一，均屬解題反思能力，故命名為解題反思。在錯誤偵測向度，答案計算錯誤偵測、運算程序錯誤偵測、題目語意脈絡錯誤偵測、數量訊息擷取錯誤偵測、正確無誤適合歸屬因素二，均屬解題錯誤偵測能力，故命名為解題錯誤偵測。而解題反思及錯誤偵測均屬解題檢核能力之一，故命名為解題檢核。

如此結果不僅與前述文獻所提之解題反思及錯誤偵測向度歸屬檢核能力相符，且進一步發現了其間隱含之彼此群聚隸屬共同結構及向度。

(三) 整合試探性因素分析結果之高階驗證性因素分析

前述結果雖已浮現因子及項目間之包含隸屬型態，但因子整體階層組織結構及徑路關係並不明朗，為深入了解其合理與可靠性，將參照李茂能 (2006) 建議，針對第二樣本群 ($N2 = 432$) 以表 4、5、8 各向度內項目為單位 (觀察指標變項)，並依各向度內因子及項目間之包含隸屬型態 (試探性因素分析結果架構)，使用 Amos 程式內定之最大可能性估計法 (Maximum Likelihood, ML) 進行驗證性因素分析。

第二樣本群描述統計及相關，解題反思二分層面 M 界於 3.926 至 4.146 間， SD 界於 1.414 至 1.589 間， r 值 .494，解題錯誤偵測五分層面 M 界於 2.537 至 3.676 間， SD 界於 .899 至 1.613 間， r 界於 .291 至 .757 間 (平均 .498)，各分層面彼此與各分層面總分間相關界於 .029 至 .953 間 (平均 .428)，94.44% 達 $p < .05$ ，91.67% 達 $p < .001$ ，代表各分層面內及彼此間大部份具有良好關聯、內部一致性良好，如表 9 至 10 所示。表 10 中「正確無誤」與「檢查行動、方式及解題反思」間幾近零相關，其因可能為個體進行正確無誤題型作業與平常未經錯誤設計作業之解題經驗接近，故未出現強烈覺察需要解題反思檢查之動機信念，與緒論所提解題檢核能力含反思向度運作可能受後設認知信念影響之前提相符。

表 9 解題檢核評量不同向度分層面描述統計 ($N2 = 432$)

統計量	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤	解題錯 誤偵測
M	4.146	3.926	8.072	3.225	2.926	2.537	3.245	3.676	15.609
SD	1.414	1.589	2.793	1.131	1.480	1.613	1.168	.899	5.098

表 10 解題檢核評量不同向度分層面相關 ($N2 = 432$)

	檢查行動	檢查方式	解題反思	答案計算錯誤偵測	運算程序錯誤偵測	題目語意脈絡錯誤偵測	數量訊息擷取錯誤偵測	正確無誤
檢查方式	.494**							
解題反思	.953**	.733**						
答案計算錯誤偵測	.355**	.212**	.351**					
運算程序錯誤偵測	.269**	.195**	.278**	.671**				
題目語意脈絡錯誤偵測	.406**	.190**	.384**	.757**	.647**			
數量訊息擷取錯誤偵測	.208**	.196**	.231**	.526**	.558**	.494**		
正確無誤	.029	.096*	.056	.291**	.303**	.267**	.464**	
解題錯誤偵測	.355**	.233**	.359**	.871**	.853**	.858**	.756**	.494**

註：*概率值為 $p = .046$ ；**概率值為 $p = .000$ ，前述均為 $p < .05$

結構參數如圖 3、分析指標參數如表 11（圖 3 及其說明因軟體原估算輸出關係，部分輸出數據僅至小數二位）。

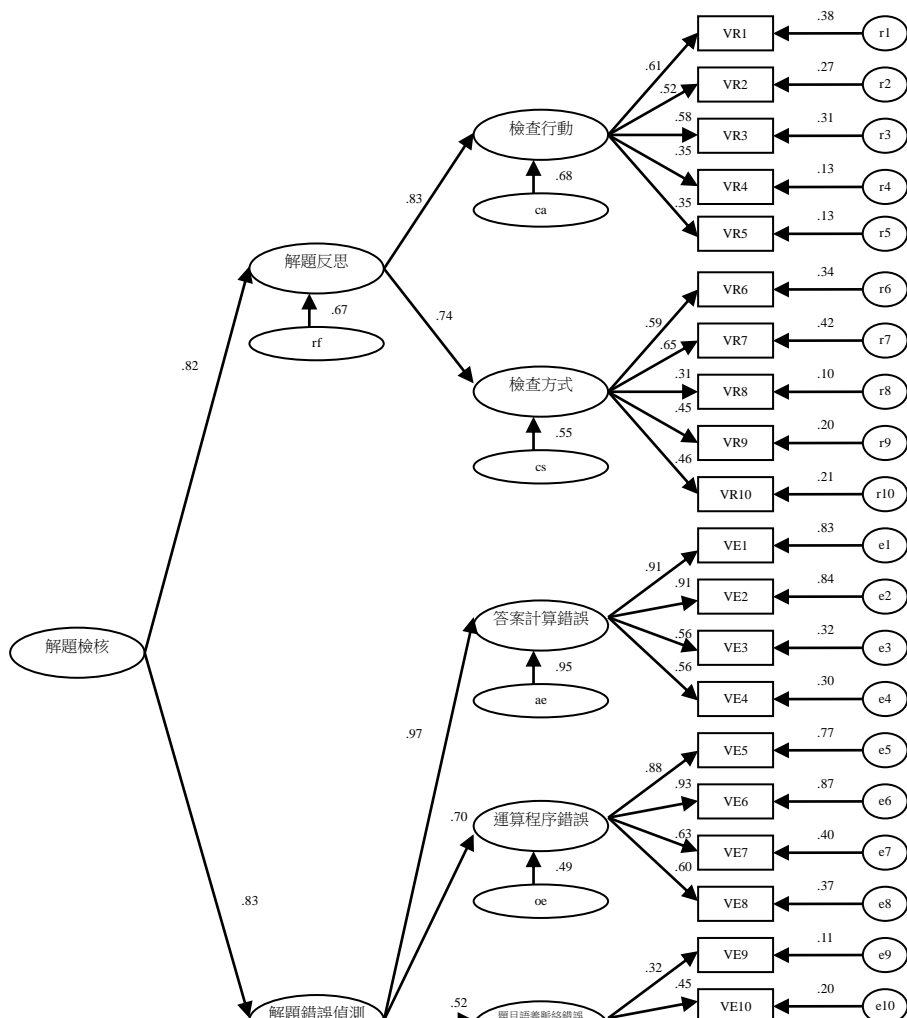


圖 3 解題檢核評量反思及錯誤偵測二向度驗證性因素分析結構 ($N2 = 432$)

首先自圖 3 結構看來，因素與觀察指標變項相關（標準化因素負荷量）及測量變異解釋貢獻量觀之，反思向度第一分層面因素負荷量及變異解釋量界於 .35 至 .61、12%至 37%間、共同萃取變異解釋比為 24.23%，第二分層面因素負荷量及變異解釋量界於 .31 至 .65、10%至 42%間、共同萃取變異解釋比為 25.62%。所有觀察指標變項對二個潛在因素的共同萃取變異解釋比為 24.92%，而二分層面因素負荷量及變異解釋量界於 .74 至 .83、55%至 68%間、向度共同萃取變異解釋比為 61.83%。

錯誤偵測向度第一分層面因素負荷量及變異解釋量界於 .55 至 .91、30%至 84%間、共同萃取變異解釋比為 56.81%，第二分層面因素負荷量及變異解釋量界於 .60 至 .93、37%至 87%間、共同萃取變異解釋比為 59.91%，第三分層面因素負荷量及變異解釋量界於 .32 至 .97、11%至 94%間、共同萃取變異解釋比為 53.24%，第四分層面因素負荷量及變異解釋量界於 .40 至 .97、16%至 95%間、共同萃取變異解釋比為 51.98%，第五分層面因素負荷量及變異解釋量界於 .50 至 .98、25%至 96%間、共同萃取變異解釋比為 60.11%。所有觀察指標變項對五個潛在因素的共同萃取變異解釋比為 56.01%，而五分層面因素負荷量及變異解釋量界於 .51 至 .97、26%至 95%間、向度共同萃取變異解釋比為 61.39%。

二向度所有三十個觀察指標變項對各潛在因素的共同萃取變異解釋比為 45.88%，而七分層面因素負荷量及變異解釋量分別界於 .52 至 .97、27%至 95%間、共同萃取變異解釋比為 61.70%。全評量二向度因素負荷量及變異解釋量分別界於 .82 至 .83、67%至 69%間、共同萃取變異解釋比為 68.07%。

參照上述，若以李茂能（2006）建議標準觀之，以上所提變異萃取解釋比，除反思向度之第一至二分層面觀察指標變項及總合未達 50%外，餘均達 50%以上，表示觀察指標變項對所測構念之代表性尚可接受（變異萃取解釋比之計算公式請參見李茂能，2006，頁 144）。圖 3 模式各迴歸係數、變異數參數均達 .05 顯著水準以上（均為 $p = .000$ ）。

表 11 模式整體適配度檢定項目分析 (N2 = 432)

向度	項目	絕對適配檢定				增值／相對適配檢定						精簡適配檢定			
		χ^2	GFI	SRMR	RMSEA	AGFI	NFI	RFI	IFI	TFI	CFI	PNFI	PCFI	ECVI	CMIN/DF
	適配標準	儘可能小 $p > .05$	> .90	< .05	< .05 優 < .08 良	> .90	> .90	> .90	> .90	> .90	> .90	> .50	> .50	儘可能小 (90%信 賴區間下限~上限)	1~3
解題 檢核	統計量	617.991 ($p = .000$) ($df = 398$)	.917	.0439	.04 ($p = .940$)	.906	.940	.922	.971	.962	.971	.716	.739	2.056 (1.902~.2.227)	1.553
	評估結果	-	+	+	+(優)	+	+	+	+	+	+	+	+	+	+

註：以上項目及標準係參考自王保進（2004）、余民寧（2006）、吳明隆（2008）、李茂能（2006）、邱皓政（2003）、蘇素美與吳裕益（2008）之建議；
+表符合-表不符合

由於余民寧（2006）、李茂能（2006）認為 SEM 整體模式適配度應綜合多重指標衡量之，故依前述學者所建議之模式適配度檢定項目及標準看來，表 11 十四個模式適配檢定項目中，十三項達理想標準、一項未符合，適配率為 92.86%，再就余民寧（2006）所提分類挑選重要指標比對原則而言，絕對適配檢定：GFI、AGFI、SRMR、RMSEA，增值／相對適配檢定：NFI、RFI、IFI、TLI、CFI，精簡適配檢定：PNFI、PCFI、ECVI、CMIN/DF 等均達標準。準此評估，本次模式整體適配程度應在可接受範圍之內，亦即試探性因素分析所得因子因素結構假定目前暫時得以成立、未被推翻。

整體而言，本次驗證性因素分析各項結果尚可接受，且因素結構及因子相關均符合前述試探性因素分析結果之假定，加上因素模式結構適配程度尚屬適當。因此解題檢核評量所測試之解題反思能力可暫時被確認包含檢查行動、檢查方式二項，解題錯誤偵測能力可暫時被確認包含偵測答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤、數量訊息擷取錯誤、正確無誤五項，各自涉及各項對應能力之運用。

二、效標關聯效度分析

採皮爾遜積差相關分析解題檢核不同向度、分層面與效標間之關係，如表 12 所示。

表 12 解題檢核評量不同向度、分層面效標關聯效度分析 (N1 = 432)

效標	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤	解題錯誤 偵測
單步驟加減法解題 與驗算單元課堂習 作教師批閱分數	.828**	.722**	.835**	.849**	.739**	.350**	.709**	.906**	.829**

註：**概率值均為 $p = .000$ ， $p < .05$

自表 12 可看出，無論不同向度、分層面，均與效標間具有顯著之中至高度正相關 ($r = .35$ 至 $.906$ ， $p = .000$)，自上述說明看來，解題檢核評量不同向度、分層面均具良好之效標關聯效度。

三、信度分析

(一) 傳統信度

研究者使用 SPSS for Windows 對資料進行內部一致性 α 係數分析，結果如表 13。

表 13 因素歸屬後解題檢核評量不同向度及分層面內部一致性分析 ($N1 = 432$)

統計量	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤	解題錯 誤偵測
標準化 α	.804	.803	.829	.798	.779	.787	.774	.771	.815
刪題建議	無	無	無	無	無	無	無	無	無
題數	5	5	10	4	4	4	4	4	20

註：因表中各向度任刪一題其標準化 α 均會小於原標準化 α 數據，故各向度均無刪題建議

表 13 所有因素歸屬後各分層面數值均大於 .70 以上，而二向度之標準化 α 係數 .829 及 .815，代表因素歸屬後之二向度分層面內部一致性尚屬良好。

(二) 組合信度

研究者依據李茂能 (2006) 所提使用標準化因素負荷量之計算公式 (係轉化自 Reuterberg 與 Gustafsson 於 1992 所提使用非標準化迴歸係數之計算公式) 對前述 Amos 進行 CFA 所得資料進行組合信度分析，結果如表 14。

表 14 因素歸屬後解題檢核評量不同向度及分層面組合信度分析 ($N2 = 432$)

統計量	檢查行動	檢查方式	解題反思	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤	解題錯 誤偵測
觀察指標數	5	5	10	4	4	4	4	4	20
標準化因素 負荷量	.61、.53、 .56、.35、 .35	.59、.65、 .31、.45、 .46	參照左列 共 10 項 數據	.91、.91、 .56、.55	.88、.93、 .63、.6	.32、.45、 .94、.97	.97、.9、 .41、.4	.54、.5、 .98、.95	參照左 列共 20 項數據
組合信度	.603	.619	.759	.833	.852	.793	.798	.847	.958

若以李茂能 (2006) 建議標準觀之 (頁 145)，除二向度部份分層面因負荷量不高之影響外，餘表 14 各項數據都在 .80 以上，表示大部分觀察指標變項對所測構念之測量可靠性尚屬良好。

四、鑑別力分析 (極端組決斷值檢定)

為了解因素歸屬後解題檢核不同向度各單題、分層面及分向度的鑑別力，研究者採傳統測驗理論的作法，將個體解題檢核不同向度總分由高至低排序後，分別選取排序居前 27% 及後 27% 個體為高低分組 (反思向度 $N1H = 220$; $N1L = 118$ ，錯誤偵測向度 $N1H = 161$; $N1L = 161$)，之後以此二組在不同向度各單題、分層面及分向度表現進行獨立樣本 t 考驗，結果如表 15。

表 15 因素歸屬後不同向度各單題、分層面及分向度極端組決斷值檢定 ($N1 = 432$)

項目 t 值	檢查行動	檢查方式	答案計算 錯誤偵測	運算程序 錯誤偵測	題目語意脈 絡錯誤偵測	數量訊息擷 取錯誤偵測	正確 無誤
1	12.246**	25.907**	17.439**	12.419**	29.503**	9.879**	6.300**
2	12.681**	63.893**	16.506**	13.236**	38.079**	11.094**	5.551**
3	18.502**	51.465**	4.683**	19.150**	25.397**	16.196**	6.910**
4	25.907**	12.246**	4.211**	18.010**	26.432**	16.856**	7.031**
5	23.245**	14.055**					
分層面	29.301**	45.199**	16.581**	19.397**	38.031**	21.181**	7.936**
分向度	解題反思	46.431**	解題錯 誤偵測	32.348**			

註：均採用不假設變異數相等之分析數據

**概率值均為 $p = .000$, $p < .05$.

表 15 極端組表現決斷值分析之 t 考驗結果均達 .05 顯著水準，代表因素歸屬後不同向度各單題、分層面及分向度對不同能力個體的鑑別力良好。

五、不同解題表現個體解題檢核表現之實測結果分析 ($N1 = 432$)

為了解不同解題表現個體之解題檢核表現實測結果，將以不同解題表現（單步驟解題整體表現，如第二項工具解題評量平均表現前 27% 及後 27% 個體， $N1GH = 136$; $N1GL = 107$ ）為分組自變項，解題檢核反思及錯誤偵測分層面（檢查行動、檢查方式；答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤、數量訊息擷取錯誤、正確無誤）、各向度表現為依變項，進行多變項獨立樣本平均數考驗分析。以了解不同解題表現個體在解題檢核表現之差異，以下分別說明。

（一）共線性分析 (multicollinearity)

結果發現不論以三組各依變項互為被預測及預測變項進行迴歸分析檢視，依 Belsley、Kuh 與 Welsch (1980) 判斷標準，目前允差大（均近 1 未為 0）、VIF 小（1 至 1.484）、在同一特徵值上的條件指數小（1 至 17.004，1-30 表低度、30-100 表中度、100 以上表嚴重）、變異數比例均低於 1 之狀況看來 (.03 至 .68)，較無共線性問題，適合進行後續分析。

（二）不同解題檢核表現分層面表現之實測結果分析

1. 前置分析

其多變項變異數同質性考驗結果三項 Box's M 值為 .217、.652、.957 ($p = .135$)，單變項變異數同質性考驗結果，三組各依變項 Cochran's C 值分別為 .12 ($p = .321$)、.085；.056、.048、.061、.118、.139；.096、.060 ($p = .373$) 及 .569 ($p = .286$)，Bartlett-box F 值分別為 .803 ($p = .314$)、.516；.312 ($p = .141$)、.604、.686、.147、.153；.871 ($p = .245$)、.578，多變項及單變項均未達 .05 顯著水準，代表三組各依變項的多變項及單變項變異數同質，故研究者將續行分析。

2. 正式分析

其多變項考驗結果三項 Wilks' Λ 值為 .236、.046、.128，其 $F(2,240)(5,237)$ 值為 387.697、983.511、814.804 (均 $p = .000$)，均達 .05 顯著水準，代表不同解題表現個體間在反思及錯誤偵測各分層面、各向度表現整體上均具差異， η^2 效果值界於 .764 至 .954 間，亦即自變項對多個變項具界於 76.4% 至 95.4% 間之影響效果。

單變項考驗結果，各分層面、各向度各依變項的 $F(1,241)$ 值分別為 334.972、760.971、35.726、298.604、616.051、2264.038、2745.033、663.266、786.189 (均 $p = .000$)，均達 .001 顯著水準 (Bonferroni 多重檢定顯著水準校正，概率應分別小於 $.05/2 = .025$ 、 $.05/5 = .01$ ，以下均依此校正)，代表不同解題表現個體間在反思及錯誤偵測各分層面、各向度表現上均存有差異， η^2 效果值界於 .129 至 .919 間，亦即自變項對各單變項具界於 12.9% 至 91.9% 間之影響效果。因前述均僅有二組，故可直接以其平均數進行表現之差異比較，均為高解題表現組優於低解題表現組，反思向度 M 為 $4.981 > 2.449$ 、 $4.971 > 1.682$ ， SD 為 0.104、1.627、0.170、1.378，錯誤偵測向度 M 為 $3.963 > 3.336$ 、 $3.919 > 1.813$ 、 $3.934 > 2.037$ 、 $3.988 > 0.533$ 、 $3.990 > 0.346$ ， SD 為 0.256、1.189、0.323、1.375、0.327、0.812、0.140、0.850、0.103、0.814，二向度整體 M 為 $9.971 > 4.131$ 、 $11.882 > 5.495$ ， SD 為 0.170、2.639、0.531、2.589。

上述各組各表現平均數原始分數，參照其平均數及標準差數據線性轉換後之共同量尺 T 分數表現剖面圖 (profile) 如圖 4。

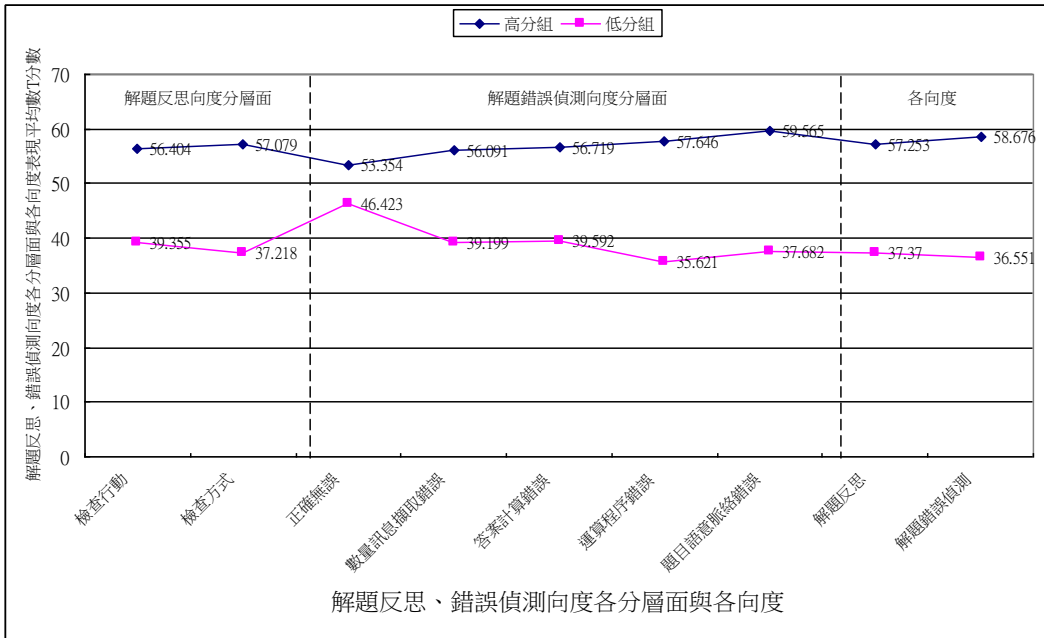


圖 4 不同解題表現解題反思、解題錯誤偵測各分層面與各向度表現 (平均數線性轉換 T 分數)

再進行 Roy-Bargman 降步 F 考驗，結果發現反思二分層面、反思及錯誤偵測二向度，均為第一依變項（檢查行動、反思向度） $F(1,241)$ 值為 334.972、663.266（均 $p = .000$ ），達 .001 顯著水準。在排除第一依變項（檢查行動、反思向度）影響後，第二依變項（檢查方式、錯誤偵測向度） $F(1,240)$ 值為 184.864、258.278（均 $p = .000$ ），仍達 .001 顯著水準。

而錯誤偵測五分層面，第一依變項（正確無誤） $F(1,241)$ 值為 35.726（ $p = .000$ ），達 .001 顯著水準。自第二依變項起依序漸次排除前列各依變項（數量訊息擷取錯誤、答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤）影響後，各依變項（數量訊息擷取錯誤、答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤） $F(1,240$ 至 1,237) 值為 245.319、270.119、314.620、211.819（均 $p = .000$ ），仍均達 .001 顯著水準，代表在排除變項重疊變異後，自變項對各依變項均具影響，亦即不同解題表現個體間在正確無誤、數量訊息擷取錯誤、答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤表現上確具差異，亦即不同解題表現對正確無誤、數量訊息擷取錯誤、答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤表現均具影響。

此外 Wilks 聯合多變項 (joint multivariate) 及個別單變項 (individual univariate) 95% 信賴區間考驗結果，前述各依變項區間值均不含零，代表前述各依變項，在不同解題表現個體間存有真正差異。

綜上可知多變項考驗顯著之因，乃為各依變項之貢獻，故可確認不同解題表現個體間在解題反思、錯誤偵測向度及分層面表現上存有真正差異（由其平均數看來，高解題表現組優於低解題表現組）。

六、綜合歸納與討論

(一) 評量發展

觀諸各項性能指標分析結果，無論內部相關與決斷值（鑑別力），或構念效度（試探性及驗證性因素分析，含因素萃取、萃取變異比及模式適配度），抑或內部一致性 α 係數及組合信度測試數據尚屬良好，亦即本評量在測驗工具編製上具有相當程度之良好品質。

再就因子結構而言，共分反思、錯誤偵測二向度，前者含二個分層面，後者含五個分層面，此項證據清楚指出，本評量內容架構已達成原初編製研究目標（二向度範疇解題檢核），且與整合自 Abbas (2016)、Jitendra 等人 (2013)、Morin 等人 (2017)、Peltier 與 Vannest (2016a)、Peltier 與 Vannest (2016b)、Van Boxtel (2016) 所提之解題反思向度內容，以及整合自 Bunt 等人 (2013)、Pratt 與 Urbanowski (2016)、Yin 等人 (2016)、Zhang (2017) 所提之解題錯誤偵測向度內容相符，同時意指解題反思與錯誤偵測間之檢核整合運作，故本評量在理論基礎及目標達成上均具良好適配性。

(二) 實測結果

觀諸對不同解題表現個體之實測分析結果，發現其在反思、錯誤偵測二向度表現上均具多變項及單變項差異，且其多變項降步分析結果顯示，二向度及各分層面表現均具有真正差異，亦即其具顯著差異之基礎確係來自於其各自之貢獻，無論排除何者，其各自表現仍具顯著差異。

故不同解題表現學生，在解題反思、錯誤偵測二向度表現上具有真正差異，若就解題表現而言，此項亦符應 Abbas (2016)、Bunt 等人 (2013)、Jitendra 等人 (2013)、Morin 等人 (2017)、Peltier 與 Vannest (2016a)、Peltier 與 Vannest (2016b)、Pratt 與 Urbanowski (2016)、Van Boxtel (2016)、Yin 等人 (2016)、Zhang (2017) 研究結果，亦即解題反思與錯誤偵測對個體解題表現應具有影響，這也為融入前二者之檢核策略教學具有更佳學習促進效果的發現提供了基礎性支持證據。易言之，若能以解題策略教學形式來提昇個體解題檢核能力，對其學習表現將有良性影響，故將解題檢核策略融入於教學內容中應有其必要性。

不過若欲提昇或促進個體解題檢核表現，恐不能僅止於短期教學策略介入 (strategy intervention) 因素之考量，而必須自個體認知發展的寬廣視野進行思考，因此對個體解題檢核表現進行長期連續性追蹤研究實屬必要。

結論與建議

一、結論

基於前述結果發現，在本研究所提之各項問題上，暫已獲致相當程度之釐清，分述如下：

(一) 解題檢核評量具有良好之解題檢核能力測量效度（向度內及各向度）

本評量具良好構念效度，除各潛在特質測量向度性良好外，並隱含「檢查行動、檢查方式，偵測答案計算錯誤、運算程序錯誤、題目語意脈絡錯誤、數量訊息擷取錯誤、正確無誤」七個各向度內分層面結構，及「反思、錯誤偵測」二個分向度運作結構，意指本評量之構念測量效度良好。已達成原初編製研究目標（二向度範疇解題檢核），且與整合自 Abbas (2016)、Morin 等人 (2017) 所提之解題反思向度內容，以及整合自 Bunt 等人 (2013)、Zhang (2017) 所提之解題錯誤偵測向度內容相符。

(二) 解題檢核評量具有良好之解題檢核能力測量信度（向度內及各向度）

本評量具良好傳統內部一致性及組合信度，前者意指本評量各向度項目、分層面、各向度的內部一致性良好，後者則意指本評量在因子結構下之各分層面構念測量上的內部一致性良好，意指本評量之構念測量信度良好。已達成原初編製研究目標（二向度範疇解題檢核），且與整合自 Peltier 與 Vannest（2016a）、Peltier 與 Vannest（2016b）所提之解題反思向度內容，以及整合自 Pratt 與 Urbanowski（2016）所提之解題錯誤偵測向度內容相符。

（三）解題檢核評量具有良好之解題檢核能力測量鑑別力（向度內及各向度）

本評量具良好構念測量決斷值，意指各項目、各分層面、各向度在構念測量上之鑑別力良好，能有效區分不同能力個體在各構念上之表現，並能有效反應所測構念。已達成原初編製研究目標（二向度範疇解題檢核），且與整合自 Jitendra 等人（2013）、Van Boxtel（2016）所提之解題反思向度內容，以及整合自 Yin 等人（2016）所提之解題錯誤偵測向度內容相符。

（四）不同解題表現個體之解題反思、錯誤偵測向度與分層面表現具有真正差異

不同解題表現個體在解題反思、錯誤偵測向度與分層面表現上均存有真正差異（高表現組優於低表現組），亦即個體解題能力對解題檢核表現具有影響。同時亦符應 Abbas（2016）、Bunt 等人（2013）、Jitendra 等人（2013）、Morin 等人（2017）、Peltier 與 Vannest（2016a）、Peltier 與 Vannest（2016b）、Pratt 與 Urbanowski（2016）、Van Boxtel（2016）、Yin 等人（2016）、Zhang（2017）之研究結果。

二、建議

（一）教學及評量

教學者可於單步驟數學文字題解題教學歷程中後階段，運用本評量進行不同目的之輔助性評量。例如於教學中及教學後，配合教學題型內容使用以了解學習的進展或困難是否與其解題檢核能力有關，若發現解題學習落後者的解題檢核能力確有較弱之現象，則可能必須在目前或日後的解題學習活動中，加入解題檢核策略教學，例如配合程序性協助（procedural facilitator）使用各項解題檢核評量（整合解題反思、錯誤偵測活動）進行策略融入教學，此不僅有助於確認個體解題表現與其解題檢核能力之關聯，同時亦利於教學者隨時調整給予學習者之教學協助，提昇其學習效率。

（二）未來研究

後續研究可再針對影響個體解題學習之重要因素，如不同年級、不同題型（如加減乘除法之多則混合題型等）及不同學習內容（如圖形面積體積、比例、速度問題等），整合解題反思、錯誤偵測向度，進行各類解題檢核評量之編製發展（含解題經驗前導誘發問題），以測試解題檢核能力在不同解題學習內容中的表現。

同時可擴大解題檢核評量之發展及驗證範疇，如其他內容學習領域，並作為確認個體解題檢核能力狀態之研究工具，亦可搭配動態評量（dynamic assessment）進行個體解題檢核潛能（learning potential）發展評估及促進。此外更可透過不同取樣調查設計以確認不同解題表現個體之解題檢核表現，並進行發展性研究。

三、研究限制

本研究結果雖尚良好，但仍有潛在限制，一為顧及版本與進度統一，取樣數及區域受限，故研究結果可能受樣本依賴影響。二為依班級教師建議，量表評定點數應衡量國小學生閱讀及判斷能力不宜過多，故研究結果亦可能受量表評定點數影響。三為明確列出各種單一或組合可能，解題反思評量之檢查方式選項較多，學童可能難以理解其意，加上單選題形式，較難真實反應引導學童做多種不同檢查之數學教學現場實務，前述雖已設置可讓學童選擇及自由表達之其他選項因應之，但研究結果仍可能受其影響。

參考文獻

- 王保進(2004)：多變量分析—套裝程式與資料分析。臺北市：五南。[Wang, B. J. (2004). *Multivariate analysis and data analysis package*. Taipei, Taiwan: Wu Nan.]
- 余民寧(2006)：潛在變項模式：SIMPLIS 的應用。臺北：高等教育。[Yu, M. N. (2006). *Latent variable models: The application of SIMPLIS*. Taipei, Taiwan: Higher Education.]
- 吳明隆(2008)：結構方程模式 AMOS 的操作與應用。臺北：五南。[Wu, M. L. (2008). *AMOS structural equation model of operation and application*. Taipei, Taiwan: Wu Nan.]
- 吳昭容(1990)：圖示對國小學童解數學應用題之影響。未出版獨立研究，國立臺灣大學心理學研究所，臺北市。[Wu, Z. R. (1990). *The impact of schema drawing on mathematical word problem solving for elementary school students*. Institute of Psychology, National Taiwan University.]
- 李茂能(2006)：結構方程模式軟體 Amos 之簡介及其在測驗編製上之應用。臺北：心理。[Lee, M. N. (2006). *Amos structural equation modeling software in the Introduction and Application on the test preparation*. Taipei, Taiwan: Psychology.]
- 李源順主編(2016, 2017)：國小數學課本(第一至四冊)。臺南：南一。[Lee, Y. S. (editor) (2016, 2017). *Elementary mathematics textbooks (first to fourth volume)*. Tainan, Taiwan: Nan Yi.]
- 林長壽主編(2016, 2017)：國小數學課本(第一至四冊)。臺南：翰林。[Lin, C. S. (editor) (2016, 2017). *Elementary mathematics textbooks (first to fourth volume)*. Tainan, Taiwan: Han lin.]
- 林清山(1991)：多變項分析統計法。台北：東華。[Lin, Q. S. (1991). *Statistical method for multivariate analysis*. Taipei, Taiwan: Dong Hua.]
- 邱皓政(2001)：社會與行為科學的量化研究與統計分析：SPSS 中文視窗版資料分析範例解析。臺北市：五南。[Qiu, H. Z. (2001). *Social and behavioral sciences quantitative research and statistical analysis: SPSS Chinese Windows version of the data analysis sample resolution*. Taipei, Taiwan: Wu Nan.]
- 邱皓政(2003)：結構方程模式：LISREL 的理論、技術與應用。臺北：雙葉。[Qiu, H. Z. (2003). *Structural equation model: LISREL theory, technique and application*. Taipei, Taiwan: Yeh Yeh.]

- 陳正昌、程炳林、陳新豐、劉子鍵 (2012)：多變量分析方法—統計軟體應用 (六版)。臺北市：五南。[Chen, C. C., Cherng, B. L., Chen, X. F., & Liu, Z. C. (2012). *Multivariate analysis: Statistical software application* (sixth edition). Taipei, Taiwan: Wu Nan.]
- 陳蜜桃、李新民、黃秀霜 (2007)：台灣地區國民小學教師實用智能、情緒智能與專業表現之相關研究。《教育心理學報》，39 (2)，295-316。[Chen, M. T., Lee, H. M., & Huang, H. S. (2007). The relationship among practical intelligence, emotional intelligence and professional performance of taiwanese elementary school teachers. *Bulletin of Educational Psychology*, 39(2), 295-316.]
- 教育部 (2008)：國民中小學九年一貫課程綱要數學學習領域 97 年數學課程綱要修訂。臺北：教育部。[Ministry of Education (2008). *Grade 1-9 curriculum guidelines revision in mathematic learning areas*. Taipei: Ministry of Education.]
- 楊瑞智主編 (2016, 2017)：國小數學課本 (第一至四冊)。臺北：康軒。[Yang, R. Z. (editor) (2016, 2017). *Elementary mathematics textbooks* (first to fourth volume). Taipei, Taiwan: Kang Hsuan.]
- 蘇素美、吳裕益 (2008)：害羞量表之修訂及其效度研究。《教育與心理研究》，31 (4)，53-82。[Su, S. M., & Wu Y. Y. (2008). A study on the revise of the shyness scale and its validity. *Journal of Education & Psychology*, 31(4), 53-82.]
- Abbas, R. S. S. (2016). A program based on developmental mathematics approach to develop higher order mathematical thinking levels and mathematics appreciation for primary stage students. *Education*, 136(3), 378-390.
- Bunt, A., Terry, M., & Lank, E. (2013). Challenges and Opportunities for Mathematics Software in Expert Problem Solving. *Human-Computer Interaction*, 28(3), 222-264.
- Bandalos, D. L., & Finney, S. J. (2001) Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). New Jersey: Lawrence Erlbaum Associates.
- Belsley, D. A., Kuth, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley.
- Byrne, B. M. (2001). *Structural equation modeling with Amos: basic concepts, applications and programming*. New Jersey: Lawrence Erlbaum Associates.
- Crocker, L., & Aligina, J. (1986). *Introduction to classical and modern test theory*. Orlando, Florida: Holt, Rinehart and Winston, Inc.
- Fennell, F. M., & Speer, W. R. (2013). *Defining mathematics education: presidential yearbook selections 1926-2012: Seventy-fifth yearbook*. Reston, VA: National Council of Teachers of Mathematics.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

- Jitendra, A. K., Rodriguez, M., Kanive, R., Huang, J. P., Church, C., Corroy, K. A., & Zaslofsky, A. (2013). Impact of small-group tutoring interventions on the mathematical problem solving and achievement of third-grade students with mathematics difficulties. *Learning Disability Quarterly*, 36(1), 21-35.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kong, J. E., & Orosco, M. J. (2016). Word-problem-solving strategy for minority students at risk for math difficulties. *Learning Disability Quarterly*, 39 (3), 171-181.
- Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R. L. Linn (ed.), *Educational measurement* (3rd ed, pp. 335-366). Phoenix, Indiana: Oryx Press.
- Morin, L. L., Watson, S. M. R., Hester, P., & Raver, S. (2017). The use of a bar model drawing to teach word problem solving to students with mathematics difficulties. *Learning Disability Quarterly*, 40(2), 91-104.
- Özcan, Z. Ç. (2016). The relationship between mathematical problem-solving skills and self-regulated learning through homework behaviours, motivation, and metacognition. *International Journal of Mathematical Education in Science and Technology*, 47(3), 408-420.
- Peltier, C., & Vannest, K. J. (2016a). Schema-based strategy instruction and the mathematical problem-solving performance of two students with emotional or behavioral disorders. *Education & Treatment of Children*, 39(4), 521-543.
- Peltier, C., & Vannest, K. J. (2016b). Utilizing the STAR Strategy to Improve the Mathematical Problem-Solving Abilities of Students with Emotional and Behavioral Disorders. *Beyond Behavior*, 25(1), 9-15.
- Peltier, C., & Vannest, K. J. (2018). The effects of schema-based instruction on the mathematical problem solving of students with emotional and behavioral disorders. *Behavioral Disorders*, 43(2), 277-289.
- Pratt, S. M., & Urbanowski, M. (2016). Teaching early readers to self-monitor and self-correct. *The Reading Teacher*, 69(5), 559.
- Powell, S. R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice (Wiley-Blackwell)*, 26(2), 94-108.
- Van Boxtel, J. M. (2016). REASON: A self-instruction strategy for twice-exceptional learners struggling with common core mathematics. *Teaching Exceptional Children*, 49(1), 66-73.

- Verschaffel, L., Greer, B., & DeCorte, E. (2007). Whole number concepts and operations. In F. K. Lester (Eds.), *Second handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 557-628). Charlotte, NC: Information Age Publishing Inc.
- Yin, H.-Z., Li, D., Yang, J., Li, W., Qiu, J., & Chen, Y.-Y. (2016). Neural bases of unconscious error detection in a chinese anagram solution task: Evidence from ERP study. *PLoS ONE*, *11*(5), 1-12.
- Zhang, D. (2017). Effects of visual working memory training and direct instruction on geometry problem solving in students with geometry difficulties. *Learning Disabilities-A Contemporary Journal*, *15*(1), 117-138.

收稿日期：2017年08月23日
一稿修訂日期：2017年10月31日
二稿修訂日期：2018年03月04日
三稿修訂日期：2018年06月11日
接受刊登日期：2018年06月12日

Bulletin of Educational Psychology, 2018, 50(2), 243-266

National Taiwan Normal University, Taipei, Taiwan, R.O.C.

Development and Analysis of a Checking Assessment for Word Problem Solving in Primary School Learners

Chia-Hua Hsu

Department of Education

National Chiayi University

This study developed a Checking Assessment for Single-Step Word Problem-Solving for Addition and Subtraction (CA-SSWPS-AS) using componential and content analyses. The checking ability component was divided into two dimensions: reflection and error detection. The reflection dimension consisted of checking actions and checking skills, whereas error detection comprised errors in answer calculation, operational procedure, problem semantic context, and quantitative information catching, as well as a correct (no error) facet. The single-step word problems for addition and subtraction were derived from various mathematics textbooks for first and second graders in Taiwanese elementary schools. Componential and content analyses were used to examine the checking operations and semantic schema involved in single-step word problem-solving for addition and subtraction. The results were then used in development of the CA-SSWPS-AS. Considering the homogeneity of learning contents and experiences in different schools, 864 second graders from Chiayi City and Chiayi County in Taiwan were selected as participants through convenience and cluster sampling. All participants were tested using the CA-SSWPS-AS combined with experience preactivated tasks in single-step word problem-solving for addition and subtraction. The data were divided into two clusters for use in random cross-validation statistical analyses. Validity analysis using exploratory and confirmatory factor analysis produced a number of factor structures that matched the aforementioned two dimensions and their facets. The variance proportions of factor extraction and model fitness were positive, as was criteria-related validity (using students' scores assessed by teachers in classroom performance-checking as criteria). Reliability analysis demonstrated a positive α coefficient and composite reliability. Regarding item analysis, the critical ratio was strong, regardless of whether a single item, a subassessment, or the whole was analyzed. Finally, reflection and error detection performance varied significantly when checking individuals with different word problem-solving performances.

KEY WORDS: Checking, Learning assessment, Single-step word problem, Word problem-solving