

TIMSS八年級與國中基測物理試題 認知成份之探討

盧思丞 / 涂柏原

TIMSS八年級與國中基測物理試題認知成份之探討

盧思丞

台南大學測驗統計所博士生

涂柏原

台南大學測驗統計所副教授

摘要

本文統整了TIMSS科學認知架構與文獻中對於認知成份研究所得到的結果，以總字數、認知需求層次與解題所需概念數三個成份做為認知成份來描述認知成份與TIMSS八年級與國民中學基本學力測驗物理試題難度之間的關係；同時藉由不同能力水準考生在不同試題難度水準上之答對率，說明各種能力水準考生與所精熟的認知需求層次之間的關係，以提供教學與評量的具體參考資料。研究使用TIMSS八年級物理試題共69題，與國民中學基本學力測驗物理試題75題試題，進行認知成份分析，發現三個認知成份能夠解釋TIMSS八年級及國中基測物理試題77%以上的難度變異。

關鍵字：認知成分、TIMSS、國中基測

The cognitive component analysis of TIMSS's and BCTest's physics items

Szu-Cheng Lu

Ph. D Student, Graduate Institute of Measurement & Statistics National University of Tainan
cheni2345ster@gmail.com

Bor-Yaun Twu

Associate Professor, Graduate Institute of Measurement & Statistics National University of Tainan

Abstract

Based on the TIMSS science framework and the literature perspectives, the study used three cognitive components, including number of words, cognitive level, and number of science concepts, to explain the variation of item difficulty. Sixty-nine items of TIMSS science assessment and 75 physics items from the BCTEST were used for this analysis. It was found that 77% of item difficulty variance was explained by the three cognitive components for both TIMSS and BCTEST. And for the students with different ability levels did perform differently on the items which called for different cognitive requirements. The more difficult items did require more advanced cognitive abilities in order to give correct responses.

Keywords: cognitive components, TIMSS, BCTEST

壹、緒論

一、研究動機

為了在國際比較的脈絡中瞭解科學教育的成效，台灣目前參加兩種國際評比：國際數學與科學教育成就趨勢調查（Trends in International Mathematics and Science Study, TIMSS）和學生能力國際評量計畫（The Programme for International Student Assessment, PISA）。其中，TIMSS偏重學科知識學習成就的評量，包含數學和科學兩個內容，調查對象為四年級和八年級兩個年級的學生（Olson, Martin, & Mullis, 2008）。

TIMSS成就測驗是在知識內容與認知能力二維架構下設計發展的，八年級科學的知識內容包含生命科學、化學、物理、地球科學等四個內容領域，四個內容領域各自又區分為三至六不等的若干主題；科學的認知能力則分為知識、應用與推理三個認知領域（Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberber, 2005）。其目的在於瞭解各國學生是否學會了學校科學與數學課程中預定的學習目標，因此其成就測驗基本上是以學校學科知識內容為架構設計的（Martin, Mullis, Gonzalez, Gregory, Smith, & Chrostowski, 2000）。對TIMSS八年級的試題來說，在49個國家中有80%的國家的課程涵蓋率高於70%，而台灣的涵蓋率更高達91%（Martin, Mullis, & Foy, 2008）。

國中生基本學力測驗（以下簡稱國中基測）是一種學科成就取向的測驗，測驗內容包含國文、數學、社會、英文與自然五個科目。其中自然科內容領域含括物質與能、生命世界、地球環境、生態保育、資訊科技等的學習、注重科學及科學研究知能等，主要在評量學生的科學知識與技能，評量對象為國內九年級的學生。

TIMSS與國中基測之評量目的基本上是相同的，都是評量科學學習成就。TIMSS的題型包含選擇與問答兩部分，比較能評量出學生較完整的科學知識，不被選擇題的選項所局限，而國中基測僅有選擇題。相較於TIMSS的問答題，國中基測係以題組的形式來評量學生對於同一主題的不同概念，雖受限於選擇題的題型，效能可能不如問答題，但仍可以測驗出學生的科學知識與閱讀理解能力。雖然TIMSS與國中基測仍有些差異存在，二者的試題之主軸與核心概念方面仍有許多相同之處，二者皆致力於以科學的方法發展更創新的測驗工具，以有效評量學生對科學概念的理解與應用，並幫助教師或研究者更加認識學生的科學學習成就，且施測的對象年齡相近，約為15到16歲的學生。

台灣學者有關TIMSS的研究，探討的主題涵蓋了：1.了解學生數學及科學學習成就，提供規劃數學及科學課程之參考（張殷榮，2001；林碧珍、蔡文煥，2003）；2.了解學生數學及科學學習成就與家庭背景、學習環境、教師因素等影響因子的關係，並作國際比較分析（張殷榮，2001）；3.與TIMSS 1999及TIMSS 2003之結果作比較；4.了解學生在數學及科學學習成就之趨勢（羅珮華，2000；林鳳女，2000）；

4.了解國際上評量學生學習成就的趨勢與新的評量方法，提供教師參考（李濟國，2001；陳竹村，2003）；5.提昇研究人員資料分析研究能力（劉春初，2004）。

國中基測的研究則大致上包含以下四類：1.實施成效以及計分方式（涂柏原，2007；余民寧、賴姿伶、劉育如，2005）；2.試題品質與造成試題差異功能之原因（盧雪梅、毛國楠，2008）；3.影響國中基測分數的原因，包括學校因素、同儕群體因素及個人和家庭背景特徵因素等（陳吉仲、郭曉怡、李佩倫，2007）4.國中基測分數與未來表現之相關及預測情形，對教師教學之建議（宋曜廷、許福元、曾芬蘭、蔣莉蘋、孫維民，2007）。

以TIMSS或國中基測當前的研究趨勢來看，主要仍停留在學生、教師、或學校相關的背景變項上之研究，探討其對學生學習成就之影響，對於試題認知成份對試題難度之影響等主題，尚未有相關的研究產出。再者，在認知成份分析領域中，大多研究仍將焦點至於閱讀或數學，且偏重以題幹字數或選項字數為研究之核心，似乎過於強調閱讀負荷量對試題難度之影響。部分學者也曾研究科學認知成份之特徵（Enright, Allen, & Kim, 1993; Rosca, 2004; Yepes-Baraya, 1997），提出一些影響科學試題難度、或受試者成就表現的認知特徵，例如題幹或選項字數、內容領域、誘答項、以及認知歷程……等等，在在顯示當試題所使用的認知特徵不同時，會造成試題難度有所差異，Yepes-Baraya（1997）就將這些特徵稱為「認知成份」，因此認知成份與試題難度之間的關係是一個值得探討的議題。Leong（2006）認為控制試題難度不僅可以有效的測量到測驗本身所要測量的構念，還能減少因為試題較困難或較容易時，導致學生的分佈情形有偏態產生，並且可降低每個試題都進行預試的成本。

Enright、Allen和Kim（1993）進行難度成份分析，指出難度成份有文本與選項特徵、認知需求、知識層次、與科學教育者對試題難度的判斷四類，難度成份可解釋52%的難度變異量。Rosca（2004）使用17個難度成份對1999年TIMSS科學領域104題選擇題，進行難度成份分析，發現可讀性指標、Bloom的認知層次、誘答項的平均字數、答案字數與平均誘答項字數的比率等四個難度成份可解釋的變異量約28%，加上圖表的呈現，五個難度成份約可解釋30%的變異量。張銘秋等人（2010）分析PISA 2006的103題科學試題，結果顯示以知識類別的數量、知識層次、科學能力及字數四個成份預測科學素養試題難度，其認知模式可解釋52%的試題難度變異。

綜上所述可知認知成份與試題難度間的關係極為密切，且研究也顯示認知成份可以解釋難度的變異情形（Enright, Allen, & Kim, 1993; Rosca, 2004; 張銘秋、謝秀月、徐秋月，2010），因此對教育者而言是不容忽視的。若事先知道哪些認知成份會影響試題難度的變異，則可針對這些認知成份對學生進行教學，進而改善學生的成就表現。由於TIMSS與國中基測都屬成就測驗，且施測的對象年齡相近，因此研究者採用兩種大型測驗探討認知成份對試題難度之解釋變異情形，並比較兩種大型測驗的認知成份模式是否有差異，期待能歸納出科學試題的認知成份，以幫助教師瞭解學生所需具備的認知需求，提升學生的學習效果和成就。

二、研究目的與問題

本研究以TIMSS八年級與國中基測驗科學試題為主，因科學試題面向過於廣泛，故筆者只以物理的試題為研究對象，使用TIMSS八年級1999到2007年三個年度的物理試題，以及國中基測2001年到2003年三個年度的物理試題，進行認知成份分析研究。藉由試題的成份特徵和試題難度之間的關係，探討TIMSS與國中基測物理試題認知成份模式，並評估認知成份對試題難度變異的解釋力，針對由TIMSS和國中基測所得到的結果，進行比較。根據研究目的，主要的研究問題如下：

1. 依據TIMSS八年級物理測驗內容架構，國中基測物理試題的分配特徵為何？
2. 本研究所提之認知成份分析架構對TIMSS八年級與國中基測物理試題難度變異解釋力為何？
3. 依據認知成份架構，不同物理能力水準的學生其認知運作差異為何？

貳、文獻探討

一、認知成份分析之重要

依據認知取向編製測驗是目前評量趨勢，雖然現代測驗理論對測驗實務有許多影響，但心理計量和認知理論的連結仍然不大（Embretson, 1993）。認知心理學重視變異來源與理論，忽略個別差異，鮮少提及測量穩定性與精確性；而心理計量則重視個別差異，講求測量精確性和穩定性，缺乏效度驗證。因此結合兩取向之優點進行試題認知成份分析，將有很大的學術產出性（Dimitrov & Raykov, 2003; Embretson & Gorin, 2001）。

以往的認知心理學家強調題目的解題歷程，忽略題目內容對試題難度的變異來源，也未探討題目內容之變異與解題歷程的交互作用（丁振豐，1995）；林世華、葉嘉惠（1999）指出認知心理學強調解題歷程，題目是由其包含的認知成份所組成，因此測量的基本單位不是題目，而是所涉及的認知成份。當測量的單位愈詳細，試題所能提供的訊息量也愈豐富；藉由認知成份分析，可協助區辨測驗架構缺點，提供不同形式試題與不同測驗間比較的基礎，對系統性的發展測驗頗有助益（張銘秋、謝秀月、徐秋月，2010）。

洪碧霞、林素微和林娟如（2006）指出認知科學研究拓廣了學習評量內涵的視野，而測量方法學的進步提昇多元而複雜學生表現的可解釋性。美國評量基礎委員會（the Committee on the Foundations of Assessment）提出一項評量的三角架構，來統整學生在學習領域表徵知識和發展能力的認知模式、用以觀察學生表現的作業或情境設計、及針對學生表現資訊進行推論的解釋方法等三者之間的關係。認知與學習模式具有比以往更重的比重，藉由對學生成就目標與學習進展內涵性質之描述，以協助學生在學校進行成功的學習（Pellegrino, Chudowsky, & Graser, 2001）。Dimitrov and Raykov（2003）認為一個試題難度預測模式中認知與程序操作的知識，能夠允許

測驗發展者（1）在試題施測之前，編製已知難度的試題；（2）避免在研究團體進行個別試題的預試，以減少開支成本；（3）使試題難度與學生的能力值配合一致（match）；（4）發展以特定認知和處理特徵為目標的教學策略。

綜合上面的說法，可歸納出使用認知成份的七個優點：（1）藉由認知模式解釋試題難度來源，以便預測新試題之參數；（2）以認知模式預測試題特徵，所以不需對每個試題進行預試；（3）可說明試題的構念效度，且試題的意義與解題歷程的認知模式相互關聯；（4）藉由試題特徵（難度）的重新組合產生新試題於電腦化測驗中；（5）針對受試者能力配對產生的試題難度，促進電腦化適性測驗的發展；（6）不需考慮安全性問題，因為系統中只有認知因素是可知的；（7）可針對認知及處理的特徵，協助教師理解學生學習表現的認知過程，發展更有效的教學策略（Dimitrov & Raykov, 2003; Dimitrov, 2007; Embretson & Daniel, 2008; Embretson & Gorin, 2001）。

因此，倘若對於認知成份瞭解越多，對於測驗編製者而言，可省下繁複、費時的工作，即可操弄所編寫的試題之難易度；對於教學者而言，可更明確知道什麼對學生而言是困難的，進而發展更好的教學策略。故本研究希望能透過認知成份分析將影響測驗試題難度的成份分解出來，以提供教師或測驗編製者的參考。

二、科學難度成份的相關研究

Yepes-Baraya（1996）使用1993年NAEP科學評量研究學生接受測驗時所使用的認知程序。Yepes-Baraya將難度成份分為認知特徵與試題特徵兩種，其中認知特徵又可分為內容知識、推理和說明、以及假設公式和測試；而試題特徵可分為程序比喻資訊及試題形式和閱讀難度。在這些內容下，各含有4~11個試題特徵，共36個試題特徵。與考生面談進行放聲思考，分析答案藉以確定該測驗確實有測得其評量所欲測量的構念。

後來Yepes-Baraya（1997）使用1996年NAEP科學評量所有試題的資料進行分析，包含4年級、8年級及12年級等共三個年級45個試題區塊（block），總共超過500題試題，每個年級有將近2500名受試者作答。這個研究中所使用的難度成份與Yepes-Baraya（1996）的不同，包含了內容知識、含內容與解釋的推理、假設形成與考驗、處理圖表訊息、試題格式和閱讀難度、與實作任務的處理技巧等六類。在這六類之下，又分別含有4至11個特徵，共38個特徵。然而，作者並未使用統計方法來考驗難度成份的有效性，但是作者指出「含內容與解釋的推理」是最重要的特徵。

Enright、Allen和Kim（1993）使用1985年與1986年44題NAEP生命科學的選擇題來進行難度成份分析，他們將難度成份分為四類：文本與選項特徵、認知需求、知識層次、與科學教育者對試題難度的判斷。他們的研究結果顯示這四個難度成份可以解釋52%的變異情形，若結合試題特徵與科學教育者對試題難度的判斷，則可以提升7%至15%的解釋力。在個別難度成份的貢獻上，知識層次的解釋力高達38%。

Rosca (2004) 使用1999年TIMSS科學領域的104題選擇題進行分析，其難度成份包含題幹的字數、題幹中的句子數、每一段的字數、每一句的字數、圖表的呈現、Bloom的認知層次、TIMSS的表現層次、TIMSS的科學內容分類、答案的字數、誘答項的平均字數……等共17種。Rosca採多元迴歸與線性邏輯斯模式 (Linear Logistic Test Model, LLTM) 進行分析，發現可讀性指標、Bloom的認知層次、誘答項的平均字數以及答案字數與平均誘答項字數的比率等四個難度成份可解釋的變異量約28%，加上圖表的呈現只可增加解釋約1%的變異量，五個難度成份約可解釋30%的變異量。

涂柏原、梁恩琪、翁大德與楊毅立 (2004) 針對2001到2003年國中基測自然科學試題共345題，進行試題分析，提出影響自然成就表現的難度因素有五個，分別為試題資訊量、選項異質性、圖文推測程度、轉化程度與知識量。其中知識量表示試題本身所需具備的概念基礎多寡，當所需的概念基礎愈多，試題愈難。

由上面的描述可歸納得到影響試題難度的成份包含了題幹字數、各誘答項字數 (或是誘答項平均字數)、可讀性指標、認知層次與圖表呈現、試題資訊量、選項異質性、圖文推測程度、轉化程度與知識量等。因為本研究使用的TIMSS八年級物理試題包含了選擇題和問答題兩種，而國中基測物理試題只有選擇題，故僅以各試題的總字數、認知需求層次、以及解題所需概念數三者，作為本研究進行分析時所用的認知成份。

參、研究方法

一、認知成份分析架構

本研究所用的認知成份係參考Rosca (2004)、Yepes-Baraya (1997) 與Enright 等人 (1993) 之難度成份架構、以及涂柏原等人 (2004) 對國中基測科學試題的分析結果而得到的，認知成份包含各試題的總字數、認知需求層次、以及解題所需概念數等三者。其中認知需求層次包含成七個層次，包含了解簡單訊息、了解複雜訊息、應用科學原則解決問題、科學公式解決問題、解釋現象、辨識熟悉的實驗控制、與設計調查等，整個架構如表1所示。

表1 本研究之認知成份分析架構表

成份	變項的內容	編碼的值
總字數	1-n	1-n
認知需求層次	1.了解簡單訊息 2.了解複雜訊息 3.應用科學原則解決問題 4.科學公式解決問題 5.解釋現象 6.辨識熟悉的實驗控制 7.設計調查	0 1 2 3 4 5 6
解題所需概念數	根據解題所需的觀念數而來	0-k

註：研究自行整理歸納得到。n代表試題的總字數，k為所需要的概念數，若為常識則概念數編碼為0。

認知需求七個層次的內容說明如下：

1. 了解簡單訊息：知道科學事實、關係、程序、概念所代表的意義，對於科學名詞、所用單位、測量方法、科學儀器和設備有所了解。
2. 了解複雜訊息：了解物質之特質，科學方法間之相同或相異處，進而比較差異，並對於物質的一些特質和特徵進行區辨、排序。
3. 應用科學原則解決問題：運用圖表或模式來解釋科學概念、原則、結構、關係、程序或科學中的循環（如電子電路），使用基本的科學原則或相關知識，進行觀察與推論。
4. 科學公式解決問題：解釋文章、表格與圖片中與科學概念和原則有關的資訊，並使用科學公式找到量化的解決方法，此步驟涉及科學原則的直接應用或證明。
5. 解釋現象：利用一些科學原則、概念和理論，對觀察的或自然的現象提出合理的解釋、或科學驗證。
6. 辨識熟悉的實驗控制：分析問題以決定相關概念和問題解決步驟，發展與解釋解題策略，且考慮不同因素或概念所造成的影響；利用科學資訊、問題解決經驗、物理條件之改變，對科學實驗進行預測、並進一步檢驗研究假設，獲得證明或證據。
7. 設計調查：能自行設計合適的科學探究活動，藉以回答科學問題或檢驗假設，並自行操控影響實驗結果的變項或程序，探討在不同情境下，可能產生的實驗效果。

二、研究的資料

TIMSS的 1999、2003與2007年等三個年度八年級物理試題根據TIMSS技術報告（Martin, Mullis, & Foy, 2008）將主要分成六項內容領域，分別為物理現象（包含物質分類與特性、物質的物理狀態與變化）、能量轉換（例如熱的傳播與溫度）、光與聲音、電與磁力、及力與運動。以台灣來說，TIMSS所公開釋出的試題共有69題，筆者根據上述的認知需求層次以及六項內容領域將所有69個試題歸納整理得到如表2的雙向細目表。這69個試題包括1999年的18題、2003年的25題和2007年的26題，題型包含選擇題與問答題兩種。而台灣在TIMSS 1999、2003與2007年三個年度的測驗中，參與測驗的學生人數分別為5889人、5379人與4046人。國中基測資料來自2007到2009年的自然試題，因研究之需要筆者只取物理試題共75題，三個年度分別為23題、25題與27題，題型均為選擇題。筆者亦依照與TIMSS相同的分類方式，得到表3的國中基測試題之雙向細目表。國中基測試題資料是由「國中基測委員會」所提供，每個年度每試均隨機抽取5000人，三年度總共為30000名考生。

表2 TIMSS物理試題之雙向細目表

	物理現象	能量轉換	光	聲	電與磁	力與運動	總題數
了解簡單訊息	1	2	2	0	1	4	10
了解複雜訊息	2	1	4	0	4	1	12
應用科學原則解決問題	2	4	3	0	0	4	13
科學公式解決問題	6	0	1	1	1	4	13
解釋現象	4	2	4	1	2	0	13
辨識熟悉的實驗控制	2	0	0	1	0	1	4
設計調查	3	0	0	0	1	0	4
總題數	20	9	14	3	9	14	69

註：研究者自行整理歸納得到，橫列是認知需求層次，縱行是物理內容領域。

表3 國中基測物理試題之雙向細目表

	物理現象	能量轉換	光	聲	電與磁	力與運動	總題數
了解簡單訊息	4	0	2	0	1	0	7
了解複雜訊息	2	1	1	3	1	2	10
應用科學原則解決問題	6	0	2	2	8	6	24
科學公式解決問題	2	3	0	0	3	8	16
解釋現象	2	2	0	0	2	3	9
辨識熟悉的實驗控制	1	1	2	0	5	0	9
設計調查	0	0	0	0	0	0	0
總題數	17	7	7	5	20	19	75

註：研究者自行整理歸納得，橫列是認知需求層次，縱行是物理內容領域。

三、認知成份編碼說明

研究中所用的三個認知成份分別為試題總字數、認知需求層次、與解題所需概念數等三者，其意義及說明如下：

- 1.字數：包含題幹與每個選項的總字數，若有圖、表，則圖表內的字數也列入計算。
- 2.認知需求：認知需求層次共有七個層次，包括了解簡單訊息、了解複雜訊息、應用科學原則解決問題、科學公式解決問題、解釋現象、辨識熟悉的實驗控制與設計調查。
- 3.解題所需概念數：在解決科學問題時，學生會使用到一些科學概念，研究者根據試題的內容、敘述，歸納出學生需要的科學概念，包含0個科學概念（一般知識，不需科學概念解題）到4個以上的科學概念。

筆者根據表1所列編碼的值，對研究中所探討的每一個試題進行這三個認知成份的編碼。總字數這個成份直接以試題之總字數表示之；認知需求部分，因為是有次序性的，所以依照0到6的編碼方式將其進行編碼，故了解簡單訊息編碼為0，了解複雜訊息編碼為1，應用科學原則解決問題編碼為2，科學公式解決問題編碼為3，解釋現象編碼為4，辨識熟悉的實驗控制編碼為5，以及設計調查編碼為6；解題所需概念數中，分0個科學概念到4個以上的科學概念。另外，試題難度部分，則是經由單參數試題反應理論求得（即Rasch model），故試題難度之範圍大致上為-4.0到4.0之

間，當難度值為負值時表試題趨於簡單，難度值為正值時表試題趨於困難，而難度值接近0表試題難度適中。

實際編碼的方式，以底下四個TIMSS的物理試題作為編碼示例（見圖1至圖4）。圖1是難度值小於-1.0的範例試題，總字數50個字；難度值-1.206；主要是評量學生是否能根據題幹的說明，找出有作功的圖示，因為已經給作功的定義，所以學生不用再另外尋找相關的科學概念或知識，所以屬於認知需求層次的第一個層次「了解複雜訊息」，故編碼為0，其解題所需的概念數則編碼為0。

範例一 (b = -1.206, 50字, 層次一-了解複雜訊息, 0個概念)	
當一個物體沿著作用力的方向移動時，這就是對它作了功。有一個人作了下圖所示的各種活動，在哪一個圖中的人有作功？	

圖1 難度值小於-1.0之試題範例

圖2是難度值介於-1.0~0.0的範例試題，該題的總字數為34個字，難度值為-.768；以題幹來看，是在評量學生是否能夠比較不同顏色之間對光的反射差異，因此在認知需求層次分類是屬於第二個層次「了解複雜訊息」，故編碼為1；而解題所需的概念則是色光的反射概念，只需要一個解題概念，所以編碼為1。

範例二 (b = -0.768, 34字, 層次二-了解複雜訊息, 1個概念)	
欲將建築物的牆壁塗上油漆以盡量反射光線，應該用什麼顏色的油漆？	
A.白 B.紅 C.黑 D.粉紅	

圖2 難度值介於-1.0~0.0之試題範例

圖3則是難度值介於0.0~1.0的範例試題，總字數為52個字；難度值為0.553；以題幹來看，是希望學生利用圖表資料、與科學概念或原則解決問題，並使用公式獲得一個量化的答案，所以研究者將之歸類為第四個層次「科學公式解決問題」，故編碼為3；而解題所需的概念包含兩個，一為距離和時間，二為速率，所以編碼為2。

範例三、(b = 0.553, 52字, 層次四-科學公式解決問題, 2個概念)	
瑪莉騎腳踏車的途中發生爆胎，她修補後立刻繼續向前騎。下圖表示她的整個騎車過程。瑪莉花多少時間修補車胎呢？	
A. 20分鐘 B. 30分鐘 C. 40分鐘 D. 70分鐘	

圖3 難度值介於0.0~1.0之試題範例

範例四 (b = 1.854, 308字, 層次六-辨識熟悉的實驗控制, 4個概念)				
某個科學班級要找出裝有汽水的罐子的密度。全班分為四組來完成這項工作。每組各取得一罐汽水。當各組完成任務後, 他們發表的實驗結果如下表所示。				
	A組	B組	C組	D組
密度(g/mL)	1.04	0.04	2.77	1.05
全班同學很驚訝, 每組對罐子的密度所測得的結果都不同。他們檢視每組所用來測汽水罐質量和體積的方法。表2表示每組如何測得汽水罐的體積。				
表2: 體積				
組別	方法			體積(ml)
A	我們把燒杯加水到1400 mL的刻度。我們把沒打開的罐子放入燒杯裡。罐子沉下去。然後水位到達1776 mL。			376.00
B	我們把燒杯加水到1400 mL的刻度。我們把空罐開口朝下, 垂直放入燒杯裡。我們用一枝鉛筆把罐子壓入水中。然後水位到達1776 mL。			376.00
C	我們把燒杯加水到1600 mL的刻度。我們把空罐放到燒杯裡, 開口朝上。我們把罐子壓入水中, 看到有氣泡從罐子裡冒出來。當不再有氣泡冒出時, 罐子沉到燒杯底, 此時水位到達1605 mL。			5.00
D	我們把罐子打開, 用量筒測量罐中汽水的體積。			371.00
B和C兩組試著測量汽水空罐的體積。請解釋為什麼他們的結果不同。				

圖4 難度值大於1.0之試題範例

圖4範例試題的總字數為308個字；難度值為1.854；根據試題說明，所要評量的是測量物體體積之方法，以及對於物體質量、密度和體積間關係之瞭解程度，學生必須先知道物體之質量、密度和體積三者之間的關係，並判斷何者是正確的測量方法，以及當實驗結果不同時，是何種因素造成的。B組和C組最大的差異是空罐放入燒杯的方式（一個瓶口朝上、另一個瓶口朝下），所以在計算體積時，B組的學生會將瓶子內的空氣體積也計算進去，而C組則只計算了鐵的部分；因此，是在分析問題以決定解決步驟時，考慮不同因素或概念所造成的影響，且利用問題解決經驗、條件之改變，對科學實驗進行預測、進一步獲得證明或證據，故屬於認知需求中的第六個層次「辨識熟悉的實驗控制」，故編碼為5；解決問題所需概念為四個，包含體積、質量、密度、以及三者間的關係（質量/體積＝密度），故編碼為4。

四、資料分析

本研究根據筆者所自行得到的TIMSS八年級與國中基測之物理試題的內容架構，並以多元迴歸方法進行分析，探討總字數、認知需求層次與解題所需概念數，對由Rasch模式所估計得到的試題難度（b值）之解釋力為何，並進一步說明各種能力水準的學生在各種難度水準試題上面的表現情形以及試題之特性。

肆、研究結果與討論

一、物理試題認知成份編碼一致性

由於認知需求層次與解題所需概念數的編碼會受主觀判斷影響，導致不同的人有不一樣的編碼，為確定本研究的編碼是合理的，因此筆者除自行編碼以外，另外

邀請了嘉義大學科學教育領域專家、與現任自然科教師一同進行編碼的工作，其中科教領域專家只對認知需求部分進行編碼。三人完成編碼之後，進行一致性分析分析以確定本研究的編碼是否合理，一致性分析的結果如表4所示。相較於Enright、Allen、與Kim（1993）的50%~79%，本研究的一致性百分比無論是TIMSS的物理試題，或是國中基測的物理試題，皆在80%以上，顯示認知成份編碼之一致性是可接受的（請參見表4）。因此，本研究後續根據各個試題的編碼所進行的難度分析，所得到的結果應是可以信賴的。

表4 TIMSS與國中基測物理試題認知成份編碼一致性百分比

編碼者	TIMSS		國中基測	
	認知需求層次 (完全一致)	概念數 (完全一致)	認知需求層次 (完全一致)	概念數 (完全一致)
1 vs. 2	56 (81%)	61 (88%)	60 (80%)	64 (85%)
1 vs. 3	57 (83%)	--	--	--
2 vs. 3	59 (86%)	--	--	--

註：表中的數字是編碼一致的題數，總題數TIMSS為69題，國中基測為75題；編碼者中的1是筆者，2為自然科教師，3為科教領域的專家。

二、物理試題內容架構的分配特徵

依據TIMSS的內容架構（參見表2），以物理現象的試題數最多，約30%，而聲的試題最少約4%；若以認知需求層次來看，層次六辨識熟悉的實驗控制與層次七設計調查所佔的試題較少為6%，其餘層次的試題分佈較平均，約佔總題數的14%~19%。

相較於TIMSS的內容架構，國中基測物理試題（參見表3）中，物理現象、電與磁、以及力與運動所佔之比率較高，介於23%~27%間，而能量轉換、光、及聲之試題所佔比率只有7%~9%，與TIMSS試題之分配有差異存在；在認知需求層次上，國中基測與TIMSS架構有些微差異，完全沒有第七個層次設計調查，且在第三層次應用科學原則解決問題、以及第四層次科學公式解決問題所佔的比率最多，分別為32%與21%，其餘所佔比率則介於9%~13%之間。

三、物理試題難度變異的解釋力

本研究所提出之認知成份包含總字數、認知需求層次與解題所需概念數三個，表5和6分別是TIMSS與國中基測物理試題難度與認知成份間之相關矩陣；不論是TIMSS或國中基測，物理試題難度與認知需求層次的相關都是最高的，均在.800以上，其次是解題所需概念數、以及總字數，且三個認知成份與試題難度之相關都達顯著。

不論TIMSS或國中基測，均可用總字數、認知需求層次與解題所需概念數三個變項來預測效標變項，且三者之間的相關介於.200~.560之間，預測變項間的共線性問題應該不顯著，因此對於預測變項對效標變項之解釋力不會有影響。對TIMSS物理試題而言，如表7所示，三個預測變項的解釋力為80.2%（調整後79.3%）；對國中

基測物理試題而言，三個預測變項的解釋力為78.5%（調整後77.6%），但如表8中顯示，總字數並未達到.05之顯著水準，因此，只有兩個預測變項時，其解釋力為78.2%（調整後為77.6%）。

表5 TIMSS八年級物理試題難度與認知成份間之相關矩陣（題數=69）

	試題難度	總字數	認知需求	概念數
試題難度	1			
總字數	.398**	1		
認知需求	.800**	.295*	1	
概念數	.767**	.292**	.555**	1

* $p < .05$. ** $p < .01$

表6 國中基測物理試題難度與認知成份間之相關矩陣（題數=75）

	試題難度	總字數	認知需求	概念數
試題難度	1			
總字數	.344**	1		
認知需求	.859**	.337*	1	
概念數	.608**	.159	.497**	1

* $p < .05$. ** $p < .01$

TIMSS八年級或國中基測物理試題之迴歸方程式如下：

$$\hat{Y}_{TIMSS \text{ 試題難度}} = .002 \times (\text{總字數}) + .214 \times (\text{認知需求層次}) + .298 \times (\text{概念數}) - .770$$

$$\hat{Y}_{\text{國中基測試題難度}} = .353 \times (\text{認知需求層次}) + .197 \times (\text{概念數}) - 1.246$$

由迴歸方程式看出，TIMSS八年級或國中基測之預測方程式係數雖有不同，但仍可以發現除總字數以外，認知需求層次與解題所需概念數對試題難度的影響是很大的，當變動一個單位時，難度變動的幅度介於.197~.353之間。

表7 TIMSS八年級物理試題難度參數多元迴歸方程式係數摘要表

預測變項	非標準化係數	標準誤	標準化		
(常數)	-.770	.098	--	-7.831***	--
總字數	.002	.001	.119	2.013*	3.837
認知需求層次	.214	.029	.504	7.510***	4.756
解題所需概念數	.298	.044	.460	6.844***	5.415

註：整個模式的解釋力為 $R^2 = .802$ （調整後為.793）。

* $p < .05$. *** $p < .001$

表8 國中基測物理試題難度參數多元迴歸方程式係數摘要表

預測變項	非標準化係數	標準誤	標準化係數	t	條件指數
(常數)	-1.246	.116	--	-10.729***	--
總字數	.001	.001	.063	1.082	4.005
認知需求層次	.353	.033	.718	10.812***	5.306
解題所需概念數	.197	.052	.241	3.798***	7.414

註：整個模式的解釋力為 $R^2 = .785$ （調整後為.776）。

*** $p < .001$

四、不同物理能力水準學生認知運作之差異

為進一步探討國中基測中不同物理能力水準的學生之認知運作差異情形，同上的作法，藉由Rasch模式得到基測物理試題難度（b值）、以及受試者的能力值（ θ 值）。當難度值為負表試題趨於簡單，難度值為正表試題趨於困難，而難度值接近0表試題難度適中；受試者能力部份，也是同樣的情形，當能力值為負表受試者能力較低，難度值為正表受試者能力較高。

依據本研究所得之認知成份架構，研究者根據試題難度分佈情形將試題依難度分成三個水準（表9），水準一為試題難度小於-.5的試題，共18題，此為較簡單的試題，平均難度為-.905，總字數平均為66.831，認知需求之平均為.998，而概念數之平均為.556；水準二則是試題難度介於-.5到.5的試題，有42題，這是屬於中等難度的試題，平均難度為.035，總字數平均為81.789，認知需求之平均為2.524，概念數之平均為1.293；水準三為試題難度大於.5的試題，共15題，是指難度較高的試題，平均難度為.989，總字數平均為89.071，認知需求之平均為4.203，而概念數之平均為2.134。

若由總字數來看，發現水準一試題平均字數只有66字，比水準二和三的試題來的少，而水準二和三之字數平均差異不大。認知需求部分，水準一試題的認知需求平均是.998，故該水準試題要求認知需求必須在層次一的了解簡單訊息或層次二的了解複雜訊息之間；而對水準二試題而言，其認知需求需在層次三的應用科學原則解決問題和層次四的科學公式解決問題；至於對水準三的試題來說，題認知需求必須在層次五解釋現象與層次六的辨識熟悉的實驗控制。以概念數來看，其各水準的平均分別為.556、1.293、以及2.134，因此難度水準一的試題最容易的，只要具備0或1個概念數，就可以答對該層次的試題；而難度水準三試題最難，概念數要求為2個以上才可。

表9 不同水準試題難度及認知成份平均數對照表

試題水準	試題數	難度	總字數	認知需求	概念數
一 ($b < -.5$)	18	-.905	66.831	.998	.556
二 ($-.5 < b < .5$)	42	.035	81.789	2.524	1.293
三 ($b > .5$)	15	.989	89.071	4.203	2.134

研究者進一步將受試者能力依據分配情形分成四個水準（表10），亦即將能力低於-1.0的學生歸類為低於基礎水準、-1.0~0.0之間為基礎水準為、0.0~1.0之間為精熟水準、而能力值在1.0以上者則屬於進階水準的學生。以低於基礎水準之考生來說，不論在哪一個難度水準其答對率均在.40以下，且在難度水準二和三答對率都不到.20；而基礎水準的考生，在試題水準一的答對率為.60，且在試題水準二和三之答對率均低於.40，與研究者預期相同；若以精熟水準考生來看，在試題水準一的答對率為.85，而試題水準二的答對率為.591，試題水準三的答對率為.321，也符合預期的結果；相較於前三個水準的考生，進階水準的考生是能力最好的考生，所以在試題水準一和水準二的試題以上的答對率都在.85以上，且在最困難的試題水準上答對率

也最高，其答對率為.70。

表10 不同能力水準考生在不同難度水準試題之答對率

能力水準	試題水準		
	一 ($b < .5$)	二 ($-.5 < b < .5$)	三 ($b > .5$)
低於基礎水準	.315	.176	.109
基礎水準	.606	.372	.204
精熟水準	.855	.591	.321
進階水準	.965	.871	.703

綜合以上的分析（根據表9和表10），在基礎水準之考生，所具備的認知需求層次為一和二，亦即了解簡單訊息與了解複雜訊息；精熟水準之考生，其具備的認知需求層次為三和四，即應用科學原則解決問題與科學公式解決問題；而進階水準的考生則是具備認知需求層次五和六，解釋現象與辨識熟悉的實驗控制。此一結果可提供教師在教學上的參考，對於測驗編製者也具有一定的意義。

伍、結論與建議

本研究根據試題的內涵與認知成份分析的觀點，來對TIMSS八年級物理試題進行難度分析，發現總字數、認知需求、以及解題所需概念數等三項認知成份解釋了79.3%的難度變異。而在國中基測的試題方面，總字數、認知需求與解題所需概念數等三項認知成份對於試題難度的變異，解釋力為77.6%，但因總字數之迴歸係數不顯著，故將該變項刪除，其所得之解釋力仍為77.6%，因此最後只以認知需求及解題所需概念數來解釋國民中學基本學力測驗物理試題難度變異。

在考生能力水準與試題水準之比較，也可以發現基礎水準的考生只能了解簡單訊息與了解複雜訊息，精熟水準的考生則是具備應用科學原則解決問題與科學公式解決問題兩個認知需求，而進階水準之考生則是有解釋現象與辨識熟悉的實驗控制之認知需求。因此，在教學上，教師可以針對不同能力水準之考生，提供不同的教學輔助，以使學生的能力水準能夠有所提升。

參考文獻

- 丁振豐 (1995)。幾何圖形類比推理題目內容成份分析之研究。**臺南師院學報**，28，83-114。
- 余民寧、賴姿伶、劉育如 (2005)。國中基本學力測驗實施成效之初步調查：學校的觀點。**教育與心理研究**，28(2)，193-217。
- 宋曜廷、許福元、曾芬蘭、蔣莉蘋、孫維民 (2007)。國民中學學生基本學力測驗的回顧與展望。**教育研究與發展期刊**，3(4)，29-50。
- 李濟國 (2001)。影響高中學生物理學習成就因素之探討。**科學教育月刊**，240，21-30。
- 林世華、葉嘉惠 (1999)。數字系列完成測驗試題認知成分分析之研究。**教育心理學報**，31，139-165。
- 林碧珍、蔡文煥 (2003)。我國國小四年級學生在國際教育成就2003試測的數學成就表現。九十二學年度師範學院教育學術論文發表會論文集。台南：國立台南師範學院。
- 涂柏原、梁恩琪、翁大德、楊毅立 (2004年3月)。國中基測自然科試題分析研究。「科技化測驗與能力指標評量國際研討會」發表之論文。國立台南師範學院。
- 涂柏原 (2007)。國中生基本學力測驗量尺分數轉換之實徵研究。**教育研究學報**，41(1)，61-77。
- 洪碧霞、林素微、林娟如 (2006)。認知複雜度分析架構對TASA-MAT六年級線上測驗試題難度的解釋力。**教育研究與發展期刊**，2(4)，69-86。
- 陳吉仲、郭曉怡、李佩倫 (2007)。影響國中基本學力測驗分數的因素之分析。**教育政策論壇**，10(4)，119-142。
- 陳竹村 (2003)。TIMSS 1999 台灣名列前茅極可能因素探討。**教育研究月刊**，108，133-146。
- 張殷榮 (2001)。我國國中學生在國際測驗調查中科學學習成就影響因素之探討。**科學教育**，244，5-10。
- 張銘秋、謝秀月、徐秋月 (2010)。PISA科學素養之試題認知成份分析。**課程與教學季刊**，13(1)，1-20。
- 劉春初 (2010)。TIMSS-R架構與DEA分析法的運用—以台灣地區國民中學學校經營效率南北地區比較為例。**臺東大學教育學報**，15 (1)，167-184。
- 盧雪梅、毛國楠 (2008)。國中基本學力測驗自然科之性別差異和差別試題功能 (DIF) 分析。**測驗學刊**，55(4)，725-759。
- 羅珮華 (2000)。第三次國際數學與科學教育成就研究後續調查之抽樣設計。**科學教育月刊**，235，14-20。
- Dimitrov, D. M. (2007). Least squares distance method of cognitive validation and analysis for binary items using their item response theory parameters. *Applied Psychological Measurement*, 31(5), 367-387.
- Dimitrov, D. M., & Raykov, T. (2003). Validation of cognitive structures: A structural equation modeling approach. *Multivariate Behavior Research*, 38(1), 1-23.
- Embretson, S. E. (1993). Psychometric Models for Learning and Cognitive Processes. In Frederiksen, N., Mislevy, R. J. & Bejar, I. I. (Eds), *Test Theory for a New Generation of Tests* (pp. 120-150). Hillsdale, N J: Erlbaum.
- Embretson, S. E., & Daniel, R. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science Quarterly*, 50(3), 328-344.

- Embretson, S. E., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38(4), 343-368.
- Enright, M. K., Allen, N., & Kim M. (1993). *A Complexity Analysis of Items from a Survey of Academic Achievement in the Life Sciences* (ETS Research Report 93-18). Princeton, N.J.: Educational Testing Service.
- Leong, S. C. (2006, May). *On varying the difficulty of test items*. A paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., et al. (2000). *TIMSS 1999 international science report*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Pellegrino, J. W., Chudowsky, N., & Glaser R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Center for Education, National Research Council.
- Rosca, C. V. (2004). *What makes a science item difficult? A study of TIMSS-R items using regression and the Linear Logistic Test Model*. Unpublished doctoral dissertation, Boston College, Boston.
- Yepes-Baraya, M. (1996, April). *A cognitive study on the National Assessment of Educational Progress (NAEP) science assessment*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York. Retrieved from ERIC database. (ED404343)
- Yepes-Baraya, M. (1997, March). *Lessons learned from the coding of item attributes for the 1996 NAEP science assessment G4 results*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago. Retrieved from ERIC database. (ED409356)