

計 畫 編 號
NAER-99-12-B-2-04-00-2-08

# 國家教育研究院

## 實作評量於數感概念的應用

### 研究報告

研究主持人：謝如山（國立臺灣藝術大學副教授）  
協同主持人：謝進昌（國家教育研究院助理研究員）  
研究助理：傅心怡（國立臺灣藝術大學研究助理）  
研究助理：郭惠娟（國立臺灣藝術大學研究助理）  
研究助理：張毓珍（國立臺灣藝術大學研究助理）  
研究助理：顏瑛慧（國立臺灣藝術大學研究助理）

研究期程：民國 99 年 5 月至民國 100 年 4 月

執行單位：國家教育研究院

中華民國 100 年 7 月



# 目錄

壹、計畫源由與目的 .....	6
貳、文獻探討 .....	8
參、研究方法 .....	12
肆、研究結果 .....	15
伍、結論 .....	45
陸、建議 .....	46
參考文獻 .....	47

## 表次

表 1 各年級與試題的平均得分表 .....	15
表 2 各層面估計概況整理 .....	17
表 3 各層面模型估計結果與配適狀況 .....	18
表 4 原始總分與 Rasch 參數對照表 .....	18
表 5 電器組第一子題的解題策略分析 .....	24
表 6 電器組第二子題的解題策略分析 .....	25
表 7 電器組第三子題的解題策略分析 .....	26
表 8 聰明的柯南第一子題的解題策略分析 .....	28
表 9 聰明的柯南第二子題的解題策略分析 .....	31
表 10 聰明的柯南第三子題的圖表策略分析 .....	32
表 11 聰明的柯南第四子題的策略分析 .....	33
表 12 哈利波特第一子題的解題策略分析 .....	34
表 13 哈利波特第二子題的解題策略分析 .....	36
表 14 哈利波特第三子題的解題策略分析 .....	38
表 15 慶生會第一子題的解題策略分析 .....	40
表 16 慶生會第二子題的解題策略分析 .....	41
表 17 慶生會第三子題的解題策略分析 .....	43

## 圖次

圖一 變數分布圖 .....	16
圖二 測驗訊息曲線 .....	20
圖三 試題一試題特徵曲線 .....	21
圖四 試題二試題特徵曲線 .....	22
圖五 試題三試題特徵曲線 .....	22
圖六 試題四試題特徵曲線 .....	23



# 實作評量於數感概念的應用

## 中文摘要

本研究目的，主要在探討開放性試題於學生數感概念的可行性，以四等分制來進行分析，透過多重 Rasch 模式的分析方式，是否能發現不同年級學生對不同題目間是否有認知層次的不同。由於本文為一探索性的研究，故在題目的選擇上，先使用數感的實作評量試題，以找出現階段學生是否對數感概念在學習上是否產生落差。

有鑑於古典測驗理論下的分析所採用的指標，像是難度、鑑別度、信度，都會受樣本的影響(sample dependent)。也就是說，所得到的這些指標，會依據受測樣本的不同而有所差異，同一份試題施測於兩個學校，在 A 校所得到的試題參數，和在 B 校所得到的參數有可能會截然不同。另一方面，古典測驗理論下來估計受試者的能力值，容易受到試題的影響(item dependent)，同一位學生在一份簡單的考試卷可得高分，而在一份難的考卷卻會得低分，只要測驗的題目不同，受試者所得到的能力估計值就會有很大的差異，在不同試卷的題目，即使內容相近，不同受試者之間的成績也無法做直接的比較與對照。然而，試題反應理論能補足古典測驗理論分析的缺點，進行較嚴謹的分析。

多層面的 Rasch 模型，是延續 Rasch 模型發展而來，由於可以同時考量試題難度和評分者嚴厲度之間的關係，被廣泛的使用在許多不同的領域。

實作評量與另類評量(alternative assessment)及真實評量(authentic assessment)十分類似，但另類評量為廣泛的指有別於傳統的紙筆測驗，真實評量則著重於評量內容於現實生活的結合，實作評量則是重視學生參與建構、並進行實作的過程與結果。

數感的重要，可從 Yang 與 Wu (2010)提出四個理由。第一，數感可以表現出彈性、創造、合理與效率的思考模式 (Dunphy, 2007)。第二，對於數量、數字、運算的主要概念，能有效率及有彈性的應用至生活中。第三，成年人在的數學思考的發展也與數感有直接的關係，Dehaene (1997)與 Berch (2005) 提及成人應要進行直覺的數感來使未來的數學思考與應用更加豐富。第四，過度的計算不僅會限制孩子在數學的思考與理解，也會阻礙孩子在數感上的發展 (Burns, 1994；Kiplatrlick et al., 2001；Reys & Yang, 1998；Yang & Li, 2008)。

本研究從桃園縣取樣四、五與六年級，四年級有四個班級，其他年級各三個班級，每年級約一百位學生，共三百多位，以發現實作評量試題在估測概念的應用。本研究設計實作評量試題，試題性質為應用、分析、評鑑與創造等向度。每題均有三到四子題，每一題組以 3 分制進行評分，最高為 3 分，最低為 0 分，以進行設計。得 3 分者完全正確，2 分者為少部份錯誤，1 分等級者為部份正確，0 分者為完全錯誤。

研究結果發現，有關電器組的評量目的為要求孩子用他自己的策略選擇後，

再進行計算，在此部分的成就表現，幾乎全班都有 25 人答對。雖然子題二，進行了減法運算，學生也都能回答正確。於第三子題，雖出現加減法的混合運算，每班也約有 20 位學生都能列出正確的算式，進行解題。由此可看出，四至六年級學生在基本計算的數概念部份有一定的學習水準，學生亦能接受此種題型的出題方式。

有關聰明的科南，目的要學生能覺得問題是否適當，能提出理由、判斷圖表、製作長條圖與設計問題來問同學。學生在第一子題提出很多不同的理由來贊同或反對這一問題，不論學生的理由如何，只要合理就視為正確的答案。於此，也對課本將長條圖延後至六年級的適當性，提出證據，學生可於四年級學習，而不需延到六年級。最後，要求學生由圖表來想出一個問題，有一班約半數同學都能提出，學生表現均相當優異。如學生用折線圖，用正字標記標示等，這都是相當合理的表達方式

於哈利波特題組的目的在於測量孩子在時間的量感。在第一子題中發現，學生在時鐘的標示部份，出其意料的相當薄弱，一班少於十人答對。可能與學生沒有進行撥鐘的實作有關。如果學生沒有進行實作，即使學生到高年級，觀念還是不清楚。最後請學生自行設計一個上午的時間表，學生答對的表現也的確不多，每班約 5 人。因為本題有很多學生產生誤解，誤認為自己可以安排行程，也就是未來在設計題目時要更加明確。但若日後學校考試能針對此一方向出題，學習表現應會更佳。

慶生會的目的在於測量學生對於因倍數概念與發現乘除法的關係。由於因倍數是在高年級的教學，所以預期學生在四年級時，應尚未教到本單元。然而，在學生的表現中，學生未回答的比率不高，只有四年級一班的學生有 25 人，其它班級的表現都相當優質，如有提出一種的分法，二種以上分法者。在第三子題的難度雖然增加，學生仍可以用分法的方式進行解題，雖然答對人數不多，但可鑑別出較高層次的學生。

本實作評量的結果發現，學生在開放性問題的表現相當多元，可以測量出學生在數學方面的能力，期望在未來的大型考試中，看到這樣有深度的題目在國內中出現。

**關鍵字：實作評量、數感**

# The Application of the Performance Assessment in the Concept of Number Sense

## Abstract

The purpose of the current study is to propose the possibility to design the open-end question to examine students' number sense. Four ranking level were design to analyze students' achievement. Via the multi-facet Rasch Model approach, researcher aims to examine whether different graders' on different questions perform differently. Because this is an exploratory research, the questions were focused on the number-sense field to find whether students have difficulty in fourth to sixth grade levels.

In terms of the parameters of classical test theory were influenced by the sample size, such as difficulty and discriminatory. That is, the parameters will perform differently because students are in different schools. The results of students in A school are different from those in B school. Besides, the estimates of students ability are depend on items. The same student score higher on one exam, but lower on another. The parameter of ability estimator varied dependent on items. However, the item response theory can be used to avoid the problems as shown above.

The assumption of multi-facet Rasch model is based upon Rasch model. The model is able to take into account the item difficulty and the grader's aspects, which can be applied to different fields.

The performance based assessment, alternative assessment and the authentic assessment are similar. The alternative assessment is extensively used differently with respect to the paper-pencil test. The authentic assessment stressed on the contents related to daily life, but the performance assessment emphasize on that students need to construct the knowledge by themselves.

The importance of number sense can be driven from Yang and Wu (2010) with four reasons. First, number sense can help students to show flexible, creative, reasonable and efficient thinking model (Dunphy, 2007). Secondly, with respect to the concept of quantity, number and operation, students can apply to their daily life. Thirdly, adults thinking development is directly related to number sense. Dehaene (1997) and Berch (2005) proposed that adults should utilize intuition to develop their mathematics thinking ability. Fourthly, over practice not only limits students' thinking and understanding ability, but restricts students' development in number sense (Burns, 1994 ; Kiplatrck et al., 2001 ; Reys & Yang, 1998 ; Yang & Li, 2008).

Total were 300 participants were sampled from Chungli, Taoyuan County. Four classes students were from fourth grade, and three classes were from fifth and sixth

grades, respectively. The aspects of items were designed for students to analyze, apply, evaluate and create questions. The score ranged from 3 to 0. Students graded 3 were totally correct. They scored 2 were mostly correct. Students graded 1 were partially correct and 0 were totally incorrect.

The results found that about the items for electricity aims to ask students to take their own strategy to solve the problems. For each class, over 25 students were correct. In the second question, most students answered correctly. For the third question, although students need to add and subtract, more than 20 students' answers were correct. Most students from fourth to sixth grade levels performed correctly in the first item.

The second item is related to statistics data. Students need to propose the reason, evaluate the graph, to draw a bar chart and to design a question from the graph. Students gave many different reasons to approve and against the question. As their answers are reasonable, they were graded correct. The results provided the evidence that statistics chart can be taught at fourth grade level instead of sixth grade. Finally, students were asked to propose a question based on the chart. About one half of students perform well.

The third item aims to measure students' concepts of time. Most students calculate the time correctly, but few draw the clock time precisely. They may lack of experiences to set a clock. If students in low grade level do not have the experiences, they will make the same mistake in high grade levels. The question also asked students to design a schedule. Few than 5 students made the answers correct. It may results from that students misunderstood the problems, so that most students arranged the schedule by themselves regardless of the questions suggested.

The item of birthday party is to evaluate the concept of factors and the relationship of multiplication and division. Due to related contents are taught in the high grade levels, fourth graders may have difficulty to deal with the problem. However, students in fifth and sixth grade levels did not perform well.

The results found that students have shown different ideas in these four items. The design of the open-end questions can help us to diagnosis students' mathematics ability in multiple aspects. We expect to have these open-end questions applied in Taiwan Assessment of Student Achievement (TASA) in the near future.

Key words: performance assessment, number sense



## 壹、計畫源由與目的

選擇題是應用最廣的一種測驗題型，可以有效的測量各種不同知識與學習成果。選擇題之所以應用廣泛的原因眾多，其中包括評分容易，可以很快的辨認學生作答反應的型態，評分也不會因為其他因素，如字跡潦草、寫錯字、或是其他因素造成評分的不公平。選擇題也會有較高的信度，因為測驗題數相對而言，可以比簡答題題數來的多，因而容易提高信度。且選擇題的題目可以互相獨立，所以測量到的構念，可以廣泛的覆蓋多元的教學內容。再者，選擇題比其他種題型(例如簡答題、是非題、配合題)較易建構高品質的題目，而且就分析角度而言，有較多軟體，可以快速的為選擇題進行題目分析。

雖然選擇題具有許多優勢，但有其限制，例如，選擇題並不適合測量所有的知識與能力，因為這種題型只能要求學生選出正確的答案，但是對於一些高階的數學或科學方面的問題解決與組織思考的能力，或是對於音樂、藝術等表演藝術的能力，是無法進行測量的。此外，設計選擇題型時，較不易找到看似有理卻不正確的誘答選項，一些較機靈的學生，可能會使用答題的技巧，來猜測出問題的答案，有可能學生對於答案要如何計算不清楚，卻可以得分。反觀來說，開放性問題不易猜測，而從答題中可以看出，學生對於知識理解的困惑之處，並依據所填答的內容，而給予適度的補救教學。再者，選擇題型式的考題，不見的適用所有的學生，尤其對於圖像式思考或是需要藉由引導步驟來循序漸進式找出答案的學生，可能並不合適。

相對應於選擇題的考試型式有很多種，而實作評量為一種新興的考試型態。然而，實作評量有很多種形式存在，例如藝術與音樂領域，對於可以要求學生演唱一首民謠或做一件手工藝品，並依據學生的演唱技巧、或是手工藝品的精細程度來進行評量。對於職業與工程教育課程，可以要求學生進行汽車維修、做一件木工來進行評量。當然，一般的學科測驗，例如數學、科學與語文等，也可以藉由實作評量，來評斷學習成果。就數學而言，可以測試學生是否能夠運用數學，來解決現實生活中所遇到的問題，並利用與人溝通與協調的技巧，來進行問題解決。而科學方面，可以藉由觀察、形成假設，蒐集資料以獲得科學的素養與能力。至於語言，雖說一般的辭彙與用法可以藉由筆紙測驗來測量，但是對於口語表達或是演說的能力，則是需由實作評量來完成的。

在多元智慧理論 (Theory of Multiple Intelligence) 中，Gardner (1993) 提出了人類的智慧，包括聽覺智慧(音樂)、語言、數學邏輯、空間智慧、肢體智慧、內省、自然觀察與人際互動等八種智慧，對於學校的教學，應著重的應為各個面向，而只著重在智能上的評估。而對於評量方面，也應可採多元的評量方式，方能測出學生真實的能力。雖然選擇題的題型有其優勢存在，但對於某些面向的智慧評估來說，是很困難的。

近年來，IEA (International Association for the Educational Achievement)所主辦的國際數學與科學研究(TIMSS, Trends in International Mathematics and Science Study), TIMSS 主要是將學齡的學童數學與科學的成就進行國際間的排名比較，在 TIMSS 的測驗中，除了進行一般傳統的選擇題式的筆紙測驗之外，亦有採用實作評量的題目，要求學生藉由操作教具並將結果寫在評量單，評分者會依據學生所填答的內容進行評分，以了解學生的數理能力。由 TIMSS 的做法可得知，實作評量可以運用在大型考試中的，只要能夠建立評分的公平性、進行準確的時間控制、選取適當的評分人員，亦可以得到公平的結果，對於學生的成就表現，也可以藉由實作評量的評估，有更深入的了解。

實作評量的題目大多是採多分題的形式，例如 4 分，學生答對部分概念得到 1 至 2 分，略有錯誤但大致正確得 3 分，完全正確得 4 分，而對於這種試題的分析，則多局限在古典測驗理論的框架中，鮮少有文獻採用試題反應理論，來探索如何有效測量受試者的潛在能力。古典測驗理論下的分析所採用的指標，像是難度、鑑別度、信度，都會受樣本的影響(sample dependent)，也就是說，所得到的這些指標，會依據受測樣本的不同而有所差異，同一份試題施測於兩個學校，在 A 校所得到的試題參數，和在 B 校所得到的參數有可能會截然不同。另一方面，古典測驗理論下來估計受試者的能力值，容易受到試題的影響 (item dependent)，同一位學生在一份簡單的考試卷可得高分，而在一份難的考卷卻會得低分，只要測驗的題目不同，受試者所得到的能力估計值就會有很大的差異，在不同試卷的題目，即使內容相近，不同受試者之間的成績也無法做直接的比較與對照。然而，試題反應理論能補足古典測驗理論分析的缺點，進行較嚴謹的分析。

多層面的 Rasch 模型，是延續 Rasch 模型發展而來，由於可以同時考量試題難度和評分者嚴厲度之間的關係，被廣泛的使用在許多不同的領域。例如，姚漢禱和姚偉哲(2008)將多層面的 Rasch 模型應用分析在雙不定向飛靶優秀選手的射擊技術上，結果發現多層面的 Rasch 模型可以提供教練在選手訓練建議，並找出選手的弱點，進而提升選手的成績表現。張新立和吳舜丞(2008)則將多層面 Rasch 模型應用在學術研討會論文評分，總共有 131 位評審，每篇論文之評分工作則由三名相關領域專家進行同儕評審。他們的研究中發現不同學術分組的評審委員之間，存在嚴厲度的差別，也就是某些組的評審委員，會評分偏高，有些則一致性的給分較低，若是使用原始總分來評定論文表現優劣，會造成不公平的現象。測驗領域也廣泛使用此理論進行分析，像是分析寫作考試 (Engelhard, 1992; Twing & Williams, 1992) 與臨床醫學上的證照考試等 (Lunz & Stahl, 1990; Lunz, Wright, & Linacre, 1990; Lunz, Wright, Stahl, & Linacre, 1989)。多層面 Rasch 模型在實作評量領域也有許多應用(Smith & Kulikowich 2004; Basturk, 2008)，例如，Basturk (2008) 用此模型來評斷學生在做報告時的表現，學生報告分成六組，每組會依據六個向度：緒論、內容、描述、整體架構、圖表呈獻與表達力來做評分，在其研究中探討三個層面之間的關係，包括

每一組的表現能力，六個向度的難度、與評審的嚴厲度，研究結果指出多層面 Rasch 特別有利於分析多分題的數據，尤其是牽涉到需要評分者來進行評分，能夠有效的把評分者個人的嚴厲進行考量，而對受試者的能力進行客觀的比較。此外，透過此分析亦能將原本順序尺度的數據，轉換成具有可進行加減的等距量尺 (interval scale)，進行轉換後的 logit 尺度，可以展現能力在各個向度的強弱，與進行各組別之間的比較(例如性別差異)。分析後所得到的參數更可進一步的分析，雖然 Rasch 模式廣泛的為教育及心理計量領域所使用(王文中, 2004)，但是國內對於多層面 Rasch 的模型分析，尤其是在實作評量上的應用，使用的相當有限，因此，本研究將利用多層面的 Rasch 來分析實作評量的題目。

受試者在實作評量上的分數大多由評分者來進行給分。評分者之間的異質性可以透過嚴密的訓練，與完整的評分歸準，使其盡量達成一致，然而，人和人之間還是存在明顯的差異性，即使用客觀的框架將評斷的標準限制住，主觀的意識與個人好惡還是會在評分的過程中造成影響。尤其是在大型測驗上的應用，幾千人的考題，需要上百位評分者進行評分，某些嚴格的評分者，給的分數偏低，而某些寬鬆的評分者，則給的分數偏高，這些評分者嚴厲度的差別，會嚴重影響到學生成績的公正性。然而，這種評分者嚴厲度差異的問題，能夠藉由多層面的 Rasch 分析來解決，使學生的成績，不再受到評分者個人嚴厲度的影響，而能夠公平的進行比較。本研究將使用此模型，並針對四個數學實作評量的題目進行分析，施測對象為國小四年級、五年級、及六年級共三個年級 313 位學生，其中，三位評審依據評分歸準與學生的作答，進行 0~3 分的評分，本研究希望能藉由分析結果，提供未來分析實作評量數據的參考。

除了量化的分析之外，本研究亦針對每個實作評量題目，依據學生的解題策略，歸納出學生的迷思概念類型。實作評量提供教師學生的學習過程、與結果之間的連結，並不是單純的測量表面之的簡單測驗。結果或許沒有一個唯一簡單的答案，然而，評量所呈現的，是一個多元的表現與問題解決的過程，問題的形成、知識的組織、與統整多樣化的證據，以及原創性，都可能是實作評量的層面，本研究將對實作評量的各個層面加以評估與考量，其中包括實作評量的實施步驟、題目規劃、施測結果評估與試題分析的分析與研究。

## 貳、文獻探討

文獻探討包括兩個部份，前半部對實作評量的基本原理與相關研究作簡介，後半部則對數感概念作一概述。

### 一、實作評量的簡介與基本原則

以往教學為了顧及大規模施測的方便性與評分的客觀性，大多是使用選擇題的方式來評鑑學生成效，但近年來，社會變遷與提倡多元智慧的教學目標，使原本著重於基礎記憶知識的教學，逐漸轉變為強調學生高層次的知識與問題解決能

力的培養，因此，傳統的選擇題式的題型，無法完全的因應所有評量的需求。張麗麗(2002)指出，選擇題型在評量高層次與統整能力上有其困難度。而很多的選擇題目，指將知識進行分割與去情境化，使的測驗內容與生活脫節，無法將所學應用在生活中。此外，選擇題的考試只重視考試結果，不重視歷程，使得學生無法在學習歷程中自我建構出有意義的學習，並逐漸成為被動的學習者。為了克服這些問題，實作評量逐漸受到測驗界的重視。

實作評量與另類評量(alternative assessment)及真實評量(authentic assessment)十分類似，但另類評量為廣泛的指有別於傳統的紙筆測驗，真實評量則著重於評量內容於現實生活的結合，實作評量則是重視學生參與建構、並進行實作的過程與結果。實作評量的種類有很多，例如，可以使用具有結構限制的紙筆形式試題，要求學生完成個人或小組的計畫，進行日誌與省思，或由教師對於學生表現進行正式與非正式的觀察記錄，或是藉由科學實驗、展示或檔案評量的方式進行。

Shavelson, Baxter, & Pine (1992)提出了一些關於實作評量實施的基本原則：

1. 實作評量的內容需要超脫記憶與零碎的片段性知識，或是從多個變項中，選出一個單一正確答案的模式。評量要能夠捕捉學生對於學科知識的理解、認知與展顯問題解決的能力，並能夠鼓勵學生提供新穎、具有創造力的答案。
2. 學生必須依據操弄性與實驗性的教具來回答評量的問題，一些需要動手實作的評量內容必須要客觀且標準化。某些需要長期進行的計畫案或許無法以一個測驗時段來完成。
3. 長時間進行的計畫案與動手做的實作評量是較為昂貴且費時的，因此若能加上電腦化科技的輔助是較佳的。
4. 實作評量要能夠反映認知方面的發展，特別是要著重於心智方面，以用於評鑑學生的知識架構與了解學生的迷思之處。
5. 實作評量須與課程改革緊緊相扣。因為測驗分數主要是依據考試內容來做解釋，因此測驗分數應要能在學科的指標中做有意義的解釋。

## 二、實作評量的步驟

實作評量講求學生透過學生的動手操作，讓學生展現所學來解決問題。Stiggins(1994)建議在建構實作評量的試卷時，應考量下面的步驟：

### 1. 確定評量的原因與目的

必須先確立為何要實施實作評量，例如，評量的結果是要進行學生的分組以鑑定個別能力、還是僅供教學的參考。

### 2. 設定所欲評量的實作表現

依據評量教學中重要的概念與技能，設計出能夠測量學生真實能力的內容，並選擇所要評量的表現，是針對結果、過程還是對過程和結果兼顧來進行評鑑。

依據評量領域中所需表現的知識或技能，設定評量表現的標準。

### 3. 設計實作評量的內容

需決定作業的形式，實施評量的情境，與確定所要編制的試題數量。所選擇的內容應具有代表性。

### 4. 確定實作評量成績計算的標準

須決定評分的方式，採用分析式評分或是整體式評分，須對評分者進行專業的訓練，並使每個人對於評分歸準有所共識，最後，要決定記錄評分的方法，方法可採用檢核表、評定量表或是軼事記錄表等。

## 三、國內外相關實作評量研究

國內外有許多論文都對數學科的實作評量進行相關研究，而大多數著重於量化的實驗研究。詹元智(2002)以小六學生為對象，以兩個月的時間，採用實驗組與控制組的方式進行研究。對於控制組的學生，進行傳統教學、學生解題與練習。而實驗組則是進行數學實做評量教學以及八次的實作作業練習。教師以生活情境佈題、小組討論的方式，針對學生的問題給予實作評量的練習，並提供課後輔導與回饋。兩組學生在實驗前後須接受筆紙測驗與數學實作測驗。研究發現兩組學生在傳統筆紙測驗上的表現無顯著差異，但是在數學實作評量表現上，接受實作評量教學與練習的實驗組學生表現較佳。

曾安如(2004)則是探討數學日記寫作活動對於數學成就與態度的影響。研究亦採準實驗設計，實驗組的教學內容包括回憶上課摘要、解決生活情境的應用、對每個數學單元的感想、找出題目中錯誤的概念並進行澄清，將數學概念融入故事進行寫作等。結果發現實驗組學生在數學成就與數學態度的表現上均優於控制組學生。

Jurdak 及 Zein (1998)以中學生為對象，探討數學日誌對學生數學成就及學習態度的影響。其準實驗設計中，實驗組在每周數學課結束後進行數學日誌寫作，學生必須完成题目的寫作、閱讀與數學學習有關的內容，並寫出答案。控制組的學生則是進行課內測驗的計算與練習。研究發現實驗組學生在數學理解、程序上的知識與數學溝通上的表現均優於控制組，但是對問題解決、數學態度及在校數學成績上而言沒有顯著差異。

蔡正濱(2006)則探討實作評量中的評分者一致性。採用不同類型的計分歸準、不同複雜度的作業及受試者是否熟悉實作評量計分歸準的經驗三個面向來探討對評分者一致性的影響。研究發現對影響實作評量一致性的主要變異來源為受試者與作業項目之間的交互作用(pxt)變異，其是為受試者本身的變異(p)，再其次為受試者、作業與評分者三者之交互作用(pxtxr)。此外，不同的計分歸準對評分

的一致性的影響不大，高複雜度的作業比低複雜度的作業不易達到一致性，對於分析式的計分影響小於整體式的計分歸準。

整體而言，實作評量對於學生的學習是有助益的，而且，只要能夠對評分人員進行專業的訓練，實作評量也可以達到良好的信度。

有鑑於實作評量已有非選擇試題的形式出現，本研究已將四至五年級實作評量試題進行研發，每一試題分為三子題，每子題將依四分等級依學生學習表現進行評分。由於四至五年級實作評量試題多數較著重於數量的估測，即針對學生於數感的概念，故本研究將先以數量估測概念切入，以了解學生於數量估測的學習層次。

#### 四、數感概念的相關研究

有關數感的重要，Yang 與Wu (2010)提出四個理由。第一，數感可以表現出彈性、創造、合理與效率的思考模式 (Dunphy, 2007)。例如，算式 $12+9=?$ ，一年級學童在做加法運算時，可以運用湊十的特性來解決，應用 $12+10-1$ 的方式來得到答案。第二，對於數量、數字、運算的主要概念，能有效率及有彈性的應用至生活中。例如，當一台玩具汽車需要\$599元，一個音樂盒需要\$399元時，小明要買兩個，需花多少張100元。學生要能理解599元需要幾張100元的概念。第三，成年人在的數學思考的發展也與數感有直接的關係，Dehaene (1997)與Berch (2005) 提及成人應要進行直覺的數感來使未來的數學思考與應用更加豐富。第四，過度的計算不僅會限制孩子在數學的思考與理解，也會阻礙孩子在數感上的發展 (Burns, 1994; Kiplatrick et al., 2001; Reys & Yang, 1998; Yang & Li, 2008)。例如，學生對於 $25 \times 0.45 = ?$ ，答案應該是25的一半左右，學生的答案如果是1.125，則他們缺乏數感的能力。Yang與Wu(2010)亦指出傳統的數學教學，若過於重視紙筆運算，會使學生缺乏數感能力(Alajmi, 2004; Markovits & Sowder, 1994; Menon, 2004; Yang, 2005; Yang & Li, 2008)。

有關數感的定義，由於並未明確，NCTM (2000) 定義數感為學生能拆解與組合數字，能使用數字如 $1/2$ 或100做為參考，來解決相關的數學問題。學生要能了解十進位數字系統來進行估測，以理解數字並確認數字的相關大小與絕對性，研究者採用Yang與Wu (2010)對於數感的定義有四方向，第一為能了解基本數字與運算的意義。例如，學生應能了解十進位的數字系統；或是學生能用不同表徵來表示同樣的數字，例如，學生能發現1000是十個100，也可以是100個十(McIntosh et al., 1992)。第二為學生要能確認數字的相關大小與絕對性。例如，在估測 $2/5$ 有多大時，學生要要了解 $2/5$ 應會界於 $1/2$ 與 $1/3$ 間，因為 $2/4$ 等於 $1/2$ ，及 $2/6$ 是 $1/3$ 。第三，能使用適當的基準點。如教室的高度，可能是兩倍的大人高 (McIntosh et al., 1992)。第四、能判斷答案的合理性，即學生要能有估測的能力，例如，學生對於學校操場跑道的長度約為一百公尺的範圍，而不會誤認為10公里或10公尺。

故本研究目的，主要在探討開放性試題於學生數感概念的可行性，以四等分

制來進行分析，透過潛在類別模式的分析方式，是否能發現不同年級學生對不同題目間是否有認知層次的不同。由於本文為一探索性的研究，故在題目的選擇上，先使用數感的實作評量試題，以找出現階段學生是否對數感概念是否在學習上產生落差。

## 參、研究方法

以下依研究對象、資料來源、課程設計與統計方法分別論述。

### 一、研究對象

本研究從桃園縣取樣四、五與六年級，四年級有四個班級，其他年級各三個班級，每年級約一百位學生，共三百多位，以發現實作評量試題在估測概念的應用。本研究取樣的學校為中壢市市中心的小學，學區屬性較為中上，四至六年級每學年有十一班，均為常態分班。四年級九班與十班，為參與研究者另一研究案，進行實驗教學課程，已有兩年。本研究同時研究評分者效度，並將有三位評分者，以進行多層面 Rasch 的統計分析。

### 二、資料來源

以下依研究工具、面談、與實作評量的建構分別論述。

#### 1. 研究工具

本研究設計實作評量試題，試題性質為應用、分析、評鑑與創造等向度。每題均有三到四子題，每一題組以 3 分制進行評分，最高為 3 分，最低為 0 分，以進行設計。得 3 分者完全正確，2 分者為少部份錯誤，1 分等級者為部份正確，0 分者為完全錯誤，評分標準請見附件。

本研究設計四題組，其設計面向有數字的運算、圖表的判讀與製作、時間的量感與數字的關係等不同目標。第一題組意為檢視學生的數與計算的概念，要求學生能正確的進行加減法計算，從題目的意義中發現數字運算的基本概念。第二題組的目的在於讓學生判讀圖表，製作圖表，進而能發現圖表中的問題，如學生可用不同的想法來贊同或反對此一圖表的應用。第三題組旨在了解學生在時間的量感，要求學生計算時間、畫出時鐘、設計行程表等，了解學生在時鐘的觀念與主動安排行程的規劃能力。第四題組為評量學生對乘除法關係的了解，如因倍數的概念與發現乘法與除法的互逆關係等。

各題目間之能力指標對應如下。

題組一：N2-2 延伸加、減、乘、除 與情境的意義，使能適用來解決更多的生活情境問題，並能用計算器械處理大數的計算。

題組二：D2-2 能將分類資料整理成長條圖，並抽取長條圖中有意義的資訊加以解讀。

題組三：N2-8 能報讀(鐘面上的)時刻以及點算兩時刻間的時間；能理解 24 時制並應用在生活中。

題組四：N2-14 能在情境中，理解乘法交換律、等號的對稱性、「 $<$ 、 $=$ 、 $>$ 」的遞移性、加法和乘法的結合律與分配律，以及乘法和除法的相互關係。

本題組設計方式，參考國際數學測驗 pisa (programme for international student assessment)，設計數與量、統計圖表、時間與乘除法關係的試題。每一題組設計三到四小題的子題，以開放性問題為主，對學生的數學認知程度進行了解。

## 2. 面談

面談的目的是為了探究學生學習困難的原因，當學生完成傳統選擇題或實作評量的測驗後，基於學童的回答分數訪談十位學生。晤談時間從 20 分鐘到 60 分鐘不等，以瞭解學生解題策略的差異。

## 三、統計方法

多層面的 Rasch 模型 (Many faceted Rasch model, MFRM, Linacre, 1994) 從試題反應理論發展而來，為單向度單參數 Rasch 模型的延伸。MFRM 的模型可以解決一些傳統古典測驗理論假設下的問題。根據 Linacre(1994)指出，MFRM 有多項優點。第一，此模型將所有需考量的層面放在同一個尺度上，例如受試者各組的表現、試題難度、評審者的嚴厲度等。第二，因為 MFRM 建構於試題反應理論下，因此具有試題獨立、與受試者獨立的優點。即使是使用不同的題目、來進行施測，受試者的能力也不會受試題難度而有影響，此外，若要分析試題參數，即使不同的受試者來進行估計，所得到的試題參數也會維持一致。因此，Rasch 模組的分析下，不同的受試者、不同的試題、甚至不同的評分者所得到的結果都能維持一致性的推論，這對於實作評量的分析特別有利，因為對於實作評量的實務應用上，不大可能同一個評分者可以評比所有的項目，而不同的評分者，即使受到良好的訓練，也會有個別差異，某些評分者評的較嚴、而某些則評的較鬆，而這些較嚴格或是較寬鬆的評分，會直接影響到受試者的成績，而透過 Rasch 的分析，則能將這種鬆嚴不一的情況考慮進去，並將受試者的能力做合理的推論。第三，Rasch 模組中所提供“適配度”的指標可以用來找出哪些評分者的評分或是那些受試者的作答反應有不一致的情形。例如，若是某一位受試者，所有難的題目都答對，卻有幾題很簡單的題目沒有答對，這種情形下，Rasch 模組可以偵測這些受試者的作答反應，研究者可以找出這些作答者，並以深入了解原因 (Engelhard, 1992; Linacre, 1999)。Infit 與 Outfit 的均方值為 Rasch 模型適配度的主要指標，兩指標乃為各個層面上是否與模組偏離之重要的檢測依據。若兩值介於 0.5 到 1.5 (Engelhard, 1992)，則數據適合 Rasch 模組的分析。此外，

亦須評量參數估計的信度 (reliability) 考驗，若是信度指數接近 1 時，代表研究所蒐集的數據具有穩定的一致性。

MFRM 透過統計的模組，來探討試題的難度和受試者能力之間的關係。此模組將試題難度與受試者的能力放在同一個尺度上，並將分數轉化成 logit 分數，logit 分數為一等距尺度的分數，可以進行統計運算，並能展現受試者能力的強項和弱項，或是對不同的族群進行比較。本研究使用 MFRM 來進行分析，有兩大主要原因。其一是由於實作評量牽涉到評分者來進行評分，雖然在本研究所使用的數據中，三位評分者對全部的受試者都進行評分，但是這種情形在一般實務界並不常見。較為常見的情形是一份試卷，經由兩位評分者來評，如果分數沒有出入，則採用兩位評分者評分的平均來當作受試者的最終分數，相反的，如果差距大於某個指標，則會再找第三位評分者來評分。但是這些評分制度雖然客觀，但還是無法避免某些評分者評分嚴厲，某些則較為寬鬆的情形。而本研究所使用的 MFRM 的方法，則可以將評分者的嚴厲度進行考量，使受試者所得的分數更為客觀。其二、在國內用來評估實作評量主要分析方法多局限在古典測驗理論下，使用 MFRM 來進行實作評量分析可說是少之又少，因此本研究所提供的分析方法，可供未來其他實作評量實務研究者的參考。

## 肆、研究結果

### 一、量的分析

#### (一) 多層面 Rasch 分析

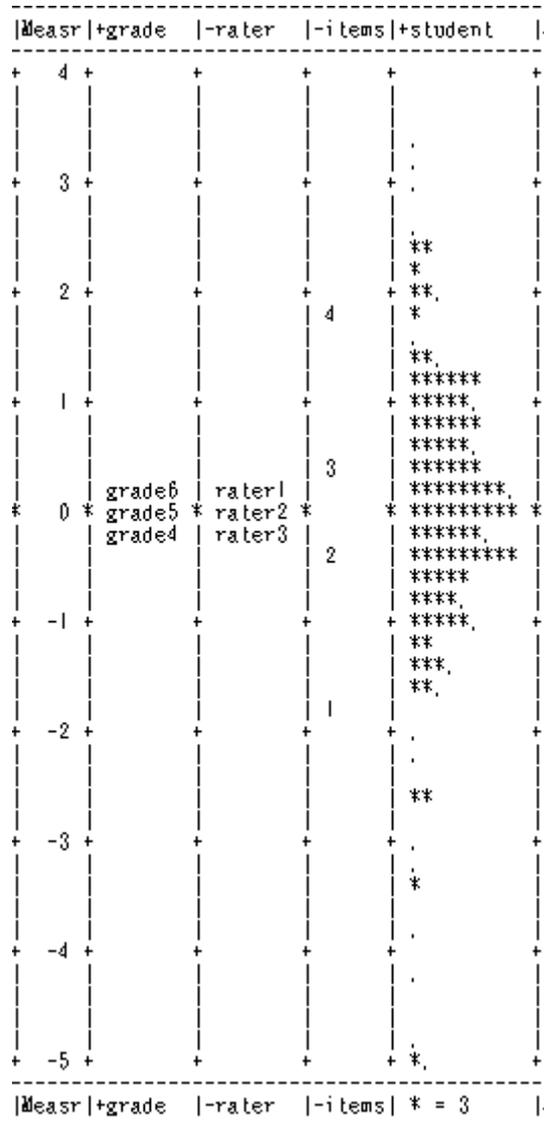
表 1 所呈現的是三位評分者，在每一個題目上所給的平均成績，例如，對於四年級的學生，第一位評分者在第一題所給的分數之平均為 2.27 分，而在第二題的給分平均分數則為 1.56。由表 1 可知，第四題的難度最高，三位評分者所給的分數在此題都是最低，而第一題是最簡單。而就評分者而言，每位評分者的嚴厲度略有不同，評分的差異性約為 0.01 至 0.05 分左右，最大的差異為四年級的第三題，第一位評分者所給分數之平均為 1.09，而第三位給的平均分數則為 1.14，而第一位評分者所給的分數，普遍來說低於第二位與第三為評分者，為較為嚴厲的評分者。

而就各年級的表現來說，則符合預期，六年級在各題的表現優於五年級，而五年級學生表現優於四年級，對於簡單的題目，年級間表現差異較小，而難的題目，差異較大，例如第一題四年級的平均得分為 2.26，和五年級的 2.27、六年級的 2.27 差異很小，而對於較難的第四題，四年級的得分平均則為 0.59，五年級為 0.68，六年級學生則為 0.73，可以看出對難的題目來說，年級之間的平均表現分數差距較大。

表 1 各年級與試題的平均得分表

年級	題 1	題 2	題 3	題 4
四年級				
評分者 1	2.27	1.56	1.09	0.59
評分者 2	2.26	1.57	1.11	0.59
評分者 3	2.26	1.57	1.14	0.62
四年級分數平均	2.26	1.54	1.11	0.59
五年級				
評分者 1	2.26	1.59	1.17	0.65
評分者 2	2.27	1.57	1.18	0.66
評分者 3	2.29	1.57	1.21	0.69
五年級分數平均	2.27	1.58	1.22	0.68
六年級				
評分者 1	2.29	1.56	1.24	0.73
評分者 2	2.27	1.57	1.24	0.71

評分者 3	2.26	1.58	1.25	0.71
六年級分數平均	2.27	1.60	1.26	0.73
總平均	2.25	1.57	1.23	0.67



圖一 變數分布圖

圖一則為變數分布圖 (variable map)，由這個分布圖中可以本研究中所考慮各個層面中的相對難易度散布狀況，最左邊的欄位是刻度，單位為 logit，logit 值越高，代表學生表現越好、評分者越嚴厲、或是題目越難，logit 的單位為等距尺度，為一具有連續性、單位又相等的數值。Logit 的大小不但能顯示大小的順序，而且數值之間具有相等的距離。第二個欄位則為年級，由此欄位可知，六年級學生平均的 logit 得分，高於五年級，而五年即又高於四年級，所以六年級的學生程度高於五年級與四年級學生，而年級之間的差距，大約是相等的，也

就是六年級學生與五年級學生在數學程度上的差距，等於五年級學生和四年級學生的差距。第三個欄位是評分者，第一位評分者所估計出的 logit 值最高，代表第一位評分者是最嚴厲的，而第三位評分者的 logit 值最低，代表評分最鬆。而第四個欄位則為每個試題之間的難度差距估計，第四題最難，所估計的難度值為 2，而第一題最簡單，難度約為-2，此外，第二題和第三題的難度介於第一題和第四題之間，第五個欄位是學生能力的估計分布，學生能力分布約為常態分布，大多集中於-1 至 1 之間，由此圖可推知，很少人可以答對第四題，因為第四題的難度為 2，已經超過大多數學生的能力範圍。

表 2 各層面估計概況整理

	評分者	年級	試題	受試者
Rasch 參數				
平均	0	0	0	-0.1
標準差	0.18	0.14	1.56	0.48
N	3	3	4	313
Infit				
平均	0.99	1.01	1	0.99
標準差	0.08	0.14	0.16	0.7
Outfit				
平均	1.01	1.04	1.01	1.01
標準差	0.07	0.15	0.17	0.8
信度	0.94	0.93	1	0.89
分離度	4.02	3.53	30.01	2.78
卡方值	34.4*	35.6*	2429*	1859.4*

表 2 為各層面分析參數整理，總共探討四個層面，並將評分者、年級、與試題估計的難度固定到 0，並觀察受試者能力的估計值。由此可知，此份試題對於受試者而言，是稍感困難的，受試者的平均能力的估計參數為-0.1，而就適配度 infit 與 outfit 來看，每個層面的適配度都在-0.7~1.3 的範圍內，代表使用的 Rasch 模式來進行估計應該是適合的。此外，每個層面的信度值都高達 0.8~1 之間，代表此數據在模組估計和觀測值之間的一致性很高。

分離度(Separation)越高，則代表越能將層面的類別區隔出來，如表二所示，試題的分離度為 30.01，代表就試題的難度而言，至少可以分成 30 個層級，也就代表難度差異性很大，有些題目很難、有些題目很容易。然而，相較之下，受試者的相似度較高，但也至少可以區分為 2 類程度的受試者。而值得注意的是評分者也出現差異性，分離度為 4.02，而卡方檢定亦達顯著，代表不同的評分者之間，評分的嚴厲度具有顯著性的不同。

表 3 進一步將所有層面的 Rasch 估計值陳列，其中參數平均值以 logit 為單位，為等距尺度的單位，可進行參數之間距離的比較，例如，四年級學生能力估計的平均值為-0.19，和五年級的 0.05 之間差距為 0.24 的 logit 的單位，而五年級和六年級學生能力的平均估計值差距則 0.09 個 logit 單位，代表 四年級和五年級之間數學能力的差距，大於五年級和六年級之間數學能力差距。

表 3 各層面模型估計結果與配適狀況

	參數平均值	參數平方差	Infit 均方	Outfit 均方
<b>年級</b>				
四年級	-0.19	0.04	0.9	0.89
五年級	0.05	0.06	0.96	1.05
六年級	0.14	0.05	1.17	1.19
<b>評分者</b>				
評分者 1	0.15	0.04	0.9	0.93
評分者 2	0.05	0.04	1.04	1.03
評分者 3	-0.2	0.04	1.04	1.07
<b>試題</b>				
試題 1	-1.92	0.05	1.24	1.26
試題 2	-0.32	0.05	0.85	0.89
試題 3	0.42	0.05	0.98	0.97
試題 4	1.81	0.06	0.95	0.94

此外，第一位評分者的嚴苛度的估計值為 0.15，高於第二位和第三位的 0.05 和 -0.2，就試題難度而言，第四題的難度最高，難度值遠高於其他三題，相對而言，第一題和第二題的難度較低，是屬於比較簡單的題目。每一個層面的適配度都座落在 0.7 與 1.3 之間，代表使用 Rasch 的模式來進行分析應該是適當的。

表 4 原始總分與 Rasch 參數對照表

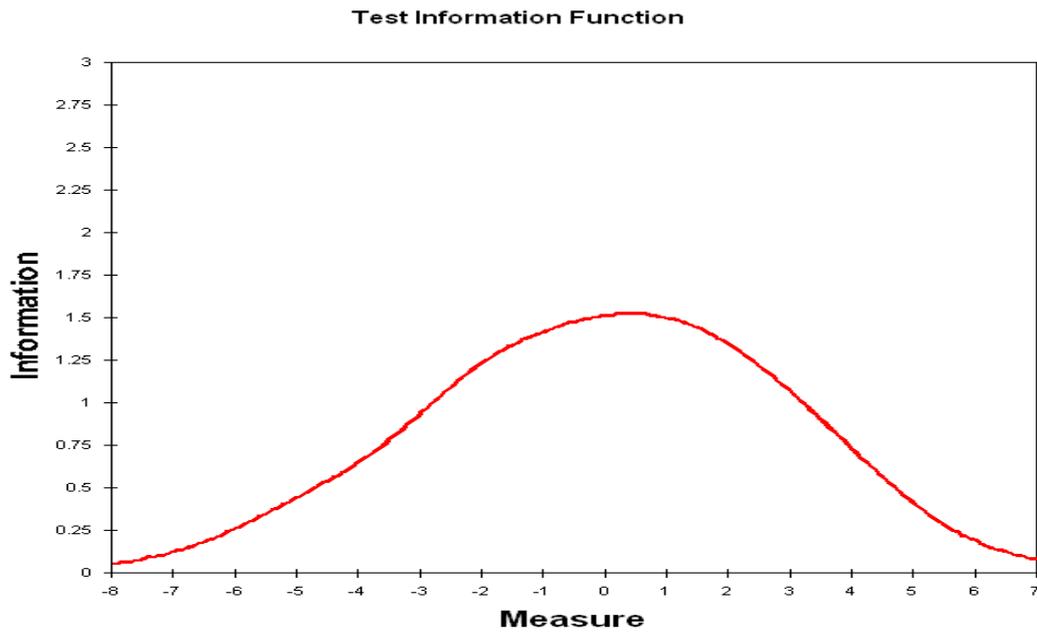
原始總分	Rasch 參數值	參數標準差	人數分布	百分比	百分位數
0	-6.05	2.01	6	1.9	1

1	-4.32	1.32	9	2.9	3
2	-2.97	1.03	10	3.2	6
3	-2.05	0.91	22	7	12
4	-1.28	0.85	43	13.7	22
5	-0.58	0.83	76	24.3	41
6	0.09	0.81	51	16.3	61
7	0.75	0.81	45	14.4	77
8	1.42	0.83	26	8.3	88
9	2.14	0.87	16	5.1	95
10	2.97	0.96	7	2.2	98
11	4.09	1.19	2	0.6	99
12	5.58	1.93	0	0	100

表 4 為原始總分和 Rasch 模式下所得到的能力估計值對照表，總共有四題，每題分數為 0~3 分，因此原始總分的範圍為 0 到 12 分，第二欄則是對應的 Rasch 能力參數值，例如原始分數 3 分，對照的 Rasch 能力參數為 -2.05。在此表中可見兩極端的分數差異，大於中間分數的差異。原始分數 11 分和 12 分所對照下的 Rasch 差異，約為 1.49(5.58-4.09)，大於中間對照的分數，像是 5 分與 6 分，Rasch 分數的差異則為 0.67(0.09-(-0.58))，這隱含要從 5 分進步到 6 分，比 11 分進步到 12 分來的容易，或是指得 5 分的人與得 6 分的人之間的能力差距比得 11 分和 12 分的人之間的能力差距來的小，因為所需要增加的 Rasch 能力值比較小。第三欄所陳列為參數估計下的標準差，兩極端分數的標準差大於中間的標準差，接下來的人數分布和百分比則為本研究樣本分布狀態，沒有人得到滿分，大多數人的分數集中在 4~7 分，最後一欄則呈現對應在各分數下的百分位數。

## (二)、試題分析

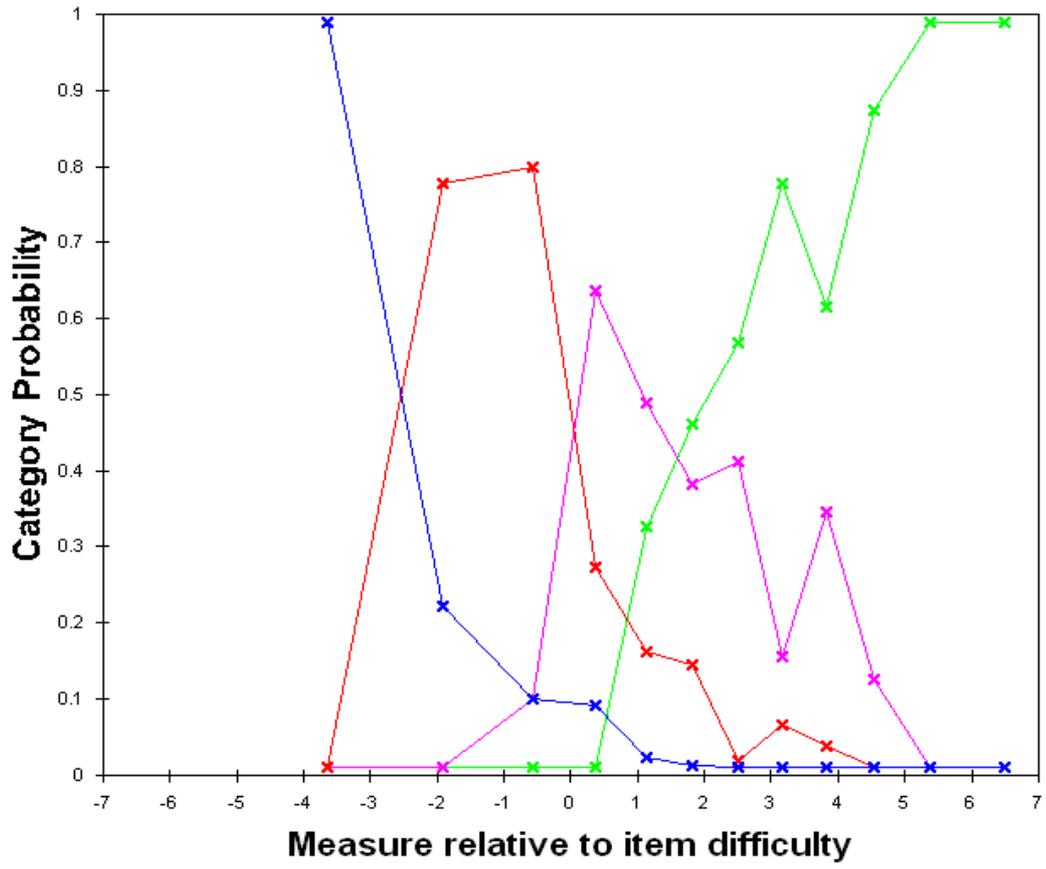
圖二為這份測驗的訊息函數 (Test Information Function)，由此分部可看出本測驗訊息量分布約為常態分布，最高峰點在能力值 0~1 之間，訊息量約為 1.5 左右，代表這份測驗難度稍難，較適合程度佳的學生來施測。



圖二 測驗訊息曲線

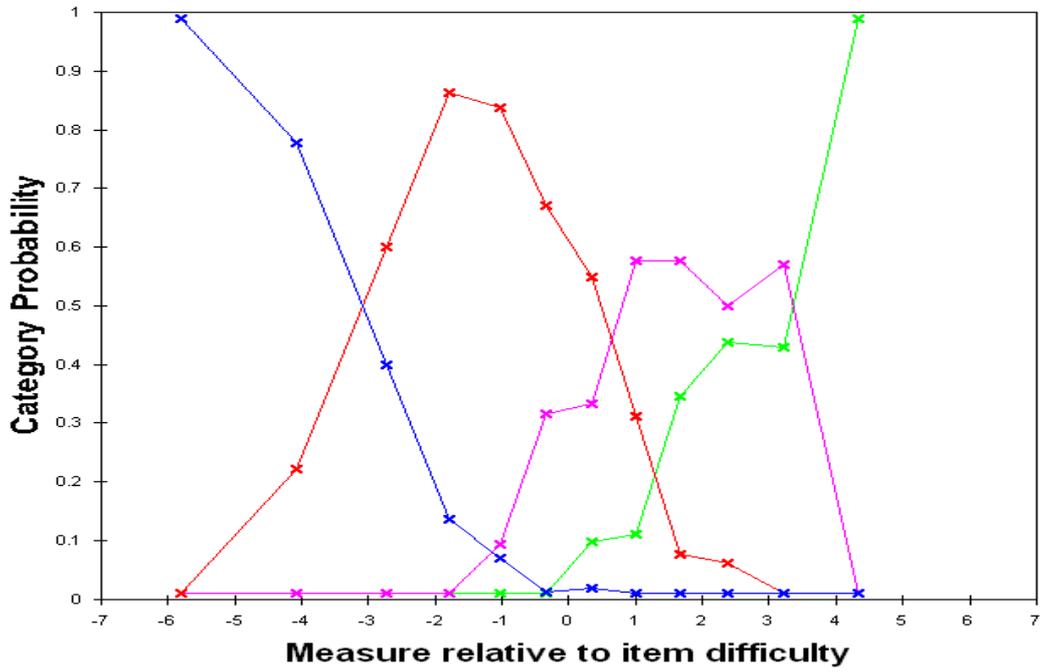
圖三至六則為每一個題目的試題特徵曲線 (item characteristic curve)，此曲線所表達的為每一個得分所得到的機率曲線，例如，圖三為第一個題目的試題特徵曲線，藍色的曲線代表得 0 分的機率分布，當能力值低於-3.5 值，得到 0 分的機率幾乎等於 1，也就是當能力值低於-3.5，這位受試者應該會百分之百在這題得到 0 分，隨著能力值上升，得到 0 分的機率就遞減，若是能力值大於 1，要得到 0 分的機率就會很低了。紅色曲線則為得 1 分的機率曲線，能力在-2 到 -0.5 之間，是最有可能得到 1 分的，得到 1 分的機率大約是 0.8 左右，然而，當受試者能力低於-2，或是高於-0.5，則得到 1 分的機率就會遞減，粉紅色曲線為得到 2 分的機率曲線，最有可能得到兩分的是在能力值 0.5 左右，機率約為 0.65 左右，而綠色曲線則為得到 3 分的機率曲線，為遞增曲線，當能力值越大，越有可能得到 3 分。在理論上，這四條曲線應該是平滑的曲線，而不應有斷層出現，但是在實務上，因為本研究所採用的數據只有 300 多人，因此在畫試題特徵曲線時，會出現斷層。例如，圖五中的得 3 分的曲線（綠色曲線），應該是一個遞增的曲線，但是卻發現在能力值 2.5 的地方突然降了下來，即是因為樣本數不夠大，且得高分的人數不夠多，所以特徵曲線的分布和理論上的預期有些出入。

1.11



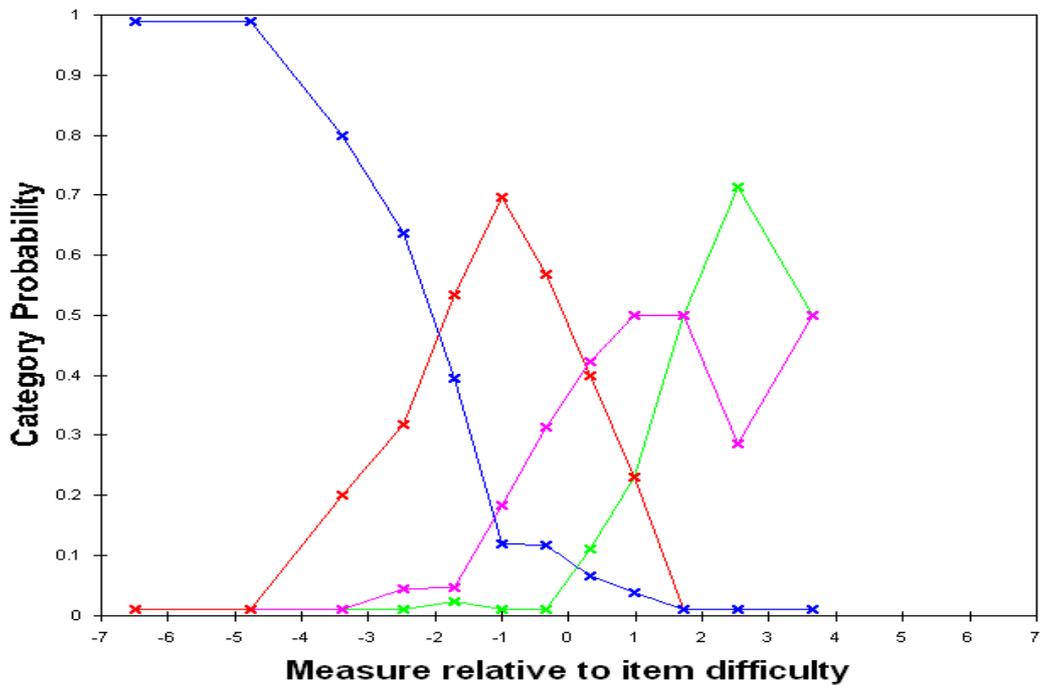
圖三 試題一試題特徵曲線

2.22



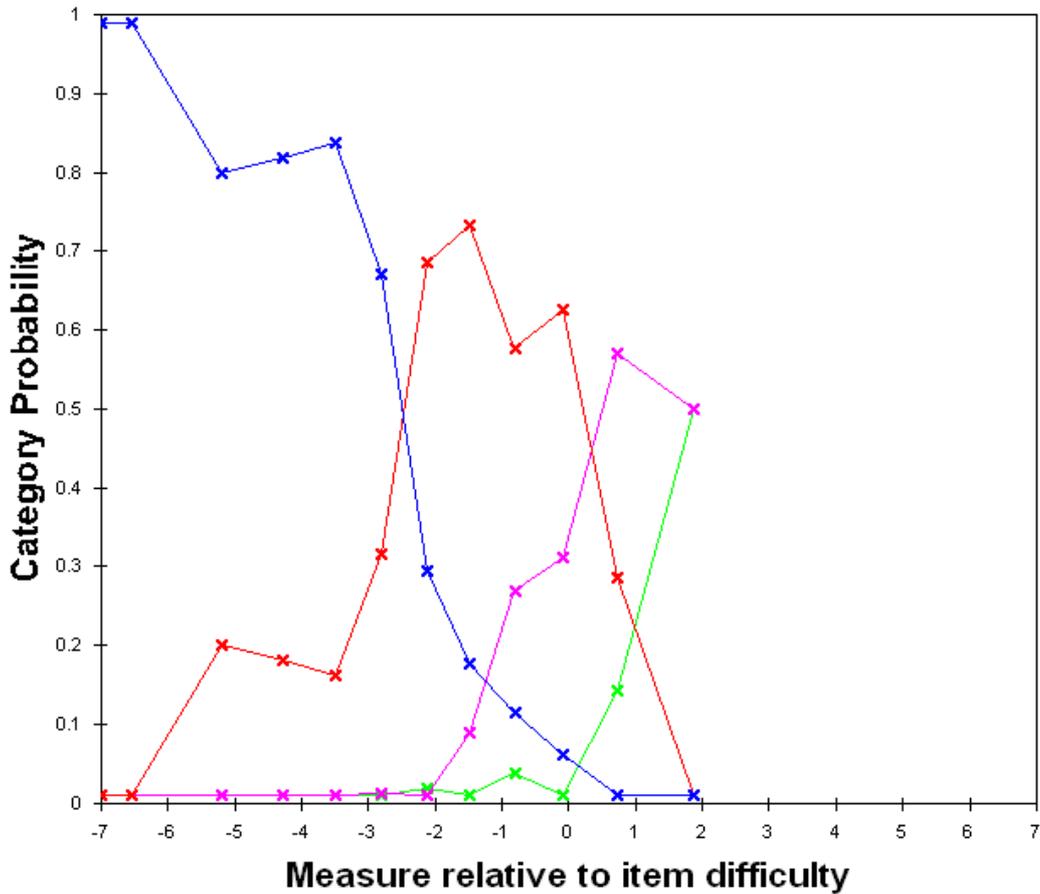
圖四 試題二試題特徵曲線

3.33



圖五 試題三試題特徵曲線

#### 4.44



圖六 試題四試題特徵曲線

## (二) 質的分析

本研究主要在測量於四至六年級學生的數感概念，有四組題目，每一組有三小題，茲將每題之研究成果，分述如下。

### 一、學生於數感概念-電器組

第一題組的題目如下所示：

小丸子準備了 70000 元要購買液晶電視、音響、錄放影機，下列表格為各項商品的價錢。

品項	液晶電視	音響	錄放影機
廠牌 A	39900 元	13000 元	3900 元
廠牌 B	38500 元	15700 元	4100 元

於本題組的設計有三面向

第一子題為小丸子液晶電視、音響、錄放影機要各買一台，請根據小丸子的預算，幫她選好廠牌，並計算價錢是多少？

第二子題為如果小丸子接受你的建議，她還會剩下多少錢？

第三子題為假如小丸子第二題剩下的錢花完了，她又另外準備 8000 元的存款購買影片，她先提出 2500 元買洋芋片，接著又存了 1880 元。請寫出小丸子提款與存款的算式，並算出所餘的存款。

以下依三子題的成果分述之：

(一)、小丸子液晶電視、音響、錄放影機要各買一台，請根據小丸子的預算，幫她選好廠牌，並計算價錢是多少？

學生在填答本問題時，只要答案合理，學生可以選擇同個廠牌或是不同廠牌的答案，於四至六年級的十個班級中，可以看出全班均有 20 人以上均可回答正確，有些同學有計算錯誤的現象，但人數都相當少，如表 5 所示。

表 5 電器組第一子題的解題策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
答對人數	25	31	24	28	24	26	29	29	27	22			
選同一廠牌	12	23	15	7	20	20	24	26	21	8			
選不同廠牌	13	8	9	21	4	6	5	3	6	14			
未作答(空白)	0	0	1	0	0	2	2	1	0	0			
選同一廠牌但計算錯誤	2	1	4	1	6	1	0	1	1	4			
選不同廠牌但計算錯誤	3	0	1	0	1	1	1	0	2	6			
買兩台錄放影機	0	0	0	1	0	0	0	0	0	0			
廠牌 a 電視減廠牌 b 電視	0	0	0	1	0	0	0	0	0	0			
選同一廠牌, 用 70000 去減 剩下的錢就是答案	0	0	0	0	0	2	0	0	0	0			
廠牌 a 的互減	0	0	0	0	0	0	0	0	1	0			
a. b 廠牌全部加起來	0	0	0	0	0	0	0	1	0	0			
合計人數	30	32	30	31	31	32	32	32	31	32			

於本子題的同學解題正確與解題錯誤，可參以下說明。

選同一廠牌	選不同廠牌

選同一廠牌但計算錯誤	選不同廠牌但計算錯誤

(二)、如果小丸子接受你的建議，她還會剩下多少錢？

於本子題，學生需要能使用減法策略進行解題，也就是用子題一的結果，進行減法計算，於本子題，學生於每一班均超過 20 位同學能正確解題。如四年九班，幾乎全班都能答對，相對的，於六年十班，有十人算錯。所以學生並未因高年級而減低錯誤的機率。

表 6 電器組第二子題的解題策略分析

答題分布	年級班別				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
全對	21	31	17	25	24	25	28	28	25	22
錯誤類型: 列式對, 但計算錯誤	3	0	6	3	0	0	2	1	2	0
空白	1	0	2	1	0	2	2	2	1	0
列式錯計算也錯	0	0	1	0	0	1	0	0	0	0
第一題算錯, 所以第二題錯	5	1	4	1	7	2	0	1	3	10
3900-4100	0	0		1	0	0	0	0	0	0
第一題的答案寫到第二題	0	0	0	0	0	2	0	0	0	0
合計人數	30	32	30	31	31	32	32	32	31	32

有關學生解題的部份說明，如照片所示。

全對	錯誤類型: 列式對, 但計算錯誤

(三)、假如小丸子第二題剩下的錢花完了，她又另外準備 8000 元的存款購買影片，她先提出 2500 元買洋芋片，接著又存了 1880 元。請寫出小丸子提款與存款

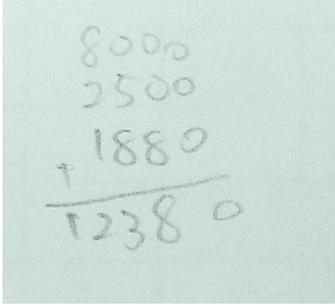
的算式，並算出所餘的存款。

於本子題，學生需能理解加減法題意後，進行解題，由表 7 可看出，四至六年級學生，每班都有 20 位學生以上，都能列出算式，能得出正確答案，每班均有 15 位以上。其他錯誤類型的答案，有些學生使用連減的策略，這是誤解題意的，均出現在四至六年級學生。

表 7 電器組第三子題的解題策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
先減後加	21	17	21	15	17	16	21	22	19	19			
列式對但計算錯誤	6	7	3	3	2	5	3	1	4	3			
先加後減	0	0	0	1	0	1	0	0	0	0			
連加	0	0	2	0	0	0	0	0	0	0			
連減	0	6	1	4	6	5	3	4	2	5			
未作答(空白)	1	1	1	3	1	4	2	2	3	0			
其它錯誤類型:2500+8000	0	0	0	1	0	0	0	0	0	0			
把第二小題剩下的算進去,先減後加	2	1	2	4	5	1	2	0	2	4			
用第二小題剩下的錢去減第三題答案	0	0	0	0	0	0	1	0	0	0			
把第二小題算進去,連減	0	0	0	0	0	0	0	1	1	0			
把第二小題算進去,先加後減	0	0	0	0	0	0	0	0	0	1			
2500-1800	0	0	0	0	0	0	0	1	0	0			
8000-2500	0	0	0	0	0	0	0	1	0	0			
合計人數	30	32	30	31	31	32	32	32	31	32			

有關學生答對與答錯的情況，請參下列照片說明。

$8000 - 2500 = 5500$ $5500 + 1880 = 7380$ $A = 7380 \text{ 元}$ $\begin{array}{r} 8000 \\ - 2500 \\ \hline 5500 \end{array}$ $\begin{array}{r} 5500 \\ + 1880 \\ \hline 7380 \end{array}$	
<p>先減後加</p>	<p>連加(列式錯誤)</p>
$8000 + 2500 - 1880$ $= 10500 - 1880$ $= 8620$ $A = 8620 \text{ 元}$ $\begin{array}{r} 10500 \\ - 1880 \\ \hline 8620 \end{array}$	<p>請寫出小丸子提款與存款的算式，並算出所餘的存款</p> $8000 - 2500 = 5500$ $5500 + 1880 = 7380$ $A = 7380 \text{ 元}$
<p>先加後減(列式錯誤)</p>	<p>連減(列式錯誤)</p>

## 二、學生的數感概念-聰明的柯南

第二題組的題目如下所示：

柯南做了一份學生飯後水果調查，他問學生：「你們飯後比較喜歡吃香蕉、百香果還是芭樂？」

(一)、你覺得柯南的問題是否恰當？請說明你贊成或否定的理由？

(二)、請問男女生分別最喜歡吃哪一種水果？哪一種水果最受全班學生的歡迎？

(三)、柯南記錄下統計的結果，如表格所示，請根據柯南的統計數據，幫他製作一個簡單易懂的調查統計圖。

(四)、請依照這個統計圖，設計出一個數學問題來問同學

以下為各子題的結果分析：

(一)、你覺得柯南的問題是否恰當？請說明你贊成或否定的理由？

於本子題的目的是要檢視學生是否對於本題目的邏輯題出問題，不論是贊同或反對均可。可看出四年級學生的表現與五、六年級的表現相當，每班均有約20人以上同學答對。於此可看出，班級同學對於邏輯批判的題型均有能力回答。於贊成者的理由與反對者之理由，請參表8。

表8 聰明的柯南第一子題的解題策略分析

年級班別 答題分布	四年級				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
答對者(贊成與反對者)	17	24	24	23	25	21	16	23	20	26
贊成且回答得有道理 合邏輯 (贊成總數)	(10)	(17)	(11)	(16)	(17)	(15)	(11)	(12)	(11)	(17)
1. 可以尊重到每個人 不同選擇	2	1	1	6	2	3		3	1	1
2. 可以讓水果不會被 丟掉				1						1
3. 可以知道他們到底 有沒有吃水果				2						
4 因為飯後吃水果很 健康	2	1	5	6			3		2	2
5 因為我不喜歡飯前 吃水果				1						
6 可以知道每班每人	5	13	4		10	5	4	2	4	2

喜歡吃的水果										
7. 少數服從多數		1								
8. 因為班上同學是奇數, 一定會有最後一個人來當最後一票		1								
9. 可以看學生喜歡吃的和老師一不一樣					1					
10. 可以在特別節日送他喜歡的水果					1					
11. 以後就知道水果怎麼分					1					
12. 可以讓大家愛吃水果					2	3				
13. 可以讓學生開心						2				
14. 這樣才知道要給學生吃什麼						1				
16. 因為我就很喜歡吃香蕉						1	1			
17. 這種方法很清楚								1		
18. 因為這樣比較好算							1			
19. 因為這樣價格比較便宜							1			
20. 這樣可以知道學生的健康知識							1			
21. 這樣可以知道哪種水果好吃	1									
22. 有些水果吃多了不好			1							
23. 問題不矛盾(問題很清楚)								6	3	4
24. 問題不難回答									1	
25. 只列出三種好, 因為水果太多了										1
26. 因為水果很好吃										6
<b>反對且回答得有道理合邏輯(反對總數)</b>	(7)	(7)	(13)	(7)	(8)	(6)	(5)	(11)	(9)	(9)
1. 因為有人喜歡吃的	1	4	7	2	3	3	4	6	3	5

不是這三項										
2. 因為有人喜歡其中好幾樣			2							
3. 不能這樣要求廚師樣樣都準備			1	1						
4. 覺得圖案很怪			2							
5 不能用問的要舉手表決		1								
6. 可以把題目再寫簡單一點		1		2						
7. 有人可能會寫跟好友一樣的, 有人會因為喜歡的人沒說所以他也不敢說		1								
8. 因為會消化不良 (應該飯前				2			4		2	
9 因為這樣浪費時間										
10 這樣浪費錢					1					
11 這樣代表他們挑食	1		1		1		1			
12 水果種類應該要讓學生提名					1					
13 因為同一種水果可能有不同口感						1				
14 如果有人不喜歡吃最高票的水果就會把水果藏起來不吃	1									
15 沒有意義	1								2	
16 很難理解	1									
17 有的統計被蓋住										
18 有的人是飯前吃水果	1		2						2	
19 很容易搞混			1							
20 應該問最喜歡吃的水果			1							
21 吃完飯很難吃芭樂, 其他兩樣比較少			1							
22 不贊成也不否定, 不否定是因為我喜歡									1	

吃巴樂, 不贊成是其他的我都不喜歡										
23 因為調查有分男生和女生									1	
24 問題和圖的標題不一樣										2
有回答贊成與否, 但沒寫原因	2	2	1	4	3	6	1	2	3	1
文不對題							4			
贊成, 因為它標示(統計)得很清楚		2	3			1	7			4
贊成, 因為香蕉比較多	1									
贊成, 因為他很認真記錄下來									1	
反對, 因為少了統計圖的名稱								1		
反對因為他們有統計全班								1		
空白	10	4	2	4	1	4	4	5	7	1
其它										
合計人數	30	32	30	31	31	32	32	32	31	32

(二)、請問男女生分別最喜歡吃哪一種水果? 哪一種水果最受全班學生的歡迎?

於本子題, 由圖表中發現答案, 每班均有 20 人以上的同學答對, 可看出學生在讀圖表的能力均有 60% 以上, 如表 9 所示。

表 9 聰明的柯南第二子題的解題策略分析

答題分布	四年級				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
全部答案都答, 且都答對	26	24	20	21	26	27	26	30	25	28
全部都答, 但部份答錯			8	1			2		1	
只有回答部份答案, 但答對	2	7		7	3	5	2	1	1	3
只有回答部份答案, 但答錯	1		1	1	2					1
全部答錯									1	
空白	1	1	1	1			2	1	3	
其它										

合計人數	30	32	30	31	31	32	32	32	31	32
------	----	----	----	----	----	----	----	----	----	----

(三)、柯南記錄下統計的結果，如表格所示，請根據柯南的統計數據，幫他製作一個簡單易懂的調查統計圖

於本子題，只有四年九班與六年級的學生對於長條圖的描繪相當清楚，其它四年級班級與五年級學生均不到十位學生可以畫出折線圖。於此可看出，教師在四年九班有進行過長條圖的教學，其它學生只有在六年級有教。於此也可看出折線圖於亦可於四年級時進行，如表 10。

表 10 聰明的柯南第三子題的圖表策略分析

答題分布	年級班別				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
長條圖	0	26	3	2	6	6	8	22	23	26
折線圖	1	0	0	0	1	0	0	0	1	2
使用代號作記號統計(正字)	4	1	1	6	2	1	1	0	0	0
使用代號作記號統計(人形.打叉打勾 圈圈.星星)	0	2	0	0	2	1	3	0	0	1
使用數字統計	4	3	15	7	5	4	7	2	0	0
創意圖	0	0	0	0	3	0	0	1	0	0
錯誤類型:照抄原表	0	0	4	4	2	1	1	0	0	0
空白	21	2	7	12	10	19	12	7	7	3
合計人數	30	32	30	31	31	32	32	32	31	32

有關學生長條圖的繪製請參下列照片說明。

<p>第他製作一個問平均值的統計圖</p> <p>14 (4)請依照這個統計圖，設計出一個數學問題來問同學</p>	<p>14 (4)請依照這個統計圖，設計出一個數學問題來問同學</p>
<p>14</p>	<p>14</p>

使用代號作記號統計(正字)	創意圖

(四)、請依照這個統計圖，設計出一個數學問題來問同學

於本子題屬於創造性的學習問題，要求學生依圖表來出一個問題，學生在四年四班的標電與六年三班的表現均偏低。於四年九班、四年十班、五年一班與六年二班的表現均有十位以上學生做對，可能與班上老師的教學有關。有很多的學生空白，可能對本子題不了解，或是不知如何作答，學生在創造的面向訓練，應要再加強。

表 11 聰明的柯南第四子題的策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
有按照題目要求出題	4	12	13	8	13	7	9	16	5	10			
出的是數學題目但不合統計圖			2										
合統計圖但出的不是數學題目	2			4	3		2	1	4	1			
空白	24	17	13	17	14		19	15	22	20			
另外自己出題目		3	2	2	1	25	2			1			
其它													
合計人數	30	32	30	31	31	32	32	32	31	32			

### 三、學生的數感概念-哈利波特

第三題組的題目如下所示：

哈利波特、榮恩和妙麗一起搭火車到士林逛夜市，他們所搭的火車原本預定晚上 6:15 抵達，但因為一些原因遲了 25 分鐘。

- (1) 哈利波特一行人吃了幾份美食後，就直接搭車回家，總共花了 2 小時 55 分才安全到家。請畫一個時鐘來表示他們回家的時間。
- (2) 妙麗打算明天的上午要用不超過 5 小時的時間，做一些她喜歡的活動。活動所需要的時間如下面的表格所示，有「\*」記號的表示她一定要做的活動。請問妙麗還可以在這個時間安排哪些活動呢？
- (3) 承上題，請用自己的方式幫妙麗製作一個明天上午的活動時間表，(從上午七點開始，包含各項活動的開始、持續時間以及結束時間)。

以下為各子題的結果分析：

- (一)、哈利波特一行人吃了幾份美食後，就直接搭車回家，總共花了 2 小時 55 分才安全到家。請畫一個時鐘來表示他們回家的時間。

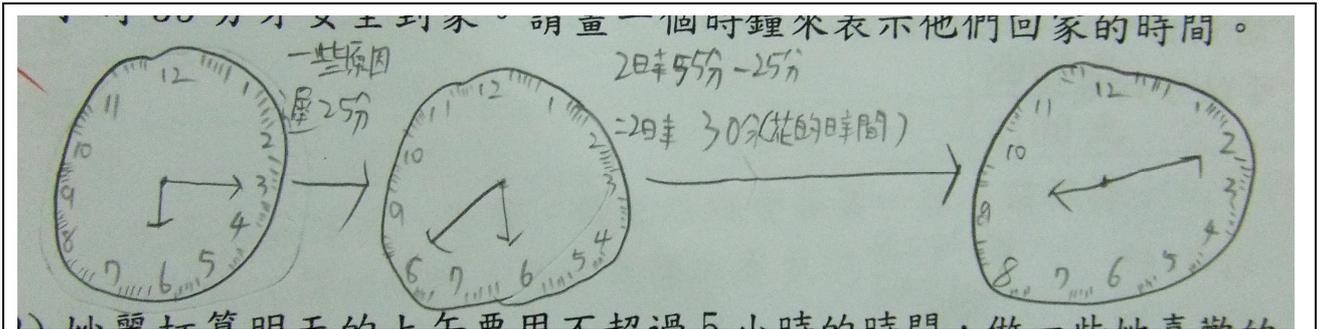
於本子題，其目的在於學生需計算與繪製時鐘的圖形，來描述出正確的時間，屬於時間的理解與應用。於表八可看出學生在答對的學生，並且計算正確的學生，四至六年級每班不到十人，可看出時間的繪製與計算對學生而言相當有難度，有些學生計算正確，繪製時鐘錯誤。有些只會以計算算出時間，但不會繪製。這些問題可能源自於教學沒有使用時鐘教具的結果。

表 12 哈利波特第一子題的解題策略分析

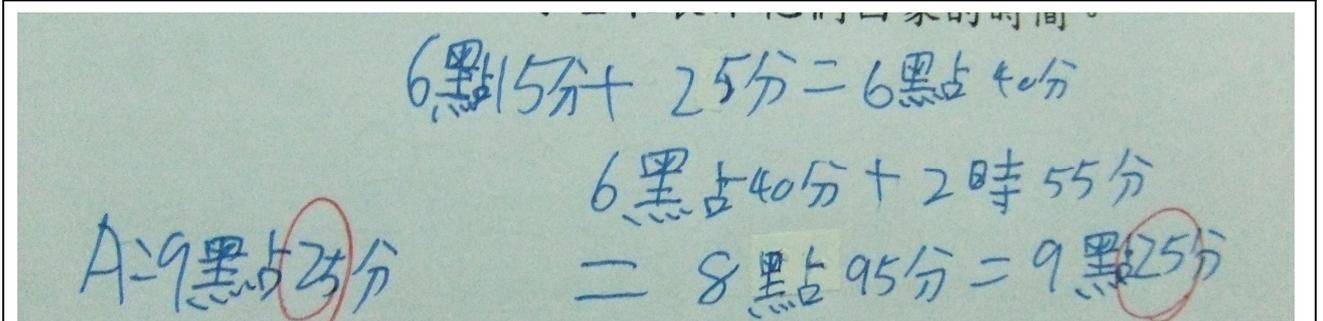
年級班別	四年級				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
答題分布										
正確總數	9	5	4	4	5	5	4	8	7	7
僅以算式標示正確的時間	0	4	0	0	1	0	0	0	0	0
僅以時鐘標示出正確的時間(時針有小誤差)	5(0)	(6)	1(4)	1	2(4)	(2)	1	2(2)	0(1)	1(6)
按題意做時間的加減計算無誤，且在時鐘標示出正確的時間(短針有小誤差)	4(3)	1(11)	3(5)	3(8)	2(7)	5(11)	3(12)	6(10)	7(9)	6(8)
未作答(時間標示錯誤)	4(9)	(6)	(12)	10(4)	(11)	4(8)	2(3)	2(3)	3(4)	0(5)
題意理解錯誤所造	1	1	0	1	2	2	3	1	3	2

成列式的錯誤										
時間加減時進退位的錯誤	3	3	5	4	2	0	5	4	4	3
時間抄錯造成答案的錯誤	1	0	0	0	0	0	3	2	0	1
其它	0	1	0	0	0	0	0	0	0	0
合計人數	30	32	30	31	31	32	32	32	31	32

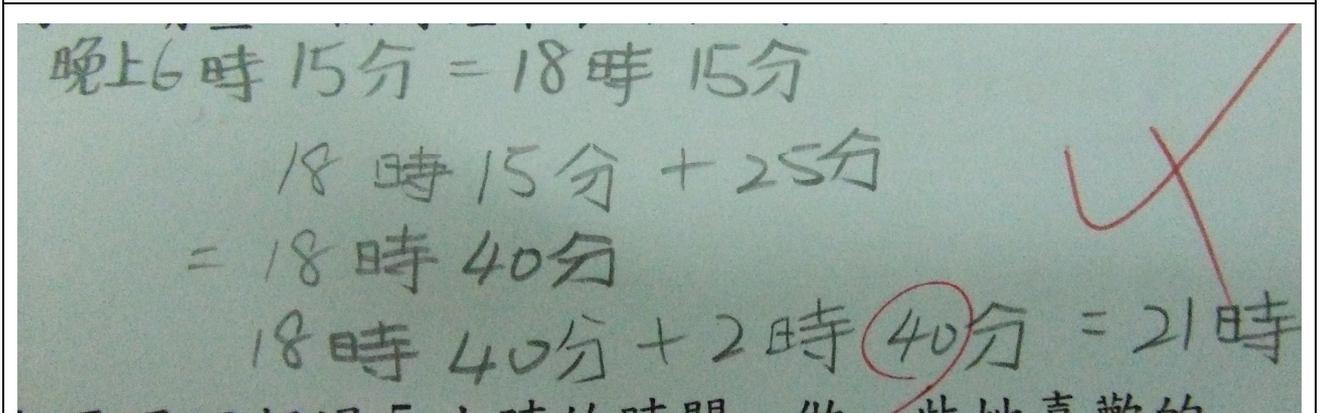
說明：



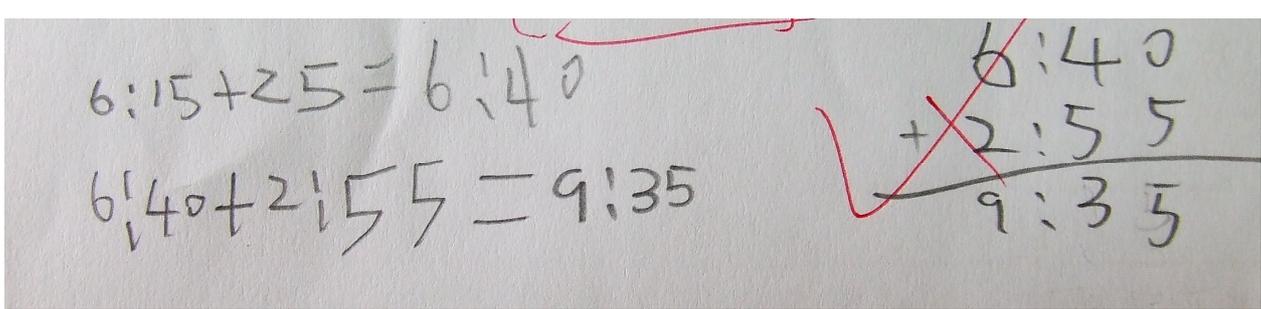
題意理解錯誤所造成列式的錯誤



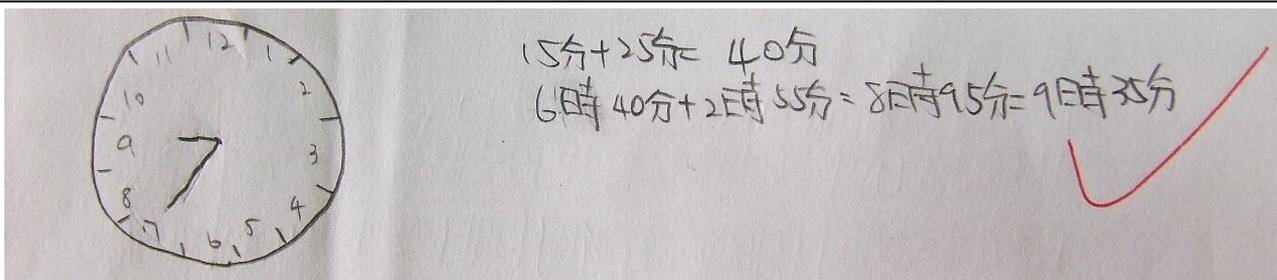
時間加減時進退位的錯誤



時間抄錯造成答案的錯誤



僅以算式標示正確的時間



按題意做時間的加減計算無誤，且在時鐘標示出正確的時間(短針有小誤差)

(二) 妙麗打算明天的上午要用不超過 5 小時的時間，做一些她喜歡的活動。活動所需要的時間如下面的表格所示，有「\*」記號的表示她一定要做的活動。請問妙麗還可以在這個時間安排哪些活動呢？

本子題的目的在於請學生規劃出必需辦理的活動，及可以再加上的活動，屬於分析與評判的層次。學生可以用活動與活動相加的策略，用五小時累減的策略、或是用加減混合的方式來得出合理的答案。表 13 即為各班所使用的策略人數。於四年級，僅四年九班的人數超過十人以上者正確，在五年級與六年級學生人數均有十人以上，可看出高年級學生在本活動的反應較四年級為佳。在解題策略上來看，用累加型的人數明顯較高，用累減型的人數每班均少於 3 人，明顯可看出累減的策略較有創意。於本子題的答對率也約只有 50%，即如這樣開放性的問題，四、五、六年級學生的表現，全班有一半學生的學生能夠答對。

表 13 哈利波特第二子題的解題策略分析

年級班別	四年級				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
答題分布										
答對人數	5	16	8	5	13	14	22	15	17	17
累加型	2	13	7	4	12	12	16	11	9	14
累減型	3	1	0	0	1	2	0	2	2	1
加減混合計算	0	2	1	1	0	0	6	2	6	2
未作答(未完成)	8(9)	1(9)	8(4)	16(5)	7(4)	6(6)	2(4)	6(4)	1(6)	1(6)
誤解題目	5	5	8	0	1	0	1	0	0	4
累加型進位錯誤	2	1	1	4	5	1	2	5	5	3

累減型退位錯誤	0	0	1	1	1	2	1	1	0	0
加減混合計算錯誤	1	0	0	0	0	3	0	1	2	1
合計人數	30	32	30	31	31	32	32	32	31	32

有關學生計算的錯誤與計算的正確部份，請參下列的照片說明。

說明：

誤解題目	(累減型)退位錯誤
累加型	(累減型)
加減混合計算	

(3)承上題，請用自己的方式幫妙麗製作一個明天上午的活動時間表，(從上午七點開始，包含各項活動的開始、持續時間以及結束時間)。

本子題的目的是要求學生能製作一個活動時間表，屬於創造性的層次。於本

子題的表現，四年級學生均降到五人以下，高年級為五至十人左右的達成率。於本題可看出學生在高年級有較良好的時間規劃能力，但四年級可能就弱了一些。學生有部份只填開始與結束的活動，但中間沒有填寫，於此部份亦給分，但較完整的答案應要列出完整的時間表，才較為完整。

本題未作答的比率佔 50%，如四年四班與四年十一班，但四年九班與十班均低於五人，相對的，大多人均自創活動，有 50% 以上，比率較其他班級為高，有可能老師的教學較為開放，但題目並非要學生自創，所以兩班學生有可能誤解本題的想法。

表 14 哈利波特第三子題的解題策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
答對人數	4	4	3	1	3	6	7	7	8	6			
僅開始至結束	2	4	2	0	3	3	6	5	6	5			
完整時間表	2	0	1	1	0	3	1	2	2	1			
未作答(未完成)	13(2)	3	1(1)	23	17	10	18	21	10	7(3)			
僅時間的加減，未按題意安排	3	3	4	1	3	7	4	0	5	8			
自創額外活動	7	17	20	6	5	5	1	1	3	5			
時間表內持續時間的錯誤	1	5	0	0	2	0	2	2	5	2			
僅註明開始時間或(結束時間)	0	0	1	0	(1)	4	0	1	0	(1)			
其它	0	0	0	0	0	0	0	0	0	0			
合計人數	30	32	30	31	31	32	32	32	31	32			

有關本題的說明，請參如下照片所示：

活動內容	持續時間
去朋友家	10分
*早上游泳	40分
吃早餐	30分
書展	1小時
屈臣氏	45分
看電視	30分
打掃	25分
看書	1小時
爬山	1小時
聊天	10分
回家	十二點

時間	事情
7:00-7:40	閱讀
7:40-8:30	游泳
8:30-9:15	購物
9:15-9:30	打掃
9:30-11:05	書展

自創額外活動

時間表內持續時間的錯誤

7點	早上游泳
9:50	書展
9:25	購物
10:10	打掃
10:35	閱讀
11:15	結束時間

時間	事情
7:00~7:50	早上游泳
7:50~9:25	書展
9:25~9:50	打掃
9:50~11:10	爬山
11:10~11:55	購物

僅註明開始時間(結束時間)

含開始至結束

活動	開始時間	持續時間	結束時間
晨跑	7:10	50 min	8:00
打掃	8:20	30 min	8:50
看書	8:50	30 min	9:20
逛街	9:20	1hr 30 min	

完整時間表

#### 四、學生的數感概念-慶生會

第四題組的題目如下所示：

班上舉辦慶生會，老師準備 48 塊餅乾，全部分給好幾位小朋友，如果每個小朋友最少得到 4 塊，且每個人拿到的餅乾數都要相同。

- (1)請列出可以有哪幾種分法(如可以分給幾個小朋友?每位小朋友可以分到幾塊?)
- (2)承上題，請把上面分配的方法，用乘法和除法的算式寫出來，並用文字或圖畫說明你所寫的除法算式恰好為乘法算式的相反運算。
- (3)承第一題，如果老師又加入 8 塊餅乾，請列出可以有哪幾種分法?

以下為各子題的結果分析：

- (一)、請列出可以有哪幾種分法(如可以分給幾個小朋友?每位小朋友可以分到幾塊?)

於本子題，其目的在於要求學生找出各種可能的分法，屬於因數概念的分析層次。如表 15，因本題可能過於困難，因為因數問題於四、五年級可能尚未較導，故解題不完整，所以只有六年級有較多的同學答對。四、五年級多有學生提出有一種或兩種以上的分法，也就是並未完全答對。於本子題，要注意的是，四年十一班學生有 25 人未答題，比率相當高，可能全班大多數學生不懂題意有關。

表 15 慶生會第一子題的解題策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
全部的分法，已扣除題目所要求的每位小朋友最少得到四塊的分法	0	1	1	0	0	0	1	6	4	5			
未答題	4	8	1	25	12	7	7	7	7	14			
只有單一分法	21	10	5	3	4	14	5	3	5	2			
兩種以上的分法	5	13	20	3	9	8	16	15	12	10			
全部的分法，但未扣除不符題目所要求的三樣分法：	0	0	3	0	6	3	3	1	3	1			
其它	0	0	0	0	0	0	0	0	0	0			
合計人數	30	32	30	31	31	32	32	32	31	32			

有關學生解題的策略，如何分配的方式，請參照片說明：

<p>全部的分法，但未扣除不符題目所要求的三樣分法：</p>	<p>全部的分法，已扣除題目所要求的每位小朋友最少得到四塊的分法</p>

(二) 承上題，請把上面分配的方法，用乘法和除法的算式寫出來，並用文字或圖畫說明你所寫的除法算式恰好為乘法算式的相反運算。

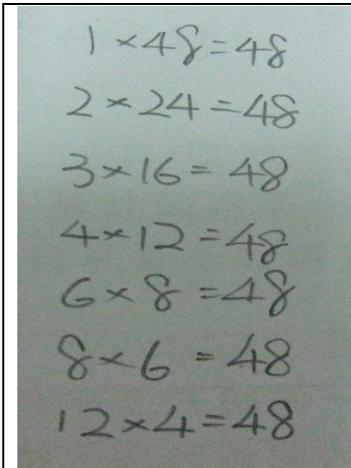
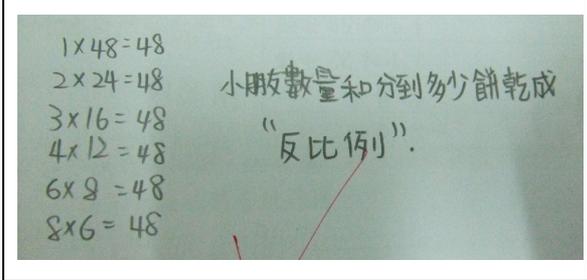
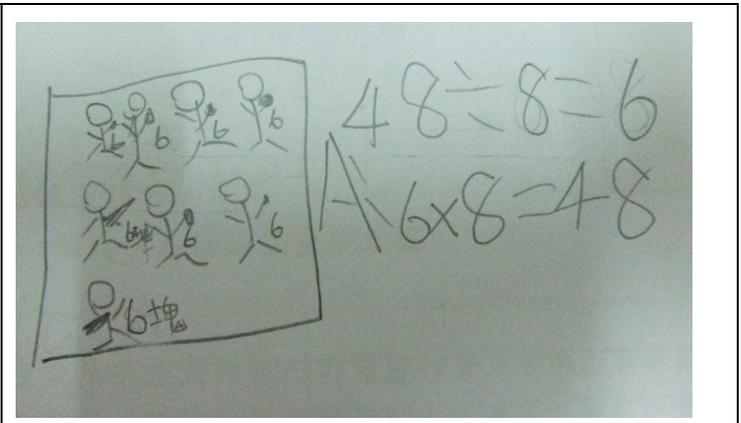
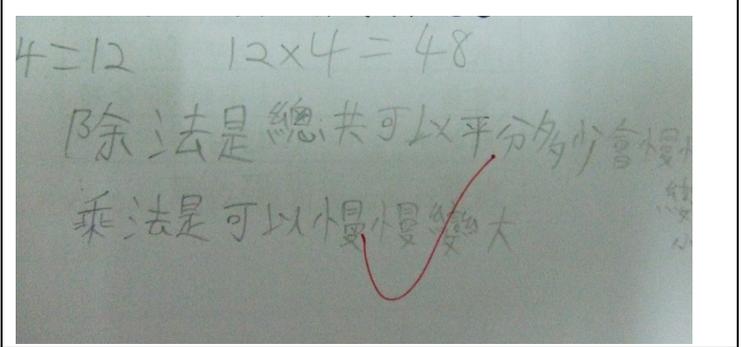
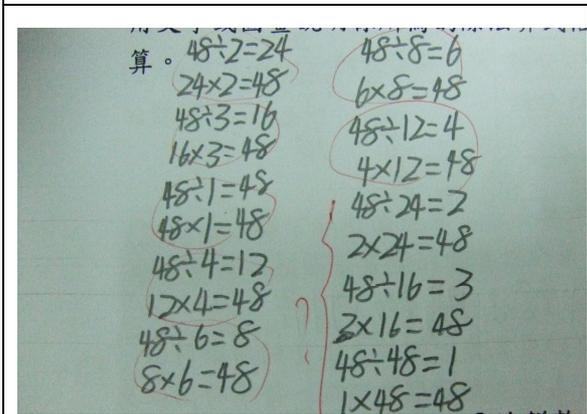
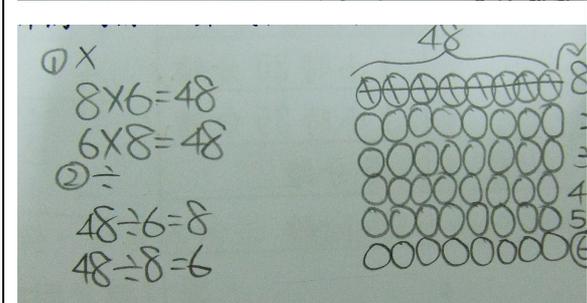
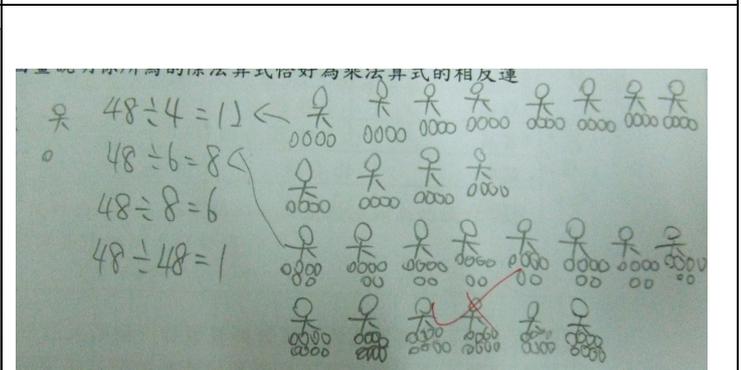
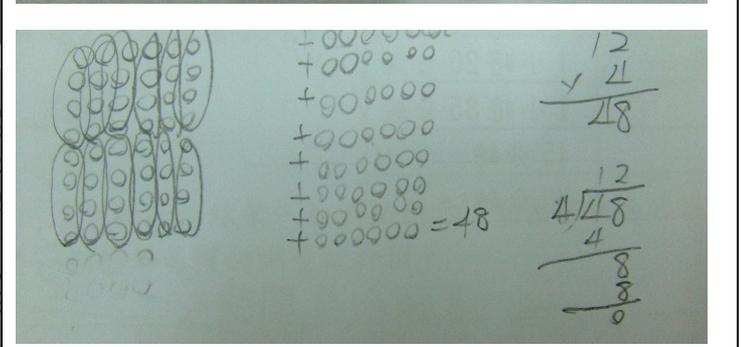
於本子題，其目的在於學生需發現乘法與除法互逆的關係，屬於創造與發現層次。所以學生若對於乘除法關係不了解，僅可能停留在計算的層次上。如表 16，不論學生應用圖示，或是用乘法或除法算則表示其相反的運算關係，都可以列為合理的答案。特別的是，於四年九班與四年十班，同學對於乘除法的關係，其表現較高年級為佳，甚至較五年級與六年級同學為佳，可能與班上的教學方法有關。

表 16 慶生會第二子題的解題策略分析

答題分布	年級班別				五年級			六年級		
	404	409	410	411	501	502	505	602	603	610
答對人數(1+2+3+4+5)	11	19	20	1	8	14	13	19	9	11
1. 單以乘法表示說明	3	7	5	0	2	5	3	5	2	2
2. 以乘法和除法算則表明一種	8	12	5	0	1	4	2	4	2	4
3. 分法的相反運算										
4. 以乘法和除法算則表明至少兩種以上分法的相反運算	0	0	10	1	4	4	8	10	4	5
5. 以圖畫標示	0	2	0	0	1	1	0	0	1	0
未答題	17	11	5	30	23	18	19	13	13	20
單以除法表示說明	2	0	5	0	0	0	0	0	9	1
其它	0	0	0	0	0	0	0	0	0	0

合計人數	30	32	30	31	31	32	32	32	31	32
------	----	----	----	----	----	----	----	----	----	----

有關學生在乘法與除法關係的解題策略，有關解題策略，請參照片說明，如下所示。

 	 
<p>單以乘法表示說明</p>	<p>以乘法和除法算則表明一種分法的相反運算</p>
 	 
<p>以乘法和除法算則表明至少兩種以上分法的相反運算</p>	<p>以圖畫標示</p>

(三)、承第一題，如果老師又加入 8 塊餅乾，請列出可以有哪幾種分法？

於本子題，解題又更為困難，當情境改變時，本題加入應用、分析與評判等層次。故於本子題，學生答對的比率均在六年級，且每班人數都在十人以下，如表 17。大多學生均未答題，或是僅提出一種或兩種分法。特殊的是，四年十班提出兩種分法的人數高達 19 人，比其它班級與年級均高。可能老師的教學方式，對學生的學習產生明顯的改變。

表 17 慶生會第三子題的解題策略分析

答題分布	年級班別				四年級			五年級			六年級		
	404	409	410	411	501	502	505	602	603	610			
全部的分法，已扣除題目所要求的每位小朋友最少得到四塊的分法	0	0	0	0	1	0	1	6	6	4			
未答題	12	13	4	29	20	15	17	15	14	20			
只有單一分法	13	6	7	1	1	9	3	1	1	1			
兩種以上的分法	5	4	19	1	4	6	8	9	8	4			
全部的分法，但未扣除不符題目所要求的兩種分法：													
其它	0	9	0	0	5	2	3	1	2	3			
合計人數	0	0	0	0	0	0	0						
	30	32	30	31	31	32	32	32	31	32			

說明：

$$\begin{array}{r} 48 \\ + 8 \\ \hline 56 \end{array}$$

$56 = (2, \cancel{4}, 7, 8, 14, 28, 56)$

A(8)種重

6

$$48 + 8 = 56$$

- 1  $56 \div 1 = 56$
- 2  $56 \div 2 = 28$
- 3  $56 \div 4 = 14$
- 4  $56 \div 7 = 8$
- 5  $56 \div 8 = 7$
- 6  $56 \div 14 = 4$

A = 6種

全部的分法，但未扣除不符題目所要求的兩種分法：

全部的分法，已扣除題目所要求的每位小朋友最少得到四塊的分法

## 伍、結論

由量化的分析可見，此研究所用的四個實作評量有難有易，是一份能夠測出各種實力學生的考題，評分者、試題的信度值也在.90 以上。然而，雖然本研究只使用四個題組，並樣本集中在四至六年級學生，但可以發現到學生在一些開放試題中所展現的數學概念。以下依四個題組的概念分別論述。

第一，有關電器組的評量目的為要求孩子用他自己的策略選擇後，再進行計算，在此部分的成就表現，幾乎全班都有 25 人答對。雖然子題二，進行了減法運算，學生也都能回答正確。於第三子題，雖出現加減法的混合運算，每班也約有 20 位學生都能列出正確的算式，進行解題。由此可看出，四至六年級學生在基本計算的數概念部份有一定的學習水準，學生亦能接受此種題型的出題方式。

第二，有關聰明的科南，目的要學生能覺得問題是否適當，能提出理由、判斷圖表、製作長條圖與設計問題來問同學。學生在第一子題提出很多不同的理由來贊同或反對這一問題，不論學生的理由如何，只要合理就視為正確的答案。學生能思考出不同的理由，如「因為同一種水果可能有不同口感」，「如果有人不喜歡吃最高票的水果就會把水果藏起來不吃」等等，有大約 30 種不同的原因，可以發現學生的想法其實是很多元的。在統計圖表的製作部份，四年九班學生的表現與六年級一致，可以說明老師可能已教了長條圖的繪製，或是老師運用了不同的教學方式。於此也可看出，數學內容只要能運用適當的教學方式，孩子在學習數學的內涵是可以提升的。於此，也對課本將長條圖延後至六年級的適當性，提出證據，學生可於四年級學習，而不需延到六年級。最後，要求學生由圖表來想出一個問題，有一班約半數同學都能提出，學生表現均相當優異。如學生用折線圖，用正字標記標示等，這都是相當合理的表達方式

第三，於哈利波特題組的目的在於測量孩子在時間的量感。在第一子題中發現，學生在時鐘的標示部份，出其意料的相當薄弱，一班少於十人答對。可能與學生沒有進行撥鐘的實作有關。如果學生沒有進行實作，即使學生到高年級，觀念還是不清楚。結果也發現，即使學生會計算時鐘，但是不會撥鐘，對時鐘的認識是不紮實的。在第二子題，請同學設計一套行程，只要答案合理就可以，其結果反而較第一題答案為佳，五年級與六年級的成就表現相當好，四年九班的學生表現幾乎和高年級一樣。所以學生在創意策略的表現，相當不錯，這樣的題目也的確可以看到學生對時間的安排相當有主見。最後請學生自行設計一個上午的時間表，學生答對的表現也的確不多，每班約 5 人。因為本題有很多學生產生誤解，誤認為自己可以安排行程，也就是未來在設計題目時要更加明確。但若日後學校考試能針對此一方向出題，學習

表現應會更佳。

第四，慶生會的目的在於測量學生對於因倍數概念與發現乘除法的關係。由於因倍數是在高年級的教學，所以預期學生在四年級時，應尚未教到本單元。然而，在學生的表現中，學生未回答的比率不高，只有四年級一班的學生有 25 人，其它班級的表現都相當優質，如有提出一種的分法，二種以上分法者。雖然全對者人數不多，但可看出學生能夠努力回答這樣的開放性問題。再者，學生能發現乘除法的關係等問題，四年級至六年級學生的表現依班級而有不同。在六年級的表現亦未特別突出，可能與教師的教學有關。在第三子題的難度雖然增加，學生仍可以用分法的方式進行解題，雖然答對人數不多，但可鑑別出較高層次的學生。

綜合上述，由四題組的題目設計，在數的計算、圖表、時間與乘除法的關係面向，有別於選擇題的方式，可看出不同學生的解題策略，而學生所展現的一些錯誤的解題策略，除了可以做為認知診斷測驗的基礎與參考，對於未來國內要進行開放式的題目，也提供了一個可實踐的方向。

## 陸、建議

本研究的成果不論是從多層面 Rasch 量化分析、多分制的試題分析，及從學生質性的解題策略分析來看，均可提供國內的 TASA 等重大型考試參考，在本研究的過程中，仍有些建議需要提出。

第一，開放性試題的評分機制為四等分制，即從 0 到 3 分，在試題分析的統計來看，並不是問題，但不同評分者的嚴謹度，會出現些微的差距。即評分者的訓練要注意。

第二，開放性試題的質性分析需要很多時間，是否要將每位學生的答案均列出，可能有其限制，所以可能未來在大量施測時，僅需部份列出學生的答案即可。

第三，開放性試題的設計要能符合目前國內的數學課程，在前測時，因為學生對題目的意義不了解，所以於這次的施測中才出現更佳的结果。

第四，學生的表現可看出班級教師教學的差異，所以雖然學生在常態分班下，在教師教學後，出現各班差異很大的情況，這樣的情況，需回報給教師，進行教學面的改善。

最後，本實作評量的結果發現，學生在開放性問題的表現相當多元，可以測量出學生在數學方面的能力，期望在未來的大型考試中，看到這樣有深度的題目在國內中出現。

## 參考文獻

- 王文中 (2004)。Rasch 測量模式與其在教育與心理之應用。《教育與心理研究》，27(4)，637-694。
- 姚漢禱、姚偉哲 (2007)。應用多層面Rasch模式分析雙不定向飛靶優秀選手的射擊技術。《測驗學刊》，89-104。
- 國立編譯館 (1994)。《國民小學數學教科書》。台北：作者。
- 張麗麗 (2002)。從分數的意義談實作評量效度的建立。《教育研究月刊》，98，37-50。
- 張新立、吳舜丞 (2008)。多層面Rasch 模式於學術研討會論文評分之應用。《測驗學刊》，105-128。
- 曾安如 (2004)。《國小二年級學童數學寫作活動、數學成就與數學態度之相關研究》。國立台中師範學院教育測驗統計研究所碩士論文，未出版，台中。
- 詹元智 (2002)。《國小數學科實作評量之效度探討》。屏東師範學院教育心理與輔導研究所碩士論文，未出版，屏東。
- 蔡正濱 (2006)。《國小數學科實作評量評分者一致性相關因素探討》。國立屏東教育大學教育心理與輔導學系碩士論文，未出版，屏東。
- Alajmi, A. (2004). *Eighth grade Kuwaiti students' performance in recognizing reasonable answers and strategies they use to determine reasonable answers*. Unpublished doctoral dissertation, University of Missouri, Columbia.
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, 38, 333-339.
- Basturk, R. (2008). Applying the many0facet Rasch model to evaluate powerpoint presentation performance in higher education. *Assessment & Evaluation in Higher Education*, vol.33, No.4, 431-444.
- Burns, M. (1994). Arithmetic: The last holdout. *Phi Delta Kappan*, 75, 471-476.
- Deheane, S. (1997). *The number sense*. Oxford, UK: Oxford University Press.
- Dunphy, E. (2007). The primary mathematics curriculum: Enhancing its potential for developing young children's number sense in the early years at school. *Irish Educational Studies*, 26(1), 5-25.
- Engelhard, G. J. (1992). The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education*, 5, 171-191.
- Gardner, H. (1993). *Frames of Mind: The theory of multiple intelligences*, New York: Basic Books.
- Jurdak, M., & Zein, R. A. (1998). The effect of journal writing on achievement in and attitudes toward mathematics. *School Science & Mathematics*, 98(8), 412-419.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

- Linacre, J.M. (1994). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Linacre, J.M. (1995). Constructing measurement with a many-facet Rasch model. In M. Wilson (ed.), *Objective measurement Theory into practice*, 2, vol2, 129-144. Norwood, NJ: Ablex.
- Linacre, J.M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3, 103-122.
- Lunz M.E., & Stahl, J.A. (1990). Judge consistency and severity across grading periods. *Evaluation & the Health Professions*, 13, 435-444.
- Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 331-345.
- Lunz, M.E., Wright, B.D., Stahl, J.A., & Linacre, J.M. (1989). *Equating practical examinations*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Markovits, Z., & Sowder, J. T. (1994). Developing number sense: An intervention study in grade 7. *Journal for Research in Mathematics Education*, 25, 4–29.
- McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2–8.
- Menon, R. (2004). Elementary school children's number sense. *International Journal for Mathematics Teaching and Learning*. Retrieved January 10, 2008, from <http://www.cimt.plymouth.ac.uk/journal/ramamenon.pdf>
- National Council of Teachers of Mathematics (NCTM). (2000). *The principles and standards for school mathematics*. Reston, VA: Author.
- Reys, R. E., & Yang, D. C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education*, 29, 225–237.
- Shavelson, R.J., Baxter, G.P., & Pine, J. (1992). Performance assessment: Political rhetoric and measurement reality. *Educational Researcher*, 21(4), 22-27.
- Stiggins, R.J. (1994). *Student-centered classroom assessment*. New York: Macmillan.
- Smith, E.V., Kulikowich, J.M. (2004). An application of generalizability theory and many facet Rasch measurement using a complex problem solving skills assessment. *Educational and Psychological Measurement*, vol.64, No.4., 617-639.
- Twing, J., & Williams, K.T. (1992). An investigation of writing assessment using a many-faceted Rasch model. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Yang, D. C. (2005). Number sense strategies used by sixth grade students in Taiwan. *Educational Studies*, 31, 317–334.
- Yang, D. C., & Wu, W. R. (2010). The study of number sense: Realistic activities

integrated into third-grade math class in Taiwan. *The Journal of Research*, 103, 379-392.

Yang, D. C., & Li, M. N. (2008). An investigation of 3rd grade Taiwanese students' performance in number sense. *Educational Studies*, 34,443–455.

附件: 評分標準

分數	評分指標
3	<p><u>完全正確、清楚</u></p> <p>能使用適當的策略與步驟來完成解題</p> <p>能展現出學生有清楚的概念與解題的邏輯</p> <p>學生的解釋很清楚</p> <p>學生能使用圖像或其它方式來表達解題的過程</p> <p>能正確與合理的回答三個子題問題</p>
2	<p><u>部份正確、清楚</u></p> <p>某些問題能清楚的回答，但並不完整</p> <p>學生在某些概念是很清楚的，但並不是完全清楚</p> <p>有些學生的解釋很清楚，但需再多些解釋才會更加完整</p> <p>學生的圖像或是其它的表現方式，對解題並未有明顯的幫助</p> <p>有些答案是正確的，但有些答案是錯誤的</p>
1	<p><u>少數正確、清楚</u></p> <p>有些解題的過程省略或是不正確，僅有少部份的答案是正確的</p> <p>少量的證據顯示學生了解這些概念和步驟</p> <p>學生的解釋和表現很難理解，有些解題的過程需再多做說明</p> <p>學生的圖像或是其它的表現方式，對解題可能無關</p> <p>大多的答案都是不正確的</p>
0	<p><u>不正確，不完整、不清楚</u></p> <p>所有的問題都是錯誤的</p> <p>學生並未了解關鍵的數學概念</p> <p>學生的解題策略與方法令人無法理解</p> <p>學生的圖像或是其它的表現方式，對解題無關</p> <p>所有的答案都是不正確的</p>