

九年一貫課程數學領域能力指標「數與量」、「代數」主題軸第一、二階段表現標準適切性評估之研究

杜佳真

中華技術學院通識教育中心

林世華

國立臺灣師範大學教育心理與輔導學系

摘要

本研究以九年一貫課程數學領域暫行綱要為例，利用 Angoff (1971) 所提出的表現標準設定程序和得懷術技術來評估由研究者重組之第一、二階段「數與量」和「代數」主題軸之數學基本能力表現標準的適切性。依據三階段程序進行，第一階段為準備階段，主要為學生實徵資料的收集和專家的邀集；第二階段為執行階段，主要進行三回合專家評估；第三階段為後續階段，主要以描述專家評估結果和收集專家回饋意見為主。經過 15 位專家代表進行 3 次評估問卷，結果發現：第一階段能力指標所指陳的學生表現標準較合乎我國學生的實徵結果，第二階段能力指標所指陳的表現標準較為簡單。

關鍵字：表現標準設定程序、得懷術、數學領域表現標準

壹、緒論

九年一貫課程能力指標系統的陳述，除了學習內容之外，有些學習領域尚包含學生表現標準的描述。這些標準多半以先前所訂定的中小學課程標準（教育部，1993）為基礎，並根據課程發展的經驗進行修訂，大多是以經驗校準的方式由指標撰寫的專家學者來制定，所以標準的高低位置的訂定，並沒有比較廣泛的討論和實徵性的學生資料的支持。

能力指標的建構應該要能以學生現況為依據，除了界定學生「應知」和「應能」的內容標準之外，尚需要運用實徵的方式來了解學生的實際能力表現，再訂定或調整符應學生實際表現的成就標準。因此，研究者以

九年一貫課程的數學領域暫行綱要（教育部，2002）進行重組所得的「數與量」和「代數」主題軸的指標內容為主，¹並且依據我國學生的實際表現邀請專家群體來評估暫行綱要所指陳的第一、二階段學生表現標準的適切性，作為未來能力指標系統修正和調整的參考。

基於上述研究動機，本研究的主要目的為：

透過標準設定的程序來評估由研究者重組之第一、二階段「數與量」和「代數」主題軸之數學基本能力表現標準的適切性。

1 在本研究中以由研究者以暫行綱要為基礎進行重組之後的能力指標細項為評估依據，指標細項對照詳見附錄。

貳、文獻探討

一、能力指標系統的建構和規劃

(一) 基本能力和基本學力的辯證

目前的九年一貫課程總綱並未明確界定「學力指標」或是「能力指標」兩者的差異，而在國內關於基本能力和基本學力的論爭中，對所謂「基本」的定義觀點較為一致，皆認為「基本」是指必要、基礎和核心周延等。較大的爭論在於「能力=學力」？就研究者收集整理相關文獻發現：能力和學力的概念相似，只是著眼點不同。整體來說，透過系統化學校教育的陶冶，讓學生學習到應有的學力，進而培養基本能力，而這些基本能力又成為進一步學力發展的基礎，所以，基本能力和基本學力乃是一體兩面、交感互動的發展（杜佳真，2004）。因此，在本研究中，將基本能力視為包含基本學力的總體集合，以下也將以「能力指標」統稱。

(二) 能力指標建構的原則

因應九年一貫課程的實施，中小學能力指標系統規劃的重要性相對提昇。指標規劃所涉及的問題分為技術性層面和政策性層面。黃政傑（1996）提出關於我國中小學基本學力指標的綜合性規劃研究，為目前國內較有系統討論能力指標的相關文件。

有關基本學力指標建構的原則包括以下四項：

1. 統整性原則：統整性表示內容是整體的，不是分化及割離的個別之物，為了顧及知識及能力的整體性發展，應

特別考慮統整性原則。

2. 可測性原則：指標內涵雖是敘述性內容，但是應可以透過調查或測驗了解是否達成指標所定的水準，以作為修訂課程及改進教學的參考。
3. 可行性原則：學力指標的建構，應考慮是否可行的問題，不可使其流於口號之吶喊或抽象之思辯內容，而缺少可驗證與實踐的可行性。
4. 前瞻性原則：學力指標具有時間性，應不斷加以更新，以符應時代的改變，原則上一套學力指標以十年為試用期限；至於學力指標與課程標準應是互相參酌之依據，也可作為修正的參考基礎，學力指標和學力調查的關係亦然。

(三) 能力指標的發展程序

透過研究小組，以及兩次的專家諮詢會議，關於我國中小學力指標發展程序大致分為四個階段。

第一階段為規劃階段，需時六個月；第二階段為指標發展階段，需時約兩年，包括十個步驟；第三階段為學力調查階段，約需時一年半的時間。包括七個步驟；第四階段為評鑑及修正階段，約需六個月，旨在由「學力指標發展統合委員會」定期召集各學習領域的學力指標發展研究小組和學力指標調查委員會小組成員，共同進行學力指標的總結性評鑑及修正事宜。整體發展的流程，如圖1所示。

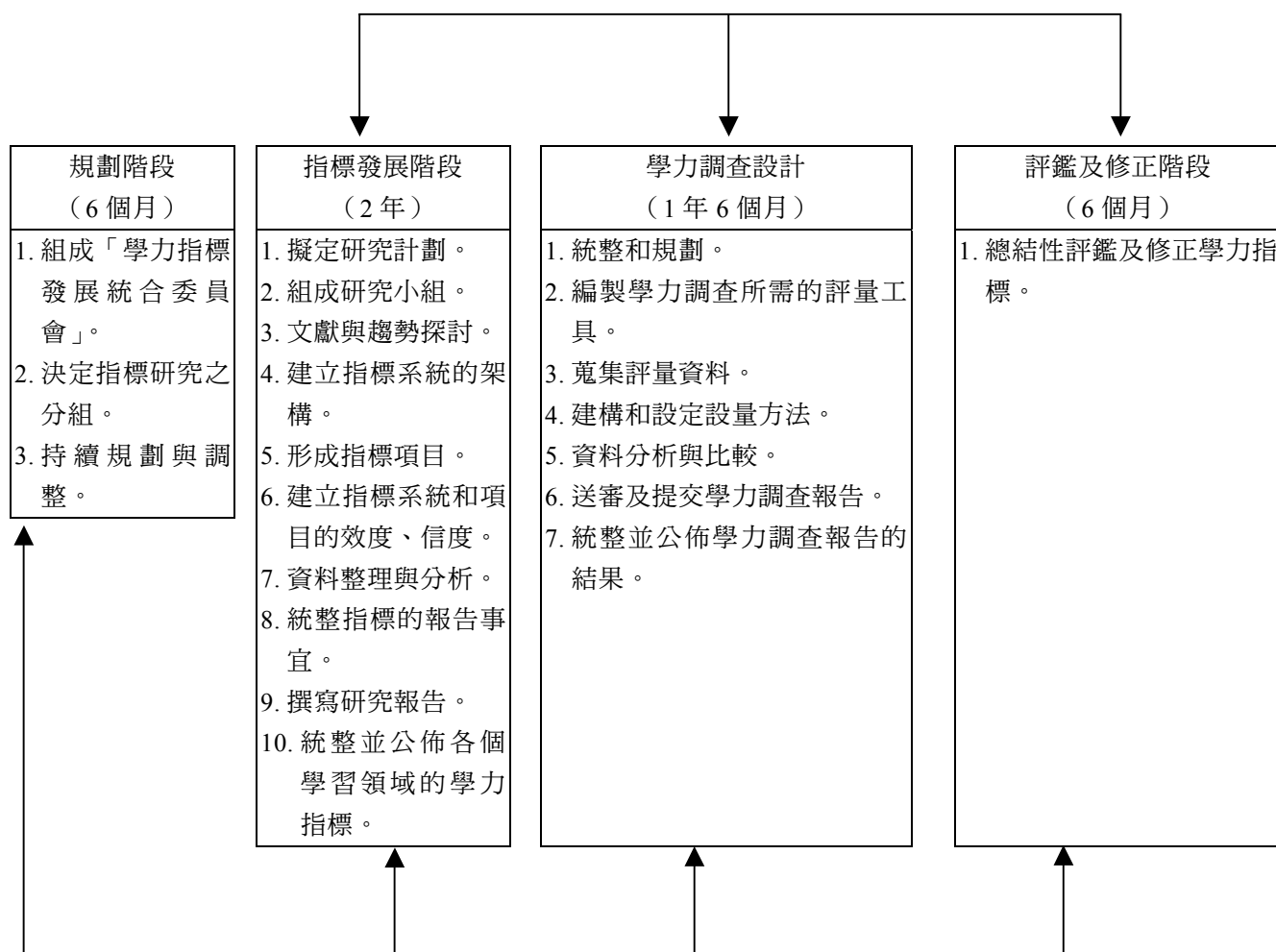


圖 1 國中小學基本學力指標發展流程圖 (資料來源：引自黃政傑，1996：120)

若以黃政傑（1996）對中小學學力指標的規劃原則和建構程序和目前能力指標系統進行對照，我們可以發現目前七大學習領域的指標系統在這些部分仍有待加強和整合；再從發展流程概念看來，目前各學習領域能力指標系統幾乎都在第二階段徘徊，指標系統的發展需要長時間的規劃，就黃政傑（1996）的架構為例，至少需費時 4 年半。但就現實面而言，因為政策面的限制和影響，並未能適切的考量技術層面。

因此在本研究中，擬透過學力調查的實徵資料來評估數學領域能力指標暫行綱要表現標準的適切性。

二、得懷術

得懷術(Delphi method)的主要目的在於獲取專家們的共識，尋求專家們對特定預測對象之一致性意見，此法不但可以收集思廣益之效，也可以得到專家獨立判斷的品質。得懷術的基本原則為：匿名性(anonymity)、操控的回饋性(controlled feedback)、意見反芻(iteration)、群組的反應統計(statistical group response) (張紹勳，2000)。

得懷術因為具有集思廣益的效果，同時也能夠得到專家們獨立判斷的品質，因此目前被廣泛應用到科技預測、公共政策分析以

及方案規劃的相關領域之中。

三、表現標準訂定的方法

九年一貫課程能力指標系統的陳述，除了學習內容之外，有些學習領域同時包含表現標準的描述。在學習的表現或是一些政策的擬定，表現標準所該具備的條件，可不是僅靠各自表述，或是一些約定俗成的門檻來做判斷的依據，必須要考量到學生的能力以及其他的相關因素，如特定政策或是整體的社會環境發展甚至是一些學生的實徵資料。所以，從 70 年代開始，就不斷的有學者提出一些標準設定的方式，企圖利用一些科學的程序來建立一個最合適的標準。

Angoff 法為目前最普遍被運用的標準設定方法，其實施程序為先依照目的選定一組具有代表性的評定專家組成委員會，讓各個專家了解其主要工作和所利用的資料以及實施程序，然後就每個細項逐項評定該表現層次中的學生可能表現，例如：基本層次中最低能力的學生是不是會答對這一題或是估計該題的通過率等，第一次的評估之後，專家們再根據一些實際的資料，甚至是透過和其他專家討論自己的評比結果，修正前一次的判斷進行第二次的評比，依照最後資料結果重複相同程序，然後求得專家群體評定結果的切割分數(cut score)，來代表該表現層次的標準點 (Angoff, 1971; Jaeger, 1989)。

四、標準設定的程序概述

無論是採用哪一種程序來訂定表現標準，都有一個典型的標準設定程序，像受試群體的訂定以及組織一個評定委員會來進行標準的設定等。而這個委員會工作的程序如下 (Hambleton, 1996)：

1. 檢視所依據的評量目的。
2. 展示評量內容和題目。

3. 發展和討論相關的表現類目。
4. 訓練和練習設定標準的方法。
5. 透過各種不同的方法讓委員訂定初步的標準。
6. 委員們相互討論他們的標準。
7. 重複標準設定的程序，直到委員之間的差異趨近平均。

(一) 評估專家委員會的組成

委員會要有哪些成員組成則是依照相關政策訂定。例如 NAEP (National Assessment of Educational Progress, NAEP) 的委員中有 70% 來自教育相關人員，30% 則是一般的社會大眾；而州際的教育評量，則在政策制定者和教育工作者之間採相等比例。無論是何種組合方式，主要的考量就是盡量是多元化以及具有代表性 (Reckase, 1998)，因為政治決策是一項複雜的行為，如果全體一致性太高，可能僅是一言堂，美其名為共識，但實質上卻是獨裁 (曾志朗, 2003)。因此，在本研究的評估委員會的組成，也盡量以不同領域的多元組合來進行考量。

不過，雖然多元組合的委員會具有推論的代表意義，在推選時仍要注意這些委員們是否具備一些相關的內容知識，因為這會影響評估結果的可信度。以 NAEP 為例，評定專家委員的產生有兩個階段：第一個階段在每一區的學校進行分層隨機抽樣，在每一區中，同時也提名出非教育領域的評估委員。而這些非教師的評估委員包括區域中的主管、教師組織的領導人、和州教育政府的職員等；以及非教育工作者包括市長、學校的領導人、管理階層或商業界等。並且將評估委員分為兩組分別進行 (Raymond & Reid, 2001; Reckase, 1998)。

至於專家要多少人才合適呢？有些程序只有 5 人，也有些程序邀集了 20-30 位專家

參與 (Hurtz & Hertz, 1999; Jaeger, 1991)，而 Angoff 程序一般而言以 10-15 人為最合適。

在本研究中，因為受限於研究者個人資源，無法透過像 NAEP 委員組成的方式來進行，僅能以個人名義邀集各領域具代表性的專家 15 人進行表現標準的評估。

(二) Angoff 法的程序

Arrasmith 和 Hambleton (1988) 提出 Angoff 法的四階段程序，概述如下：

1. 初始階段

主要提供用來進行標準設定測驗的背景訊息。在此階段中必須思考 6 個主要的問題，包括：測驗的主要目的是什麼？測驗結果所代表的意義是什麼？如果不應用標準設定的程序，測驗仍然可被運用嗎？各項資源的限制為何？是否有一些會影響測驗結果或是標準設定程序的顯著或是環境因素？例如僅限於某些地區使用等。測驗的發展歷程為何？

2. 標準設定的準備階段

在此階段的主要工作為：標準設定任務背景描述、專家的選擇以及專家資料的準備。

3. 標準設定專家會議的召開

此為執行的階段，包括以下工作：相關背景說明、標準設定程序說明等

4. 標準訂定會議的後續階段

主要為程序資料處理的階段，包括專家評定資料的處理和回饋。並同時進行最後文件的描述，例如把資料送給專家參考，並保留備分供未來使用的參考。

儘管標準設定的方法有一些差異，但核心的部分是相同的：包含優質的專家、合宜的訓練、數次的逐題評比過程以及提供讓專家進行更精確判斷的回饋，以及依據目的持

續的回顧和修正設定的程序 (Cizek, 1996; Reckase, 2000)。在標準設定的程序中，Angoff 法是目前最為廣泛應用的方法，甚至是日後修正標準設定方法的程序都以此方法為依據。

(三) 本研究的程序

雖然本研究的主要目的並不在於訂定明確的切割分數 (cut score)，但仍運用 Angoff 法的主要概念，以及得懷術的程序進行設計。由於在本研究中並不召集專家共同召開會議進行，因此，將 Arrasmith 和 Hambleton (1988) 所提出 Angoff 法的步驟，加以整理修改成三階段來進行。

1. 準備階段

包括提供來進行標準設定測驗的背景訊息描述，以及標準設定程序進行之前的準備工作。此階段可再區分為專家資料和文件資料兩大部分，文件資料則又再包含測驗的發展以及實測結果。

2. 執行階段

本階段主要描述標準適切性評估的步驟和執行過程。主要利用 Angoff (1971) 程序進行三回合的專家評估，第一回合的主要目的為瞭解專家的評估意見，透過專家全體和實徵研究的資料回饋，再於第二回合的評估中修改初始的評估結果，並在第三回合整理專家的評估資料，再比對能力指標中所指陳的表現標準適切性。

3. 後續階段

後續階段主要描述標準設定歷程的資料處理以及結果的分析，同時提供回饋問卷，進一步瞭解專家群體的意見和作為未來研究改進的建議。

參、研究方法和程序

本研究以實徵的方式來評定九年一貫課程數學領域「數與量」和「代數」主題軸中第一階段和第二階段學生表現標準的適切性。因此本研究運用 Angoff 法的主要概念，以及得懷術的程序進行設計，並將 Arrasmith 和 Hambleton (1988) 所提出 Angoff 法的步驟，加以整理修改成三階段進行說明：

一、準備階段

本階段以表現標準評估程序進行之前的準備工作為主。除了表現標準評定的專家學者邀請之外，尚包括評估文件的發展和結果描述。

(一) 試題的編擬和發展

本研究的表現標準設定程序，主要以研究者依據數學領域能力指標所指陳的第一、第二階段學生的表現標準所編擬的試題作為評估的依據。包括命題、審題預試等階段，並採用垂直等化設計來同步了解三、四、五年級學生在某一項數學能力的表現。

1. 命題

測驗的試題編寫主要以研究者根據所重新整合的九年一貫數學領域能力指標系統「數與量」和「代數」兩主題軸之內容為主，此重組指標乃是研究者依照目前所實施的數學領域暫行綱要（教育部，2002）的內容進行整合。試題以第一階段（三年級）、第二階段（五年級）的指標項目中的能力表現標準為主，由研究者參酌現行各版本教科書及參考書（如康軒、南一、翰林、國立編譯館等版本）、若干小學現有之形成性評量以及定期評量試卷、以及教育資源網站上相關資料（例如思摩特網、九年一貫課程網站等），和國際

間大型評量方案如：美國的「全國教育評量」（National Assessment of Educational Progress, 2002）、「國際數學和科學趨勢研究」（Trends in International Mathematics and Science Study, TIMSS）1995 年以及 1999 年（Trends in International Mathematics and Science Study, 1995; Trends in International Mathematics and Science Study, 1999）、澳洲學科評量（教育測驗中心，1998）等的相關試題所編擬而成，然後透過審題和預試的程序篩選適合該年級學生的試題所組合而成。

試題內容強調原創性，盡可能和坊間流傳和學生熟悉的試題類型區隔，並取材自學生的生活經驗，著重數學知識的理解和應用，而非評量學生的記憶或應試技巧。試卷題型由四選一選擇題以及填充題所組成，每份試卷的作答時間以不超過 40 分鐘為主。

2. 審題

題目初稿編擬好之後，先邀請數學領域能力指標暫行綱要中「數與量」和「代數」內容主題編擬之專家代表對試題內容進行初步審核。然後邀請同時具有豐富教學經驗（10 年以上）、熟悉測驗相關理論及評量實務工作背景的碩士層級以上、並有超過 10 年以上的教學資歷之專家教師共 4 人，連同研究者 5 人組成審題小組，針對題目的敘寫和核對所評量的能力逐題進行檢視；並且同時邀請經驗豐富且任教於中、高年級之現職教師共 6 名為試題的敘寫和呈現方式提供具體建議，最後再分階段彙集成預試試卷。

3. 預試

試題的預試對象為台北市長春國小和台北市五常國小的學生，分兩次實施。首先由三、五年級的學生進行，然後再由四、六年

級的學生進行第二次預試，最後以通過率 and 二系列相關係數作為選題的指標，因為主要是了解學生的基本能力表現，故刪除通過率小於.75、二系列相關係數小於.40，以及試題反應不符合學生發展之試題。預試的學生第一階段共 95 人，第二階段共 114 人，而預試題本第一階段約 55 題，第二階段 59 題，最後保留第一階段 43 題 and 第二階段 54 題作為正式試題。

4. 試題連結的等化設計

本研究有關於學生數學能力的表現標準評估，採用垂直等化的設計來同步瞭解三、四、五不同年級學生在某一項數學能力的表現，換句話說，先讓三年級學生 and 五年級學生分別回答在能力指標中屬於該階段能力的試題，而利用四年級學生的題本把三年級 and 五年級題本中的試題進行部分組合的設計，即可同時進行三、四、五年級的學生反應估計。三年級 and 四年級的共同題目有 19 題，四年級 and 五年級的共同題有 21 題，為考量學生作答時間 and 答題負荷，因此各年級的試卷均再分為甲、乙兩卷，同時調整選擇題答案，盡量讓在四個選項的正確答案均勻分布且不呈現規律，以降低學生僥倖猜測的機率。

5. 資料分析與統計處理

上述的資料分析分別以 SPSS11.0、SPSS12.0 計算學生的表現通過率以及 BILOG-3.1 了解試題的二系列相關係數。

(二) 實測情況

1. 測驗的實施狀況

測驗實施的時間為九十一學年度的第二學期，約在 2003 年 4 月下旬至 5 月中旬。各校施測時間考量 and 配合各學校 and 施測班級狀況，但最早施測學校 and 最晚施測學校的前後間隔不超過 3 週，並且在施測的同時了解學生課程的實施狀況 and 進度，以及所使用的教

科書版本，經研究者分析後發現各校並沒有太大的差異。

為讓學生能有充裕的答題時間，並再將試卷分成甲、乙兩卷，兩卷施測的時間各為一節課 40 分鐘，且兩卷並非連續進行施測，以排除學生的疲勞效應，甲、乙兩卷施測的間隔也不超過一週的時間。

2. 抽樣程序以及施測樣本分佈

本研究以教育部統計處所公布的國小學生統計概況資料為主（教育部統計處，2003），採用分層隨機抽樣，先將母群區分為北（苗栗以北）、中（中、彰、投地區）、南（高、高、屏地區）、東（宜蘭、花蓮、台東地區） and 離島（澎湖、金門、馬祖）等五區，並按照學生人數比例採用分層隨機抽樣，並同時考量受試學生之城鄉背景差異是否平均，不過因為受限於學校行政運作，所以實際抽測人數以原班級作為主要抽樣單位，因此與預期人數會有所差距，最後抽取三年級學生 513 人、四年級學生共 517 人、五年級學生共 522 人，合計 1552 人進行施測。施測學校以及受試人數等詳細資料 and 比例如表 1 所示。

3. 試題分析和學生答題反應的估計

運用三參數模式來進行學生答題反應的估計。並利用 Ultra Edit-32 進行學生資料處理編輯 and 整理、EXCEL2000 進行學生的資料統計 and 樞紐分析、SPSS11.0 以及 SPSS12.0 進行古典通過率計算並考驗受試樣本差異性、描述學生樣本的特性和因素分析等統計處理、以及 BILOG-3.1 計算 a、b、c 參數以及能力估計值(θ)並同時進行 model-fit 的試題檢定、利用 MATLAB5.3 進行三參數估計的矩陣運算、試題特性分析、繪圖等的資料處理 and 統計工作。

表 1 施測學校及人數一覽表

地區別	總人數		比例		抽樣學校			三年級抽樣人數		四年級抽樣人數		五年級抽樣人數	
			預計	實際	比例	預計	實際	比例	預計	實際	比例		
台北市	194795		國語實小	70	63		70	61		70	64		
台北縣	317301		自強國小	30	28		30	30		30	30		
			興穀國小	30	31		30	33		30	33		
基隆市	33454		隆聖國小	60	57		60	61		60	59		
桃園縣	177859												
新竹縣	43222												
新竹市	35898												
苗栗縣	8337		苑裡國小	30	28		30	28		30	30		
北區	810866	0.422759		220	207	0.403509	220	213	0.413592	220	216	0.413793	
台中市	99560		建功國小	70	65		70	62		70	66		
台中縣	142518												
彰化縣	111811		南郭國小	30	30		30	30		30	30		
南投縣	44225		中寮國小	30	41		30	39		30	43		
雲林縣	53901		安慶國小	30	30		30	34		30	28		
嘉義市	25343												
嘉義縣	39380												
中區	516738	0.26941		160	166	0.323587	160	165	0.319149	160	167	0.319923	
台南市	63710		安順國小	30	34		30	33		30	32		
台南縣	87351												
高雄縣	98997												
高雄市	122953		漢民國小	70	70		70	70		70	64		
屏東縣	73223												
南區	446234	0.232652		100	104	0.202729	100	103	0.199226	100	96	0.183908	
宜蘭縣	39920												
花蓮縣	29315		宜昌國小	35	29		35	31		35	29		
台東縣	18792												
東區	88027	0.045894		35	29	0.05653	35	31	0.059961	35	29	0.05556	
澎湖縣	6452												
金門縣	4108												
連江縣	573		北竿 板里國小	10	7		10	5		10	14		
離島	11133	0.005804		10	7	0.013645	10	5	0.009671	10	14	0.02682	
學生總數	1918034	1		525	513	1	525	517	1	525	522	1	

(三) 表現標準適切性評定專家的組成

在本研究中，評定學生第一第二階段的表現標準適切性的專家共有 15 位，為考量專家群體的代表性，除了教育界的先進和特定領域的專家學者們，尚包括家長、社會人士和教師組織代表等，詳如表 2 所示。

表 2 表現標準適切性評定專家諮詢小組一覽表

類別	評定專家	職別或相關背景	全體成員比例
教育相關專家學者	A1	教育相關政策和學習領域專家	20%
	A2	測驗評量和課程專家	
	A3	教師團體代表	
教學實務工作者	A4	現任教師	53.3%
	A5	現任教師	
	A6	現任教師	
	A7	現任教師	
	A8	現任教師	
	A9	現任教師	
	A10	現任教師	
	A11	現任教師	
其他領域代表	A12	家長代表	26.6%
	A13	家長代表	
	A14	社會賢達人士	
	A15	社會賢達人士	

二、執行階段

本階段主要描述標準適切性評估的步驟和執行過程。主要利用 Angoff (1971) 程序進行三回合的專家評估，第一回合的主要目的為瞭解專家的評估意見，透過專家全體和實徵研究的資料回饋，再於第二回合的評估中修改初始的評估結果，並在第三回合整理專家的評估資料，再比對能力指標中所指陳的表現標準適切性。

(一) 專家評估的實施步驟

本階段主要進行三個回合。第一回合包括專家的評估任務的訓練說明以及專家評估的進行。首先由研究者當面向專家說明研究目的、相關理論依據和專家們的任務，同時介紹測驗的發展背景和進行評估練習，同時界定基本能力，在這個過程中並且回答和澄清所有專家提出的問題。此步驟主要是希望讓專家群體對評估任務有完整和一致性的共識，盡量降低評估結果的誤差。

第二回合的實施主要是邀請專家學者們依據研究者所提供的回饋資料修正前一次的評估結果。問卷由研究者親自送交給參與的專家學者，然後再向專家學者說明資料的處理，和解釋評估問卷的進行，相關文件包括說明函和第二次評估結果問卷。

在第二回合實施完畢之後，研究者利用因素分析的方式進行專家意見共識一致性的考驗，發現專家整體的意見已趨於一致，不需進行第三回合的逐題評估，因此除了相關的結果提供給專家們參考外，在第三回合專家評估主要目的乃邀請專家進行表現標準適切性進行整體評估和審核。

(二) 實施程序的回饋資料

表現標準適切性評估的回饋資料分為兩個部分，一個是提供給專家群體作為進行第二回合評估的參考依據，以及第二回合的逐題評估結果，提供專家第三回合能力指標表現標準適切性判斷的依據；另一個部分是專家表現標準程序回饋函，以了解專家們對於實施程序中的意見。

三、後續階段

本階段主要為標準設定歷程的資料處理以及結果的分析，並提供回饋問卷，進一步

瞭解專家群體的意見和作為未來研究改進的建議，因此，將在本階段所得到的研究結果，

詳述於下。

肆、研究結果

一、專家意見的一致性分析

專家意見是否能達到一致，此為得懷術實用性的主要關鍵。除了對問卷的型式和內容的審慎設計外，為了避免獨斷和主觀的判斷，在本研究中透過因素分析的流程衡量第二次專家評估問卷的意見一致性程度。結果發現，在專家評定三年級學生的表現標準中，15 個專家的因素特徵值大於 1 者可聚斂

成 2 個共同因素，其解釋量共為 78.161%；四年級的因素分析結果也發現 15 個專家的因素分析結果特徵值大於 1 者可聚斂成 2 個共同因素，其解釋量共為 81.607%；同樣的情形也發生在五年級的因素分析結果，其 2 個因素的解釋量為 74.794%。三個年級專家群體的因素分析結果及其解釋變異量如表 3 所示。

表 3 專家群體評定三年級學生表現標準之因素分析結果

年級	因素構面	特徵值	解釋變異(%)	累積解釋變異(%)
三	因素一	10.596	70.637	70.637
	因素二	1.128	7.523	78.161
四	因素一	11.177	74.515	74.515
	因素二	1.064	7.092	81.607
五	因素一	10.137	67.582	67.582
	因素二	1.107	7.382	74.964

接下來透過因素負荷矩陣將各專家分別歸類在不同的因素構面之下，並求得各因素構面中資料的 Cronbach's α 信度係數，作為判斷專家群體意見的一致性程度依據。其結果如表 4 所示。

表 4 三、四、五年級專家群體評估結果因素構面的信度係數

因素構面	三年級	四年級	五年級
因素一	.9678 (12) *	.9666 (12)	.9644 (12)
因素二	.3959 (3)	.5957 (3)	.7910 (3)

*()中的值表示各因素構面所包含的專家數量

結果發現三個年級的第一個因素構面間的內部一致性係數均在.90 以上，屬於高信度，而且均佔全體專家人數的 12 人，顯示專家群體在第二回合的評估意見均有 80%的共識水準，所以無需再進行第三回合的逐題回饋調查，於是第三回合的問卷評估主要是以表現標準適切性評估為主。

二、表現標準適切性的評估結果

(一) 基本能力的表現和表現標準評估適切性的界定

在本研究中，邀請 15 位專家群體逐題評估若具備數學基本能力的三、四、五年級學

生，在「數與量」和「代數」的相關表現。因此當專家在進行評估判斷時，必須先釐清專家們對基本能力表現的看法。所謂的「基本」能力可能涉及到的層面或定義略為廣泛和模糊，在本研究中，首先請專家們補充其對於基本能力的一些質性敘述，例如：「能符合日常生活所需，並應用在食、衣、住、行上」、「能真正符合學生的認知發展」、「讓中下程度學生能理解」等，並且以專家認為的整體通過率來具體化基本能力的界定，在 15 位專家所評估基本能力學生的通過率平均為 80.3%。

以專家評估通過與否的數據而言，只有一位專家認為基本能力的通過率表現為 60%，其餘皆在 70%以上，且若某一試題平均評估通過率超過 70%，大約就會有 10 位專家以上贊成該題屬於基本能力的通過表現，因此，在本研究中的專家評估適切性的討論，以 70%-80%的通過率區間作為基本能力表現標準評估適切性依據。

若將每一題項和能力指標表現標準細項

進行連結，指標細項涵蓋的試題評估表現都在.70 以上的，表示專家評估這些指標細項的表現符合學生發展的適切性，則以「◎」代表該項表現標準和學生的基本能力表現是符合的；如果某一指標細項的表現標準涵蓋了多題，例如 N-1-1.3 有三個問題，而專家評估並非都在.70-.80，則以專家評估二元資料中是否多數題通過為評估適切性依據；倘若該項指標細項專家的評估並非都在.70-.80 的通過率，「○」表示以專家評估的資料而言，該項指標的表現標準適切度可再進一步討論；「×」則表示以專家評估的資料而言，該項指標細項的表現標準和學生的能力表現可能並不適切，需要考慮修改。

(二) 專家評估結果

圖 2 為專家們評估三、四、五年級學生在 80 題試題的答題反應和通過率的關係。

而表 5 則為專家評估三個年級學生的答題表現通過率統計結果。

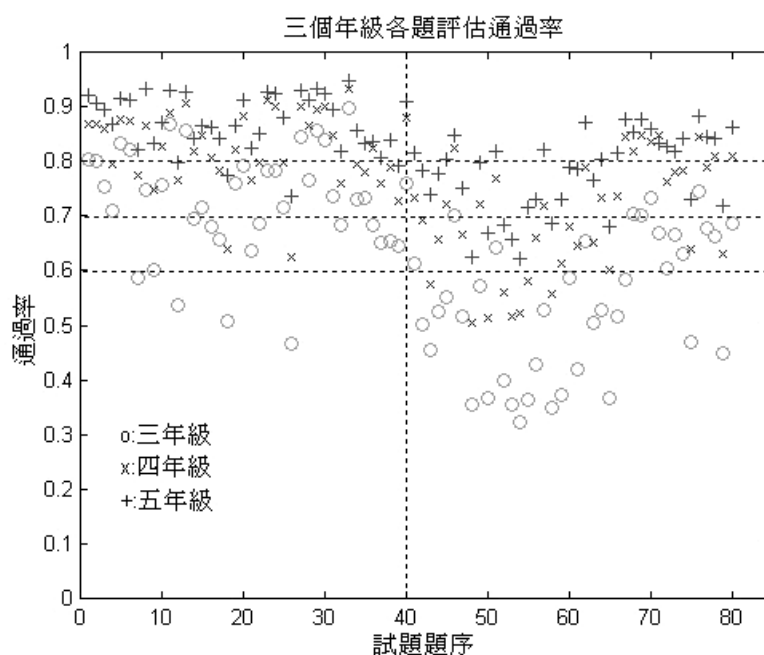


圖 2 專家評估三、四、五三年級答題通過率分布圖

表 5 專家評估三、四、五年級學生表現標準統計表

年級 (題號)		專家評估通過率						
		0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
三	1-40	0	1	3	11	15	10	0
	41-80	8	6	11	10	5	0	0
四	1-40	0	0	0	2	14	19	5
	41-80	0	0	8	11	12	9	0
五	1-40	0	0	0	0	4	20	16
	41-80	0	0	0	7	13	20	0

結果發現，就整體表現而言，專家評估三年級學生的應有的通過率表現，屬於第一階段的前 40 題，大多介於.60-.90 之間，共有 36 題，佔全部試題的 90%，不過有 4 題在.60 以下，佔全部試題的 10%；第 41-80 題，屬於第二階段的能力表現，有 5 題評估通過率超過.70，佔 12.5%。在五年級學生的評估通過率表現上，屬於第一階段的前 40 題都超過.70，而 41-80 題中屬於第二階段的題項評估表現上，大部分試題都超過.70 以上，共有 33 題，佔 82.5%，不過仍然有通過率在.70 以下，共有 7 題，佔 17.5%。

(三) 與能力指標表現標準連結的適切性評估結果

在第三回合中，同樣的由研究者先將專家的逐題評估通過率和能力指標表現標準進行連結，依照年級整理成表，請專家同時參考評估結果和不同年級的發展進行最後考量，然後再將整體評估結果中，需要專家進一步審核或判斷的題項列出，重新徵詢專家意見。

雖然本研究主要瞭解第一和第二階段學生能力指標表現標準的適切性，若從三、四、五年級學生的縱貫表現來看，將四年級學生的評估表現和三年級以及五年級學生的評估

表現進行對照，發現在第二階段中有一些指標的表現標準在四年級就可以達到.80 以上的評估通過率，例如 N-2-1.4、N-2-4.1 等，同時也有 12 位專家認為四年級學生就已經可以學會了，因此，這些指標的表現標準在第二階段結束時是否可以進行難度上的調整？或是再加深加廣？由研究者將其整理在第三次表現標準適切性評估中徵詢專家的判斷。

最後再由研究者進行最後的彙整列表和整理專家建議。表 6 為研究者所彙整專家評估數學領域能力指標暫行綱要「數與量」、「代數」學生表現標準適切性結果。

整體而言，專家們評估指標的表現標準適切性程度大致符合，在第二階段中的某些指標指陳的表現標準可能稍微簡單一些，可以再進行進一步的調整。

三、專家回饋資料分析

為能更清楚了解專家們在表現標準評定程序中的感受和意見，本研究設計了專家表現標準程序回饋函，透過專家意見的回饋，作為本次研究的檢討和未來進行相關研究的建議。

表 6 專家評估「數與量」、「代數」學生表現標準適切性結果一覽表

一、數與量

(一) 數與計算

內容標準	第一階段			第二階段		
	表現標準	題項	適切性評估	表現標準	題項	適切性評估
N-S-1	N-1-1.1	1	◎	N-2-1.1	41	◎
	N-1-1.2	2、3、4、5、6	◎	N-2-1.2	42、43	◎
				N-2-1.3	44、45	◎
	N-1-1.3	7、8、9	◎	N-2-1.4	46	×
				N-2-1.5	47	◎
N-S-2	N-1-2.1	10、11、12	◎	N-2-2.1	48	◎
	N-1-2.2	13、14	◎			
	N-1-2.3	15	◎			
	N-1-2.4	16	◎			
	N-1-2.5	17	◎	N-2-2.2	49	◎
	N-1-2.6	18	×	N-2-2.3	50	◎
	N-1-2.7	19、20	◎	N-2-2.4	51	◎
	N-1-2.8	21	◎	N-2-2.5	52	◎
				N-2-2.6	53	◎
	N-1-2.9	22	◎	N-2-2.7	54、55	◎
	N-1-2.10	23、24	◎	N-2-2.8	56、57	◎
			N-2-2.9	58	◎	
			N-2-2.10	59、60、61	◎	
			N-2-2.11	62、63	◎	
N-S-3	N-1-3.1	25	◎	N-2-3.1	64	◎
	N-1-3.2	26	×	N-2-3.2	65、66、67	◎

(二) 量與實測

內容標準	第一階段			第二階段		
	表現標準	題項	適切性評估	表現標準	題項	適切性評估
N-S-4	N-1-4.1	27、28	◎	N-2-4.1	68、69	×
	N-1-4.2	29	◎			
	N-1-4.3	30、31、32、33、 34、35、36	◎	N-2-4.2	70、71	×
				N-2-4.3	72	◎
N-S-5	N-1-5.1	37、38	◎	N-2-5.1	無題項	
	N-1-5.2	39	◎	N-2-5.2	73	◎
				N-2-5.3	74、75	◎
				N-2-5.4	76	×

二、代數

內容標準	第一階段			第二階段		
	表現標準	題項	適切性評估	表現標準	題項	適切性評估
A-S-1	無指標細項			A-2-1.1	77	◎
				A-2-1.2	78	○
				A-2-1.3	79	◎
A-S-2	A-1-2.1	40	◎	A-2-2.1	80	○

在實施程序部分，有 9 位專家認為實施程序的困難度適中，佔 60%；認為整個實施程序有點困難的有 2 位，佔 13%；認為還算簡單的有 3 位，佔 20%。讓專家們感到困擾的主要是基本能力的界定過於模糊，佔 53%，另外有 3 位專家認為整個實施程序過於複雜，同時有 2 位認為資料處理過於複雜，以及各有 1 位專家認為工作過於繁重和資料呈現不夠清楚。

由於必須要同時估計三個年級學生在 80 題試題中的表現，對專家而言相當繁重，再加上並非每一個專家實際接觸該年齡層或是有相關的教學經驗，因此有一些估測其實是相當困難的。

整個研究中，專家們相當肯定試題取材的生活化和靈活程度，提供實際學生的表現作為參考依據，並且了解學生能力的縱向發展，和專家的預期和學生實際表現之間的落差，同時也希望能提供給實際教學者參考，

能更了解學生實際表現和成人預期之間的差距。

在某些題目中四年級的通過率比五年級高，除了讓專家們感到意外之外，更有興趣探究其中的原因，同時有專家建議可以在下一年度反覆施測，以了解接受不同版本課程孩子的能力變化趨向，或是進行小學階段的全面性研究。雖然研究者在邀請專家進行評估時皆說明了研究的目的和程序，但是仍然有專家希望能更詳盡的了解相關的理論基礎，程序和目的，以及研究結果對教育可能的產生的貢獻。

整體而言，對於整個研究和實施程序專家們皆抱持肯定和支持的態度，甚至認為值得進一步探究學生的能力表現，或是進行更大範圍的研究。不過評估的工作對專家而言稍嫌繁重，這是未來進行相關研究時需要克服的部分。

伍、結論與建議

一、結論

(一) 第一階段的能力指標，大多符合三年級學生的認知表現

第一階段的 40 題評估中，15 位專家估測通過率大多在 .70 以上，未達 .60 的僅有 4 題，再就指標細項和試題進行連結，邀請專

家進行表現標準的適切性評估和審核。除了「N-1-2.6 知道加減法的逆運算關係」(第 18 題)、「N-1-3.2 推斷和檢驗整數計算結果的合理性」(第 26 題)的適切性評估為「×」外，整體而言，專家評估第一階段的學生表現大多符合學生的認知表現。

而測量「N-1-2.6」的第 18 題，專家們認

為可能是問題描述加入了代數符號，增加試題的難度，影響學生的表現，不過關於此表現指標的適切性可以再做進一步的分析和評估。「N-1-3.2 推斷和檢驗整數計算結果的合理性」為第 26 題的評估，其在五年級的通過率表現也未達.70，顯示該指標的適切性需要再進一步討論。

至於「N-1-1.3」專家們認為可能是問題的表徵形式對學生而言比較困難，影響通過率表現。但就指標而言仍應屬於第一階段學生的基本能力表現。在 N-1-2.4、N-1-2.5、N-1-2.8、N-1-2.9 等指標表現標準中也出現類似的情形，不過就第一階段所指陳的部分而言，專家評估表現標準的適切性大致符合學生的基本能力表現。

(二) 第二階段的能力指標中，某些指標表現在三年級即可達成

在第二階段 41-80 題的試題評估中，46、68、69、76 在三年級的表現評估中，總共有超過 10 位專家認為三年級就可以達成，評估的通過率也都超過.70，其依序為測量：「N-2-1.4 運用具體物、圖示和數學符號描述和比較兩位小數的大小」、「N-2-4.1 知道長度、容量、重量、面積、體積等的較大或二階的測量單位和關係（例如：千米、毫米、公升、毫公升、時、分、秒……等）」以及「N-2-5.4 發展計算長方形的面積和長方體的體積的策略」顯示以上三個指標所指陳的表現對於第二階段學生而言可能太過簡單，因此專家建議將這些指標提前到第一階段的表現標準中。

(三) 第二階段的表現標準所指陳的能力表現較為簡單

在第二階段的試題評估中，有 21 題專家評估四年級學生通過率就已經超過.70，且超

過 10 個專家認為該題項所測量的能力表現在四年級學生就可以達成，這些試題分別為 41、45、46、49、51、56、58、62、64、66、67、68、69、70、71、72、73、74、76、77、78、80，也就是說，在對應第二階段能力指標的 29 個表現標準中，有 18 個指標表現在四年級就幾乎可以達成，經由研究者再進一步分析這些試題內容細項，除了分數的理解和運算部分外，大多是屬於數學知識的理解和基本運算的認知表現，換句話說，在第二階段所指陳的學生表現中，專家們認為大多數四年級學生就可以學會，除了分數的理解和運算考量教材的邏輯性組織，還是保留在第二階段外，在第二階段能力指標所指陳的表現標準可以考慮內容難度的增加。

因此研究者再將這些試題和能力指標的表現標準進行對應，請專家在第三回合進行評估，結果發現除了「N-2-3.2 選擇和應用適當的方法解生活情境中的問題」，所指陳的表現標準涵蓋範圍較廣，難度彈性較大，因此不需調整之外，其餘指標細項如「N-2-1.4 運用具體物、圖示和數學符號描述和比較兩位小數的大小」、「N-2-4.1 知道長度、容量、重量、面積、體積等的較大或二階的測量單位和關係（例如：千米、毫米、公升、毫公升、時、分、秒……等）」、「N-2-4.2 知道日、時和分的關係，能用 24 時制來描述時間」、「N-2-5.4 發展計算長方形的面積和長方體的體積的策略」都應考量將第二階段學生的表現標準加深或加廣。

二、建議

(一) 專家表現標準評估部分的建議

1. 邀集更具代表性的專家

表現標準的評估關鍵在於專家對所執行任務的勝任程度 (Plake & Impara, 2001)，在

本研究中，因為考量專家意見的代表性，所以邀請不同領域專家和實務經驗豐富的教師組成評估小組，但因受限於資源的窘蹙，僅能以個人名義邀請專家的加入。NAEP 因為有國家政府評量委員會 (National Assessment Government Board, NAGB) 的支援，在專家的邀集上可以透過嚴謹的抽樣程序，邀請各行政區和各相關領域更具權威和代表性的專家提供建議 (Raymon & Reid, 2001)，未來的相關研究若能由國家層級進行，或許可以加入更多更具代表性和權威性的專家。

2. 試題和指標內容的連結可再加強

在本研究中，審題過程中所邀請的專家中包括日後表現標準評估的專家學者，主要即是考量試題和指標所指陳表現之間的連結，不過因為審題小組只包含所邀請 15 個表現標準評估專家中的 5 位，僅佔 $\frac{1}{3}$ 的比例，未來可以在審題過程中就邀集相關表現標準評估的專家，視為表現標準評估的程序之一，可強化試題和指標內容之間的連結。

3. 全面性及系統性的進行評估

九年一貫課程各領域能力指標所陳述的標準多半以先前所訂定的中小學課程標準 (教育部，1993) 為基礎，並根據課程發展的經驗進行修訂，大多是以經驗校準的方式由指標撰寫的專家學者來制定，以標準的高低的訂定，並沒有比較廣泛的討論和實徵性的學生資料的支持。在本研究中針對第一、二階段學生在「數與量」和「代數」兩內容主題的表現適切性進行評估，為指標的內容適切性提供實徵資料的支持，未來同樣可以考量繼續擴充縱貫和橫向的資料，瞭解九年一貫課程中一至九年級學生的能力表現評估，以及數學領域中其他內容主題的相關表現，甚至是逐年系統性的收集資料，瞭解學生的能力表現發展趨勢以及變化。

4. 不同認知能力表現的適切性評估

九年一貫課程的數學領域能力指標目前有不同版本的爭議存在，主要是對於學生的基本能力表現有不同的看法，目前研究者乃針對九年一貫課程暫行綱要中所指陳的能力表現進行評估，而預計 2005 年開始實施的修正版本在學習內容表現上呈現不同的難度和深度 (教育部，2003)，未來可以進行不同版本內容的評估，透過實徵表現和專家判斷訂定出更合宜的學生能力表現。

(二) 研究的程序部分的建議

1. 研究方法的改進和突破

表現標準設定方法雖然有效，但過程繁複，容易讓參與的專家產生不舒服的感受，這是在技術方面需要突破的部分 (Hambleton, 1996)。另外，目前國內尚未有對標準設定工作的研究，社會大眾對表現標準的基本概念仍未建立，因此本研究以目前最廣被運用和討論的 Angoff (1971) 法為核心進行，但是仍然有專家認為過程複雜，本研究的經驗或許可以作為未來相關研究的借鏡，修正或考量其他更合適的標準設定程序。

2. 召開工作坊的方式進行

在本研究中，專家的評估是以得懷術的方式，先由研究者向專家說明研究目的和程序，然後交由專家在盡可能充裕時程內完成評估，但是因為評估的工作耗力費時，若能夠以召開工作坊的方式，讓專家們在 1-2 天的時間中，能心無旁騖的專注於評估工作，將會提高評估結果的信效度，以及更迅速精確的收集資料。

3. 專家會議的召開

在本研究中，主要以得懷術的方式進行專家評估，並未召開專家討論會議，而所有的資料也是由研究者面對面拜訪專家所收集彙整而來。未來若能有更豐沛的相關資源，

可以邀集專家召開討論會議，除了可以澄清對評估任務的爭論外，更可以藉由焦點團體討論的方式讓專家相互激盪，同時整合各專家的意見。

4.分組或是利用矩陣抽樣方式進行評估，減輕專家評估負荷

在表現標準評估的過程中，需要專家逐題反覆的進行，難免讓專家產生不舒服的感覺，這也是整個表現標準程序中有待克服的部分。未來的相關研究可以考量將專家或試題進行分組甚至以矩陣抽樣方式進行 (John, Judy, & James, 1988)，以減低專家評估工作的繁重負擔。

5.基本能力的界定要更清楚明確

在本研究中，一方面因為目前各界對基本能力的爭論為達成共識，一方面因為未提供專家更深入討論基本能力界定的機會，在定義上存在著模糊的部分，若基本能力的界定不夠清楚，這將會成為評估結果誤差的主要來源之一 (Plake, Melican, & Mills, 1991) 未來相關研究進行時，可以針對基本能力的相關知識、能力和技能的表現 (knowledge、skill、ability, KSAs) 作更清楚的界定 (Plake & Impara, 2001)，提高專家評估的一致性。

6.透過討論會議進行結論的統整

本研究中僅針對試題表現邀請專家進行評估，關於評估結果的呈現和整理由研究者統整後再回饋給專家進行審核，未來可以在專家意見達成一致性共識之後，邀請專家學者召開討論會議，形成統整性的結論。

7.不同專家背景變項對評估資料影響的進一步分析

為考量專家的代表性，本研究邀集不同領域和背景的專家進行評估工作。但因為專家樣本有限，未能進一步分析是否不同領域背景的專家在估測時可能產生某些一致性的偏誤，將來若累積足夠的專家評估資料，可進行專家特性分析，避免因為不同專家背景所產生的系統性誤差而影響估測結果。

8.利用其他資料建立設定程序的信效度考驗

由於國內目前並沒有類似表現標準設定的相關實徵研究，因此本研究僅利用專家資料的一致性來作為研究程序的信度指標，未來如果可以重複進行研究，可利用不同時間專家估測的結果共識程度作為研究的效度評估 (Patrick, Plake, & Impara, 2000; Plake & Impara, 1996; Plake, Impara, & Patrick, 1999)，或是建立其他外在效標來檢驗評估結果的效度 (Goodwin, 1999)。

9.利用模糊數的概念來處理估測資料

在本研究中利用 Angoff (1971) 的程序進行，在資料的表現上是以學生是否能正確答題的二元資料 (是與否)，以及最小能力學生的通過率估測為主，兩者在邏輯概念上雖然可以相通，但卻是離散量數和連續量數兩種不同的資料形式，而在研究二研究者利用模糊技術來處理專家的評估結果，目前在標準設定的程序上尚未發展以模糊數來進行的方法，但因為人的語意變數的模糊性特質，專家們在進行估測時，也傾向以 5 人為一單位，未來或許可以利用模糊數的概念，擷取專家估測的資料特性，進行表現標準評定資料的處理。

參考文獻

- 杜佳真 (2004)。能力指標系統的重組及表現標準適切性評估之研究—以數學領域能力指標為例。國立台灣師範大學教育心理與輔導研究所博士論文，未出版，台北市。
- 張紹勳 (2000)。研究方法。台北：滄海書局。
- 教育部 (1993)。國民小學課程標準。台北：教育部。
- 教育部 (2002)。國民中小學九年一貫課程暫行綱要。台北：教育部。
- 教育部 (2003)。國民中小學九年一貫課程綱要數學學習領域。台北：教育部。
- 教育部統計處 (2003)。國民小學概況。2003 年 4 月 20 日，取自 www.edu.tw/statistics/index.htm。
- 教育測驗中心(1998)。國際學科評量 1999：數學。雪梨：教育測驗中心。
- 曾志朗 (2003)。數字會說話。科學人，22，11。
- 黃政傑 (1996)。中小學基本學力指標之綜合規劃研究。教育部教育研究委員會委託國立台灣師範大學教育研究中心研究專案，113。台北市：國立台灣師範大學教育研究中心。
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed). Washington DC: American Council on Education.
- Arrasmith, G. A., & Hambleton, R. K. (1988). *Steps for setting standards with the Angoff method*. AmherstP: University of Massachusetts. (ERIC Document Reproduction Service No.ED 299 326)
- Cizek, G. J. (1996). Standard setting guidelines. *Educational Measurement: Issue and Practice*, 13-20.
- Goodwin, L. D. (1999). Relations between observed item difficulty levels and Angoff minimum passing levels for a group of borderline examinees. *Applied measurement in education*, 12(1), 13-28.
- Hambleton, R. K. (1996). Advances in Assessment Models, Methods, and Practices. In Berliner, D. C. & Calfee, R. C. (Eds.), *Handbook of educational psychology* (pp. 899-925). New York: Simon & Schuster Macmillan.
- Hurtz G. M. & Hertz N. R. (1999). How many raters should be used for establishing cutoff scores with the Angoff method? *Educational and Psychological Measurement*, 59(6), 885-897.
- Jaeger, R. M. (1989). Certification of student competence. In Linn, R. L. (Ed.), *Educational measurement* (3rd ed) (pp.485-514). NY: Macmillan.
- Jaeger, R. M. (1991). Selection of judges for standard setting, *Educational Measurement: Issue and Practice*, 3-4.
- John, J., Judy, A., & James, C. (1988). A note on the application of multiple matrix sampling to standard setting. *Journal of Educational Measurement*, 25(2), 159-164.
- National Assessment of Educational Process (2002). Retrieved November 29, 2002, from <http://nces.ed.gov/nationsreportcard/itmrls/printproduce.asp>
- Patrick, I., Plake, S., & Impara, J. C. (2000). *Validity of item performance estimates from an Angoff standard setting study*. San Diego: the American Educational Research Association. (ERIC Document Reproduction Service No.ED 443 878)
- Plake, B. S., & Impara, J. C. (1996). *Intrajudge consistency using the Angoff standard-setting method*. San Diego: the American Educational Research Association. (ERIC Document Reproduction Service No.ED 398 251)
- Plake, B. S., & Impara, J. C. (2001). Ability of panelists to estimate item performance for a target group of candidates: An issue in judge/panelist standard setting. *Educational assessment*, 7(2), 87-97.
- Plake, B. S., Impara, J. C., & Patrick, M. I. (1999). *Validation of Angoff-bases predictions of item performance*. San Diego: the American Educational Research Association. (ERIC Document Reproduction Service No.ED 430 004)
- Plake, B. S., Melican, G. J., & Mills, C. N. (1991). Factors

influencing intrajudge consistency during standard-setting, *Educational Measurement: Issues and Practice*, 10(2), 15-16, 22, 25.

Raymond, M. R. & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In Cizek, G. J. (Ed.), *Setting performance standards* (pp.119-157). NJ: Lawrence Erlbaum Associates.

Reckase, M. D. (1998). Converting boundaries between National Assessment Governing Board performance categories to point on the National Assessment of Educational Progress Score Scale: The 1996 Science NAEP Process. *Applied measurement in education*, 11(1), 9-21.

Reckase, M. D. (2000). *The ACT/NAGB standard setting process: how 'Modify' does it have before it is no longer a Modified-Angoff process?* LA: the National Council on Measurement in Education. (ERIC Document Reproduction Service No.ED 442 825).

Trends in International Mathematics and Science Study (1995). *TIMSS Mathematics items: Released set for population 1* (Third and Fourth grades).

Trends in International Mathematics and Science Study (1999). *TIMSS 1999 Assessment-8th Grade Mathematics*.

作者簡介

杜佳真，中華技術學院通識教育中心，助理教授

Chia-Chen Tu is an assistant professor of Center for General Education, China Institute of Technology, Taipei, Taiwan. E-mail: tucc00@seed.net.tw

林世華，國立臺灣師範大學教育心理與輔導學系，教授

Sieh-Hwa Lin is a professor of Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan.

收稿日期：94.09.06

修正日期：95.04.17

95.06.06

接受日期：95.06.30

附錄 指標細項詳表

(一) 數與計算

內容標準	第一階段			第二階段		
	表現標準	題 項	對應數學領域 能力指標 (民 91)	表現標準	題 項	對應數學領域 能力指標 (民 91)
N-S-1 學生能理解數的表徵和數的系統以及數的關係	N-1-1.1 運用口述、圖示或文字等形式描述 2000 以內的數。	1		N-2-1.1 運用具體物、圖示和數學符號描述 100000 以內的數。	41	
	N-1-1.2 運用具體物、圖示或文字描述和比較 20 以內單位分數的大小。	2、3、4、5、 6		N-2-1.2 運用具體物、圖示和數學符號描述和比較假分數及帶分數的大小。	42、43	
				N-2-1.3 辨識和推演出不同等值型式的分數。(例如 $2/8=1/4$ 。)	44、45	
	N-1-1.3 運用具體物、圖示或文字描述和比較一位小數的大小。	7、8、9		N-2-1.4 運用具體物、圖示和數學符號描述和比較兩位小數的大小。	46	
			N-2-1.5 推斷數列的規則。	47		
N-S-2 學生能利用數的性質和關係進行運算	N-1-2.1 知道加、減、乘、除運算法的特性，並運用數學的方式呈現。	10、11、12		N-2-2.1 流暢的進行整數加減乘除的運算。	48	
	N-1-2.2 發展三位數加減法運算的流暢性。	13、14				
	N-1-2.3 流暢的進行積 ≤ 100 乘法的計算。(例如：單位數 ≤ 12 ，單位量 ≤ 15)。	15				
	N-1-2.4 流暢的進行除法的運算。	16				
	N-1-2.5 以 10、100 或 1000 為單位來描述 2000 以內的數的位值換算。	17		N-2-2.2 運用直式算則型式進行加減運算	49	
	N-1-2.6 知道加減法的逆運算關係。	18		N-2-2.3 知道乘除法的逆運算關係。	50	
	N-1-2.7 發展和運用加法交換律的策略。	19、20		N-2-2.4 運用加減法的逆運算關係檢驗計算結果。	51	

(附錄 續)

	N-1-2.8 知道概數的使用時機。	21		N-2-2.5 運用不同的方式描述概數(例如四捨五入、捨去和進入)。	52
				N-2-2.6 發展估算整數運算結果的策略。	53
	N-1-2.9 流暢的進行同分母真分數(分母在 20 以內)的加減運算。	22		N-2-2.7 流暢的進行假分數或帶分數的加、減和整數倍乘法運算。	54、55
	N-1-2.10 流暢的進行一位小數的加減運算(和及被減數<1)。	23、24		N-2-2.8 流暢的進行兩位小數的加、減和整數倍乘法運算。	56、57
				N-2-2.9 指出分數和小數在數線上的位置關係。	58
				N-2-2.10 應用併式的規則簡化運算。	59、60、61
				N-2-2.11 發展和應用乘法分配律、交換律和結合律進行演算。	62、63
N-S-3 學生能應用適當的方法或工具進行解題	N-1-3.1 選擇和應用適當的工具檢驗運算的結果，如電算器。	25		N-2-3.1 選擇和應用適當的方法或工具進行運算，如電算器、電腦。	64
	N-1-3.2 推斷和檢驗整數計算結果的合理性。	26		N-2-3.2 選擇和應用適當的方法解生活情境中的問題。	65、66、67

(二) 量與實測

內容標準	第一階段			第二階段		
	表現標準	題 項	適切性評估	表現標準	題 項	適切性評估
N-S-4 學生能了解測量單位及屬性	N-1-4.1 知道長度、容量、重量、角度、面積、體積等日常生活中的測量單位(例如米、厘米、分公升、千克、克和度)	27、28		N-2-4.1 知道長度、容量、重量、面積、體積等的較大或二階的測量單位和關係(例如：千米、毫米、公升、毫公升、時、分、秒……等)。	68、69	
	N-1-4.2 用直接比對或複製的方式比較物體的測量特徵(長度、重量等)。	29				

(附錄 續)

	N-1-4.3 透過看日曆、月曆、時鐘或先後順序來描述時間。	30、31、32、33、34、35、36	N-2-4.2 知道日、時和分的關係，能用24時制來描述時間。	70、71
			N-2-4.3 運用固定時間或固定距離的方式描述物體的運動狀態。	72
N-S-5 學生能應用適當的方法、工具或公式來進行測量	N-1-5.1 選擇和應用日常生活中的測量單位(例如米、厘米、分公升、千克、克和度)進行長度、重量、容量、角度的測量。	37、38	N-2-5.1 選擇和應用較大或二階單位(例如：千米、毫米、公升、毫公升、時、分、秒……等)進行長度、重量、容量、角度的測量。	無
	N-1-5.2 選擇日常生活中的測量參考架構來估測長度、重量或質量或時間。	39	N-2-5.2 在公制系統下發展適當的方法或工具進行長度、重量、質量或時間的估測。	73
			N-2-5.3 在公制系統下，計算和換算長度、重量、質量、時間的測量。	74、75
			N-2-5.4 發展計算長方形的面積和長方體的體積的策略。	76

二、代數

內容標準	第一階段			第二階段		
	表現標準	題 項	適切性評估	表現標準	題 項	適切性評估
A-S-1 學生能理解代數的表徵形式和關係	無			A-2-1.1 運用圖畫、具體物或文字符號(如□、△、甲、乙)呈現和描述數量關係。	77	
				A-2-1.2 用中文簡記式表示常用的公式。	78	
				A-2-1.3 推斷數量關係及特性。	79	
A-S-2 學生進行代數和方程式的運算	A-1-2.1 解算式填充題(如： $\square+3=8$)中的未知數。	40		A-2-2.1 列出並解算式填充題中的未知數。	80	

Evaluating the Applicability of Mathematics Performance Standards with a Competence Indicator System

Chia-Chen Tu

Center for General Education, China Institute of
Technology

Sieh-Hwa Lin

Department of Educational Psychology and
Counseling, National Taiwan Normal University

Abstract

The purpose of this study is to evaluate the Taiwanese students' performance in math, specifically in regard to number/quantity and algebra. In order to complete the study the researchers used the Delphi method to set the performance standards. The process of setting performance standards consisted of three stages. Stage 1, the preparatory stage, where as in essential background information about the exam which the participants required to take was provided by 15 experts. Stage 2 was the modified stage, those experts modified their evaluations according to the students' performance test. In stage 3, the final stage, the results of the evaluations were analyzed by the experts. The three separate evaluations demonstrated that the stage-one competence indicator was correlated with existing data we collected from our participants. In contrast, the result of the stage-two competence indicator did not correlate so closely with the performance standard.

Keywords: performance standard setting procedure, Delphi method, mathematical performance standards

