

行政院國家科學委員會專題研究計畫 成果報告

國際數學與科學教育成就趨勢調查 2007(4/4)甲 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 97-MOE-S-003-003-
執行期間：97年04月01日至98年09月30日
執行單位：國立臺灣師範大學地球科學系（所）

計畫主持人：張俊彥
共同主持人：張幼賢、楊文金、洪志明、左台益、譚克平
曹博盛、張文華、許瑛珖、楊芳瑩、吳心楷
楊瑞智、呂玉琴、林碧珍、李源順、任宗浩
張美玉、李哲迪
計畫參與人員：此計畫無其他參與人員

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 98 年 12 月 31 日

行政院國家科學委員會補助專題研究計畫成果報告

計畫類別：☒個別型計畫 ☐整合型計畫

計畫編號：第一年 NSC 94-2522-S-003 -003-【(1/4)】

第二年 NSC 95-MOE -S-003 -003 -【(2/4) 甲】

NSC 95-2522-S-003 -003 -【(2/4) 乙】

第三年 NSC 96-MOE -S-003 -002 -【(3/4) 甲】

NSC 96-2522-S-003 -003 -【(3/4) 乙】

第四年 NSC 97-MOE -S-003 -003 -【(4/4) 甲】

NSC 97-2522-S-003 -001 -【(4/4) 乙】

執行期間：民國 94 年 4 月 1 日至 98 年 9 月 30 日

第四年執行期間：民國 97 年 4 月 1 日至 98 年 9 月 30 日

計畫主持人：張俊彥教授

共同主持人：張幼賢教授

張美玉教授

左台益副教授

林碧珍教授

曹博盛副教授

洪志明副教授

李源順教授

許英珣副教授

楊瑞智教授

楊芳瑩教授

呂玉琴教授

張文華副教授

楊文金副教授

譚克平副教授

吳心楷助理教授

任宗浩助理研究員（博士級）

李哲迪助理研究員（博士級）

本成果報告內包括以下之附件：

☒出國人員出國報告書

☒TIMSS 2007 中文可公布試題

另行寄送：

☒TIMSS 2007 報告中文版和英文版各 5 冊

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

☒涉及專利或其他智慧財產權，不提供公開查詢

執行單位：國立臺灣師範大學科學教育中心

中華民國九十八年十二月三十一日

目 錄

◎ 「國際數學與科學教育成就趨勢調查 2007」期末報告·····	1
附 錄 一 國科會記者會新聞資料及簡報資料·····	11
附 錄 二 TIMSS 2007 八年級科學可公布試題（中文版）·····	51
附 錄 三 TIMSS 2007 八年級數學可公布試題（中文版）·····	104
附 錄 四 TIMSS 2007 四年級科學可公布試題（中文版）·····	161
附 錄 五 TIMSS 2007 四年級數學可公布試題（中文版）·····	202
附 錄 六 TIMSS 2007 八年級問卷（中文版）·····	242
附 錄 七 TIMSS 2007 四年級問卷（中文版）·····	315
附 錄 八 出國人員出國報告書·····	368

「國際數學與科學教育成就趨勢調查 2007」期末報告

Trends in International Mathematics and Science Study 2007

計畫編號：第四年 NSC 97-MOE -S-003 -003 -【(4/4) 甲】

第四年 NSC 97-2522-S-003 -001 -【(4/4) 乙】

主 持 人：張俊彥教授

共同主持人：張幼賢教授、張美玉教授、左台益副教授、林碧珍教授、曹博盛副教授、洪志明副教授、李源順教授、許英珣副教授、楊瑞智教授、楊芳瑩教授、呂玉琴教授、張文華副教授、楊文金副教授、譚克平副教授、吳心楷助理教授、任宗浩助理研究員（博士級）、李哲迪助理研究員（博士級）

執行單位：國立臺灣師範大學科學教育中心

執行期限：民國 94 年 4 月 1 日至 98 年 9 月 30 日

第四年執行期限：民國 97 年 4 月 1 日至 98 年 9 月 30 日

E-mail：changcy@ntnu.edu.tw

中文摘要

繼TIMSS 1995、TIMSS 1999(TIMSS -R)及TIMSS 2003之後，IEA計劃再辦理國際數學與科學教育成就趨勢調查(Trends in International Mathematics and Science Study 2007, 簡稱TIMSS 2007)。我國目前是國際教育學習成就調查委員會(IEA)正式會員，國科會和教育部共同支持補助參與TIMSS 2007調查，希望經由積極參與國際性研究計畫，為我國增添一個有利的管道。可以進一步了解各國學生數學及科學學習成就及其與各國文化背景、教育環境等影響因子之相關性，以及我國學生數學及科學學習成就與國際間之比較情形，其結果可作為我國科學教育研究與實施參考之使用。

TIMSS 2007 的主要調查對象為四年級及八年級學生，共計有 59 個國家參與並完成調查工作，其中 50 個國家參加八年級調查，26 個國家參加四年級調查。國際上TIMSS 2007 從 2005 年 2 月開始發展研究調查架構與試題，我國自 2005 年 4 月開始加入 TIMSS 2007 國際調查工作，包括提供

命題架構意見、數學和科學試題命題、參加專家問卷會議、試測(field test)資料收集、實測(main survey)資料收集、參加國際成果指標會議、國際資料分析會議等各項工作。其中我國於 2006 年 4 月 19 日進行試測，實測部分於 2007 年 5 月 21 日至 6 月 8 日舉行，國中、小各抽測 150 校，其中四年級有 27 所較大型學校抽測兩班，共抽測 327 個班級。參與受測的學生計有八年級共 4046 位學生、四年級共 4131 位學生分別完成測驗。每位受測學生需完成問卷與題本的填答，而受測學校的校長及教授受測班級數學和科學的教師則需配合填寫學校問卷與教師問卷。

測驗結果：我國八年級學生科學學習成就平均成績的國際排名為第二名，與第一名的新加坡無顯著差異；數學平均成績的國際排名為第一名，但與第二名的韓國與第三名的新加坡無顯著差異。我國四年級學生科學學習成就平均成績的國際排名為第二名，與排名第一的新加坡有顯著差異；數學平均成績的國際排名為第三名，與排名第一的香港和排名第二的新加坡有

顯著差異。各學科內容領域與認知領域平均成績的國際排名分別如下：

【八年級科學部份】

生物第三名，化學第一名，物理第四名，地球科學第一名；知識領域第二名，應用領域第一名，推理領域第五名。

【八年級數學部分】

數第三名，代數第一名，幾何第一名，資料與機率第四名；知識領域第三名，應用領域第二名，推理領域第一名。

【四年級科學部份】

生命科學第四名，物質科學第三名，地球科學第三名；知識領域第二名，應用領域第八名，推理領域第一名。

【四年級數學部分】

數第三名，幾何形狀與測量第四名，資料呈現第四名；知識領域第三名，應用領域第三名，推理領域第三名。

在我國學生成就趨勢方面，與 TIMSS 2003 比較，八年級數學有顯著的提昇，從排名第四躍升排名第一；八年級科學雖持平排名第二，但卻有退步的現象；四年級數學雖然排名第三，但相較 2003 年的結果，亦有顯著的進步，不過相較於排名第一的香港，仍有進步的空間；四年級科學，與 2003 的結果相比，有顯著的提昇。以整體來說，這個成果顯示我國數學和科學教育一直有不錯的水準，這是大家對科學教育關心與努力的結果。

總的來說，在 TIMSS 2007 的整體表現上，我國四年級與八年級學生於科學方面均名列第二；數學方面則分別名列第三和第一。而與 TIMSS 2003 的表現相較，我國

四年級學生在科學與數學方面及八年級學生在數學方面均有顯著的進步，惟八年級科學成就有下降的趨勢。我國學生在對數理具有高正向態度和對數理具高度學習自信心的人數比例上皆顯著低於國際平均。在性別差異方面，我國兩年段學生不論在數學或科學的成就上，男女生的表現均無統計上的顯著差異。

關鍵詞：國際數學與科學教育成就趨勢調查 2007、TIMSS 2007、數學與科學教育成就

Abstract

The **Trends in International Mathematics and Science Study (TIMSS)** provides reliable and timely data on the mathematics and science achievement of all the participant countries. Carried out every four years at the fourth and eighth grades, TIMSS data has been collected in 1995, 1999, and 2003. TIMSS 2007 is the fourth in a cycle of internationally comparative assessments purposed to investigate the factors which effect on teaching and learning in mathematics and science for students around the world. The results could provide the participants with information to examine the science and mathematics curriculum and education policies by comparing with other countries.

The 4th and 8th graders were two populations investigated in the study of TIMSS 2007. There are 59 countries participated and completed the investigation in this cycle. In Taiwan, the filed test was carried out on April 19, 2006; while as the main survey was processed from May 21 to June 8, 2007. In the main survey, 4131 4th graders from 150 elementary schools and 4046 8th graders from 150 junior high schools were sampled to participate and completed the test in the main survey. Each participant needs to finish the Background Questionnaire and the Achievement Test Booklet. The administrators of sampled schools and both science and math teachers of the sampled classes were surveyed to complete the school questionnaire and teacher questionnaire respectively.

Our grade 8 science performance was ranked 2nd among 50 countries, but the data were not significantly different from the 1st place country, Singapore. The performances of science content areas were followed: biology was ranked in 3rd place, chemistry was ranked in 1st place, physics was ranked in 4th place, and earth science was ranked in 1st place. In cognitive domains, knowing was ranked in 2nd place, applying was ranked in 1st place, and reasoning was ranked in 5th place.

Our grade 8 mathematics performance was ranked 1st among 50 countries, but the data were no significant different from the 2nd place country and 3rd place country, Korea and Singapore. The performances of mathematics content areas were followed: number was ranked in 3rd place, algebra was ranked in 1st place, geometry was ranked in 1st place, and data and chance was ranked in 4th place. The

performance in knowing domain was ranked in 3rd place, applying domain was ranked in 2nd place, and reasoning domain was ranked in 1st place.

Our grade 4 science performance was ranked 2nd internationally, and the data were significantly different from the 1st place country, Singapore. The performances of science content areas were followed: life science was ranked in 4th place, physical science was ranked in 3rd place, and earth science was ranked in 3st place. In cognitive domains, the performance in knowing was ranked in 2nd place, applying was ranked in 8th place, and reasoning was ranked in 1st place.

Our grade 4 mathematics performance was ranked 3rd internationally, but the data were significantly different from the 1st place country and 2nd place country, Hong Kong and Singapore. The performances of mathematics content areas were followed: number was ranked in 3rd place, geometric shape and measures was ranked in 4th place, and data display was ranked in 4th place. The performance in knowing, applying and reasoning cognitive domains were all ranked in 3rd place.

To compare with the TIMSS 2003 data, our grade 8 students had significant improvement in mathematics, but less well in science. Both our grade 4 mathematics and science had significant improvement.

In conclusion, overall performance in our grade 4 and grade 8 science were ranked in 2nd place internationally, and mathematics was ranked in 3rd place for grade 4 and in 1st place for grade 8. In trend analysis, except for grade 8 science, our grade 4 and grade 8 mathematics and grade 4 science were significant better than TIMSS 2003. The performance of our grade 4 and grade 8 in Index of Students' Positive Affect and Index of Students' Self-Confidence were below the international average in both subjects. In aspect of gender difference, there was no significant different in mathematics and science in both grade.

Keywords: Trends in International Mathematics and Science Study 2007, TIMSS 2007, Mathematics and Science Achievement

計畫緣由與目的

我國目前為國際教育學習成就調查委員會(The International Association for the Evaluation of Education Achievement, IEA)正式會員，為我國積極參與國際性研究計畫，增添一個有利的管道。國際數學與科學教育成就調查研究的主要目的在於了解各國學生數學及科學學習成就及其與各國文化背景、教育環境等影響因子之相關性，並進一步作國際間之比較分析。

第一次國際數學與科學教育成就調查於1970年舉行，共有19個國家參與。經十年後，1980年進行第二次國際數學與科學教育成就調查(SISS)，有24個國家參與。我國曾於1987年5月經IEA總部同意，引用第二次國際數學與科學教育成就調查工具，在我國進行測驗(但非正式參加)，由國立台灣師範大學科學教育中心負責執行，以了解我國國小、國中及高中學生數學及科學成就在國際上所佔的地位。

IEA自1990年開始推動進行「第三次國際數學與科學教育成就研究(Third International Mathematics and Science Study, TIMSS)」，此計畫有四十餘國參加。第三次國際數學與科學教育成就研究後續調查(稱為TIMSS REPEAT, TIMSS-R)於1999年舉辦，調查對象為八年級學生(13歲群)，共有38個國家參加。為了解我國學生數學與科學教育學習成就，並與世界主要國家相互觀摩溝通。TIMSS-R為我國首次正式參與大規模國際性學生學習成就調查。TIMSS-R的成績分析統計結果已經得知，我國學童表現優異，在科學方面的總成績位居所有參加國之第一名，數學方面居第三

名，但與第一、二名之成績無顯著差異(Mullis, 2000)。

繼TIMSS-R之後，IEA再辦理國際數學與科學教育成就趨勢調查2003(Trends in International Mathematics and Science Study 2003, 簡稱TIMSS 2003)，共有51個國家參加。TIMSS 2003的調查對象包括四年級及八年級學生，分析統計結果得知八年級科學在所有參加國中排名第二、八年級數學排名第四、四年級科學排名第二、四年級數學排名第四。

國際數學與科學教育成就趨勢調查2007(TIMSS 2007)的調查對象同樣也包括我國四年級及八年級學生，共有59個國家參與並完成調查。

基於我國難得有機會正式參加大規模的國際性學生學習成就比較測驗，以得知我國學童的學習成就情況及在國際間比較的結果，並與TIMSS-R、TIMSS 2003的結果作比較。藉此對於我國科學教育施行的成果，有一國際性指標，因此我國在國科會和教育部共同支持與補助下，由國立台灣師範大學科教中心、科教所、理學院各系、以及國立新竹教育大學、國立台北教育大學及台北市立教育大學的教授共同組成團隊，積極參與此計畫過程的各項研究工作，和國際協調工作，期能圓滿達成此種大型的調查研究。

本計畫目的在於調查我國四年級及八年級之學生數學與科學學科學習成就並作國際性比較，其結果可作為我國科學教育研究與實施之參考。探討之主題如下：

1. 了解我國學生數學及科學學習成就，

提供規劃我國中小學的數學及科學課程之參考。

2. 了解我國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，並進一步作國際比較分析。
3. 與TIMSS-R及TIMSS2003之結果作比較，了解我國學生在數學及科學學習成就之趨勢。
4. 了解國際上評量學生學習成就的趨勢與新的評量方法，提供我國教師參考。
5. 提昇國內研究人員資料分析研究能力。

研究方法

本研究屬調查（Survey）研究。在國際學者合作下，各項問卷及試題均有嚴謹的設計，以控制其品質，TIMSS 2007對於調查研究之方法有統一之規定。本計畫將依規定之方法，針對我國四年級和八年級學生之數學與科學學科學習成就進行調查研究工作。依規定參與之國家需指定一個全國性的主管單位(National Center)負責推動全部研究工作，並指派一位國家研究聯絡人(National Research Coordinator)負責完成必需的工作。我國國家研究聯絡人由本計畫主持人國立台灣師範大學理學院張俊彥教授擔任，責任協調執行計畫，並參加一年至少兩次的會議，商討有關國際試題開發、測驗架構分析、問卷修訂、施測之取樣、蒐集與處理數據、活動報告、成果分析及最後之行政工作。本計畫研究人力結合國立台灣師範大學理

學院及國立新竹教育大學、國立台北教育大學及台北市立教育大學的學者共同合作執行本計畫。

研究進行時程

本計畫於94年4月1日開始執行，至98年9月30日完成。TIMSS 2007計畫主要工作時程如下：

2005年2月~4月	發展架構
2005年5月~12月	發展工具
2005年12月	試題翻譯
2005年12月~2006年3月	準備試測
2006年4月	進行試測
2006年6月~2007年3月	準備施測
2006年10月~2007年2月	南半球施測
2007年2月~6月	北半球施測
2007年7月~12月	各國評分與處理數據
2007年9月~2008年12月	
	建立與分析國際資料
2008年12月	出版國際報告
2009年2月	公布國際資料庫與使用手冊

工作要項及完成程度

第一年工作要項

【94年4月1日至95年3月31日】

本年度為第一年研究工作，主要由研究組規劃執行，配合國際研究進行開發試題、分析我國中小學課程架構、修訂問卷

工具、受測學校班級抽樣及試測試題翻譯等。詳細成果如本報告所附。完成之工作摘述如下：

1. 配合國際時程，協助完成TIMSS Assessment Frameworks and Specifications 2007。
2. 彙整各科開發試題送至美國ISC。
3. 彙整各問卷工具修訂意見送至美國ISC。
4. 分析我國中小學課程架構。
5. 派員參加三次國家聯絡人會議(分別在埃及開羅、荷蘭的阿姆斯特丹、和斯洛維尼亞的朱布亞納舉行)。
6. 派員參加在德國漢堡舉行的資料分析工作坊。
7. 完成受測學校及班級抽樣，目前正在回收學生資料表階段。
8. 試測試題分科翻譯完成送至IEA。

第二年工作要項

【95年4月1日至96年3月31日】

本年度為第二年，主要為試測施行，將製作完成的題本及問卷寄送給抽樣出來的受測學校施測，並派共同主持人到數間受測學校視察。收回題本及問卷後，進行閱卷工作。將閱卷結果及問卷作答情形整理建檔，並送交IEA等。後半年度並開始籌備實測施行工作。詳細成果如本報告所附。完成之工作摘述如下：

1. 寄送試測題本、問卷至受測學校施測。

2. 請共同主持人至抽出之受測學校視察。
3. 回收題本、問卷之後進行閱卷工作。
4. 將試測施行結果及問題送交IEA。
5. 派一位老師參加第五次NRC 會議。
6. 翻譯2007年實測要測驗的題目、問卷。
7. 將製成的題本及問卷送交IEA 審查。
8. 調查各抽樣受測學校的班級資料。

第三年工作要項

【96年4月1日至97年3月31日】

本年度為第三年，主要進行實測，試題印製、施測與資料輸入，配合國際時程按時完成實測工作。本年度完成實測之工作摘述如下：

1. 科教中心依國際測驗總部規定，使用WinW3S軟體進行班級抽樣，依受測學校規模，分別從每校抽出1~2班作為施測對象。
2. 調查受測班級學生名單，製作受測學校之教師追蹤表與學生追蹤表。
3. 實測之工具，包括四年級測驗題本共18本、四年級學校問卷、四年級教師問卷、四年級學生問卷、八年級測驗題本共18本、八年級學校問卷、八年級數學教師問卷、八年級科學教師問卷、八年級學生問卷。各翻譯小組配合國際時程，於95年

- 11月已完成實測翻譯工作，由科教中心彙整修改，並寄至荷蘭IEA國際測驗總部確認所有翻譯之工具，並依照國際測驗總部提供之意見修改，再寄至美國Boston College檢核與確認，俟全部修改完畢在96年4月付印。
4. 由任宗浩助理研究員代表參加在杜拜舉行的第六次國家聯絡人會議，接受自由反應試題評分訓練，並帶回閱卷評分相關資料。
 5. 試卷設計中，四年級和八年級測驗題本測驗題本各有18本，測驗題本包含數學與科學試題。測驗採測驗題本輪測設計，每位學生只考其中一本測驗題本。四年級實測的測驗時間為72分鐘，分兩階段進行，每一階段測驗時間為36分鐘，中間休息10-20分鐘；八年級實測的測驗時間為90分鐘，分兩階段進行，每一階段測驗時間為45分鐘，中間休息10-20分鐘。學生於測驗後約花30分鐘填寫學生問卷，回答有關於對數學與科學的態度、教室活動、家庭背景及學校外之活動。
 6. 由學校聯絡人協助發放與收集教師及學校問卷，以收集有關教學與學校的背景資料；受測班級的數理科教師回答有關教學重點、教學經驗、使用之教科書、教育程度與專業訓練及他們對數學或科學的看法；學校校長回答有關學校教職員及資源、數理課程及教師支持情形。
 7. 96年3月底至5月初分北、中、南三區辦理實測受測學校聯絡人及監試人員講習會，邀請受測學校之聯絡人及監試人員參加，使其了解本計畫之目的與施測時進行之程序、注意事項及測驗工具保密等相關工作。
 8. 指定八年級於民國96年5月21日至6月1日舉行實測，四年級於民國96年5月28日至6月8日舉行實測，科教中心派出15位觀察員抽訪國中和國小各15所受測學校，訪視各受測學校測驗之程序是否有依規定進行。國際測驗總部另外委託花蓮教育大學林煥祥校長組國際觀察員小組，代表IEA，抽訪國中和國小各15所受測學校。因此共有60所學校有觀察員共同監試測驗。
 9. 四年級有150所學校，其中有27所較大型學校抽測2班，其餘123所學校抽測一班，共5320位學生參與測驗；八年級有150所學校每校一班，共5196位學生參與測驗(不含被排除、缺席、只作答試題或問卷之一的學生)。每位學生必須填寫一份問卷和試題，受測班級之數學與科學課程教師需分別填寫一份問卷，校長或由校長指定專人填寫一份學校問卷。測驗後由學校聯絡人負責將所有測驗材料送回科教中心。
 10. 辦理自由反應試題評分研習會，邀請數學、物理、化學、生物、地球科學等科的閱卷教授、教師、研究生，分科研討診斷式評分系統。於6

月22日開始閱卷。

11. 對於自由反應試題(非選擇題)的評分採用診斷式評分系統評分碼，每一試題的給分標準以兩位數評分碼給分，第一位數字表示回答正確性的層次，第二位數字代表診斷碼，用來定出取向的特別型式、策略、共同錯誤與迷思概念。所有得分需根據國際研究中心評分指引上的標準給分。於7月22日完成所有閱卷工作。
12. 7月初開始進行各種問卷、試題答案之輸入訓練與輸入工作。使用由DPC提供之WinDEM軟體輸入各種資料。
13. 9月中旬完成各種問卷、試題答案之輸入，經複檢資料後於九月底將資料送至DPC進行分析。
14. 9月完成資料輸入相關表格，並送至Boston College和DPC。
15. 9月完成電腦閱卷IEA Trend Scoring Reliability。
16. 9月完成電腦閱卷IEA Cross-Country Scoring Reliability。
17. 10月完成實測NRC報告，並送至Boston College和DPC。
18. 10月完成 Curriculum 問卷填寫。
19. 12月由任宗浩助理研究員代表參加在奧地利薩爾茲堡舉行的第七次國家聯絡人會議。
20. 開始進行題本掃描工作。

第四年工作要項

【97年4月1日至98年9月30日】

本年度為第四年，主要進行資料建檔、研究報告撰寫等，完成之工作摘述如下：

1. 完成 NRC 課程問卷。
2. 協助 ISC 和 DPC 校正資料庫資料與國際報告撰寫資料。
3. 參加 2008 NARST 國際年會及參訪 National Science Foundation。
4. 協助辦理於我國舉行之 IRC2008 會議。
5. 協助國科會和教育部於 97 年 12 月 10 日召開記者會公布成果。
6. 參加 IERI 舉辦之 Assessment Designs, Item Response Theory and Proficiency Estimates in Large Scale Assessment.
7. 參加 ACER 舉辦之 Data Analysis of Large-Scale Survey Data Workshop。
8. 完成學生非選擇題作答及問卷填答資料掃描建檔工作。
9. 撰寫中文和英文國家研究報告。

結果與討論

在 50 個參與八年級調查的國家/地區中，我國學生科學學習成就平均成績的國際排名為第二名，與第一名的新加坡無顯著差異；八年級學生數學平均成績的國際排名為第一名，但與第二名的韓國及第三

名的新加坡無顯著差異。科學部份各內容領域平均成績的國際排名分別為生物第三名，化學第一名，物理第四名，地球科學第一名；數學部份則分別為數第三名，代數第一名，幾何第一名，資料與機率第四名。

四年級部分共有 26 國參與，我國的測驗結果為科學學習成就平均成績的國際排名為第二名，與排名第一的新加坡有顯著差異；數學平均成績的國際排名為第三名，與排名第一的香港和排名第二的新加坡有顯著差異。科學部份各內容領域平均成績的國際排名分別為生命科學第四名，物質科學第三名，地球科學第三名；數學部份則分別為數第三名，幾何形狀與測量第四名，資料呈現第四名。這個成果顯示我國數學和科學教育一直有不錯的水準，這是大家對數學與科學教育關心與努力的結果。

在認知領域成就方面，八年級科學部分的國際排名為知識領域排第二名，應用領域排第一名，推理領域則排名第五；八年級數學部分為知識領域排第三名，應用排領域排第二名，推理領域則排名第一。在四年級科學部分，知識領域排名第二，應用排領域排名第八；推理領域則排名第一；四年級數學在知識、應用和推理各領域則均排名第三。

在我國學生成就趨勢方面，與 TIMSS 2003 的結果相比較，八年級學生在數學方面從排名第四躍升至排名第一，且在成就上有顯著的提昇；八年級科學在國際排名上雖持平排名第二，但卻有退步的現象；四年級數學雖然排名第三，但較 2003 年的

結果，有顯著的進步，不過相較於排名第一的香港，仍有進步的空間；四年級科學亦有顯著的提昇。

在對數學具正向態度指標（Index of Students' Positive Affect Toward Mathematics，簡稱 PATM）方面，我國八年級部分有 37%達高正向態度指標，四年級部分有 50%學生達高正向態度指標，均低於國際平均百分比。在對科學具正向態度指標（Index of Students' Positive Affect Toward Science，簡稱 PATS）方面，八年級部分有 40%達高正向態度指標，低於國際平均百分比，四年級部分則有 75%學生達高正向態度指標，略低於國際平均百分比。

我國八年級學生在數學自信心指標（Index of Students' Self-Confidence in Learning Mathematics，簡稱 SCM）方面，僅有 27%達高自信心指標，四年級部分有 36%學生達高自信心指標，均低於國際平均百分比。在科學自信心指標（Index of Students' Self-Confidence in Learning Science，簡稱 SMS）方面，八年級部分僅 23%達高自信心指標，低於國際平均百分比，四年級部分則有 58%學生達高正向態度指標，略低於國際平均百分比。

在性別差異對數學與科學成就表現的比較上，我國八年級與四年級學生皆無顯著的差異。

結論

總的來說，在 TIMSS 2007 的整體表現上，我國四年級與八年級學生於科學方面

均名列第二；數學方面則分別名列第三和第一。而與 TIMSS 2003 的表現相較，我國四年級學生在科學與數學方面及八年級學生在數學方面均有顯著的進步，為八年級科學成就有下降的趨勢。我國學生在對數理具有高正向態度和對數理具高度學習自信心的人數比例上皆顯低於國際平均。在性別差異方面，我國兩年段學生不論在數學或是科學的成就上，男、女生的表現均無統計上的顯著差異。

閱卷教師和研究生們、資料輸入與檢查的同學們以及科學教育中心全體同仁協助 TIMSS 2007 實測調查工作，使實測工作順利完成。謝謝國科會和教育部經費上的支持。

成果與應用

本年度工作為國際性調查實測階段，經由研究組團隊共同努力參與，完成實測與資料輸入工作，並將我國對測驗工具修訂意見呈至國際測驗總部。因我國在四年級學生與八年級學生兩個年段上已分別有 TIMSS 2003 的測驗資料，可供比較分析，而測驗資料可結合 TIMSS 1999 及 TIMSS 2003 測驗資料分析，進行趨勢研究。

本計畫之相關四年級和八年級可公布中文試題和中文問卷皆已公佈於國立台灣師範大學科學教育中心網站上，供有興趣者參考下載。不可公布試題的使用辦法由國科會另訂之。中、英版文國家報告將另行寄送。

致謝

感謝所有受測學校與相關縣市教育局的配合與協助；同時也感謝國立台灣師範大學理學院、國立新竹教育大學、國立台北教育大學及台北市立教育大學的教授們、

附錄一

1. 教育部及國科會聯合記者會新聞資料
2. 教育部及國科會聯合記者會簡報資料

教育部及國科會聯合記者會新聞資料

國民教育的世界盃 – 台灣學生的數理表現

—國際數學與科學學習成就趨勢調查(TIMSS 2007)結果—

主持人：教育部

國科會 張副主任委員文昌

報告人：國立台灣師範大學

張 俊 彥 教授

九十七年十二月十日

教育部及國科會聯合記者會新聞稿

國民教育的世界盃 – 台灣學生的數理表現

我國參加由國際教育成就調查委員會 (The International Association for the Evaluation of Educational Achievement, 簡稱 IEA) 主辦，針對四年級與八年級學生進行的「國際數學與科學教育成就趨勢調查」(Trends in Mathematics and Science Study 2007, 簡稱 TIMSS 2007)」結果出爐。我國四年級學生的數學和科學平均成績排名，在 37 個參加國中，分別排名第三和第二；而八年級學生的數學和科學平均成績在 50 個參加國中，分列第一和第二，這樣傑出的表現，再次展現我國各界對科學教育關心和努力的成果。

IEA 自 1995 年開始辦理 TIMSS 計畫，該計畫每隔四年針對參加國家或地區抽樣四年級與八年級學生，進行數學與科學的學習成就調查。我國曾參加 1999 年及 2003 年的 TIMSS 調查工作，TIMSS2007 的施測於該年 5 月 21 日至 6 月 8 日進行，共計抽測 4,046 位八年級學生和 4,131 位四年級學生；所有調查方式及作業流程均照國際標準與國際同步進行。

成就表現傑出，維持水準

本屆調查結果顯示，我國、新加坡、日本、韓國以及香港等亞洲國家，在兩個年級的數學和科學成就均名列前茅。我國四年級學生數學平均成績的國際排名為第三名，與名列第一和第二的新加坡和香港有顯著差異；科學學習成就平均成績的國際排名為第二名，與第一名的新加坡有顯著差異。我國八年級學生數學平均成績的國際排名為第一名，與第二、三名的韓國和新加坡無顯著差異，並列為最優等級國家；八年級的科學平均成績則排名第二，與第一名的新加坡沒有顯著差異，並列為最優等級國家。

此外，IEA 依據得分將學生表現分為優、高級、中級、初級等四個基準(Benchmark)：

在數學科方面，我國四年級學生列為優的有 24%，比例名列第三，新加坡和香港則分別有 41%和 40%；另一方面我國僅有 1%未達初級標準，新加坡有 2%未達初級標準，香港則全部四年級生皆達初級標準。八年級學生列為優的有 45%，排名第一，優於韓國和新加坡的 40%，但在低分群方面，我國有 5%未達到初級標準，而韓國和新加坡僅 2%及 3%，顯示我國八年級生低分群的比例偏高。

在科學成就方面，我國四年級學生列為優的有 19%，比例名列第二，新加坡有 36%；另一方面我國有 3%未達初級標準，新加坡有 4%未達初級標準。我國八年級學生列為優的有 25%，排名第二，名列第一的新加坡則有 32%；在低分群方面，我國有 5%未達到初級標準，與數學科相同，顯示我國在「帶好每個學生」的教育理念與目標上仍有努力的空間。

整體成就趨勢分析結果

我國四年級群學生曾參加過 TIMSS 2003，而八年級群學生曾參加在 TIMSS 1999 和 TIMSS 2003 調查，因此可以進行成就趨勢分析。比較結果顯示我國四年級學生在數學和科學成就的表現比起 TIMSS 2003 有顯著進步；八年級學生的數學成就比 TIMSS 1999 和 TIMSS 2003 亦均有顯著進步，然而八年級的科學成就雖名列第二，但相較於 TIMSS 2003 的成績卻有顯著下降的趨勢。

高正向數理態度學生人數比例顯著低於國際平均

除學習成就調查外，TIMSS 亦設計了問卷，收集學生的學習態度與自信心等資訊。在【數學正向態度】上，我國四年級生高數學正向態度人數百分比在 36 個列入排序的國家中敬陪末座，八年級的學生則是在 49 個列入排序國家中排名第 39，兩項均顯著低於國際平均。在【科學正向態度】的部分，四年級和八年級的高科學正向態度人數的

百分比亦均顯著低於國際平均，分列第 23 及 28 名。有趣的現象是，許多在【正向態度】排名較前面的國家，其成就表現卻都不盡理想。

高數理學習自信心人數比例顯著低於國際平均

我國學生關於數學和科學的【學習自信心】也和【正向態度】類似，均顯著低於國際平均。調查結果顯示，包含我國在內的幾個亞洲國家(如新加坡、香港、韓國與日本)在【學習自信心】的表現也都不佳。

我國男女學生在數學及科學表現無顯著差異

性別差異向來是國際調查中重要的分析指標之一。本屆調查結果顯示，我國的女生和男生在科學與數學成就上，沒有顯著差異；這結果和 TIMSS 2003 相同。

結論

綜上所述，根據 TIMSS 2007 調查結果歸納出幾點重要發現：

- 一、 我國四年級和八年級學生的表現在數學方面名列第三和第一，科學領域兩年段均名列第二；
- 二、 相較於 TIMSS 2003 的表現，我國學生在四年級和八年級的數學以及四年級的科學成就表現，均顯著進步，惟八年級科學成就有下降趨勢；
- 三、 我國學生對於數學或科學具有高正向態度的人數比例，均顯著低於國際平均；
- 四、 我國學生對於數學或科學具有高度學習自信心的人數比例，均顯著低於國際平均；
- 五、 在男女性別差異方面，我國兩個年段學生不論在數學或是科學上的成就，男女生的表現均無統計上的顯著差異。

整體而言，我國學生在 TIMSS 2007 的表現可以說是維持過去的傑出水準，但在數理學習的正向態度與自信心部分仍有待提升。

以下表格內容摘錄自 TIMSS 2007 國際報告 (<http://timss.bc.edu/isc/publications>)

表一：TIMSS 2007 四年級與八年級數學與科學成就表現前五名排序

排名	四年級				八年級			
	數學		科學		數學		科學	
	國家	平均分數 (標準誤)	國家	平均分數 (標準誤)	國家	平均分數 (標準誤)	國家	平均分數 (標準誤)
1	香港	607(3.6)	新加坡	587(4.1)	台灣	598(4.5)	新加坡	567(4.4)
2	新加坡	599(3.7)	台灣	557(2.0)	韓國	597(2.7)	台灣	561(3.7)
3	台灣	576(1.7)	香港	554(3.5)	新加坡	593(3.8)	日本	554(1.9)
4	日本	568(2.1)	日本	548(2.1)	香港	572(5.8)	韓國	553(2.0)
5	哈薩克	549(7.1)	俄羅斯	546(4.8)	日本	570(2.4)	英格蘭	542(4.5)

表二：TIMSS 2007 各國在數學及科學次領域成就表現前五名排序

排序 次領域		1	2	3	4	5
四年級數學	數	新加坡	香港	台灣	日本	哈薩克
	幾何形狀與測量	香港	新加坡	日本	台灣	英格蘭
	資料呈現	香港	新加坡	日本	台灣	英格蘭
八年級數學	數	新加坡	韓國	台灣	香港	日本
	代數	台灣	韓國	新加坡	香港	日本
	幾何	台灣	韓國	新加坡	日本	香港
	資料與機率	韓國	新加坡	日本	台灣	香港
四年級科學	生命科學	新加坡	義大利	匈牙利	台灣	美國
	物質科學	新加坡	日本	台灣	香港	俄羅斯
	地球科學	香港	新加坡	台灣	英格蘭	拉托維亞
八年級科學	生物	新加坡	日本	台灣	韓國	英格蘭
	化學	台灣	新加坡	日本	斯洛維尼亞	匈牙利
	物理	新加坡	韓國	日本	台灣	英格蘭
	地球科學	台灣	斯洛維尼亞	新加坡	韓國	捷克

表三：TIMSS 2007 各國在數學及科學次領域成就表現前五名排序

次領域 \ 排序		1	2	3	4	5
四年級數學	知識	新加坡	香港	台灣	日本	哈薩克
	應用	香港	新加坡	台灣	日本	哈薩克
	推理	新加坡	香港	台灣	日本	俄羅斯
八年級數學	知識	韓國	新加坡	台灣	日本	香港
	應用	韓國	台灣	新加坡	香港	日本
	推理	台灣	韓國	新加坡	日本	香港
四年級科學	知識	新加坡	台灣	香港	俄羅斯	日本
	應用†	新加坡	香港	英格蘭	俄羅斯	美國
	推理	台灣	新加坡	日本	香港	拉托維亞
八年級科學	知識	新加坡	台灣	日本	匈牙利	韓國
	應用	台灣	新加坡	韓國	日本	俄羅斯
	推理	新加坡	日本	韓國	英格蘭	台灣

†我國在四年級科學【應用】認知領域排名第九

表四：我國學生參加 TIMSS 2003 與 TIMSS 2007 成就表現趨勢分析

調查研究	四年級		八年級	
	數學平均成績 (標準誤)	科學平均成績 (標準誤)	數學平均成績 (標準誤)	科學平均成績 (標準誤)
TIMSS 2003	564(1.8)	551(1.7)	585(4.6)	571(3.5)
TIMSS 2007	576(1.7)	557(2.0)	598(4.5)	561(3.7)
進步成績	+12(2.5)*	+5(2.6)*	+13(6.4)*	-10(5.0)*

*代表該進步(+)或退步(-)成績具統計上的顯著意義

表五： TIMSS 2007 台灣、香港、新加坡、韓國與日本高數理正向態度
人數百分比與國際平均

參加國家	四年級		八年級	
	高數學正向態度 人數比例(標準 誤)	高科學正向態度 人數比例(標準 誤)	高數學正向態度 人數比例(標準 誤)	高科學正向態度 人數比例(標準 誤)
國際平均	72 (0.2)	77 (0.2)	54 (0.2)	65 (0.2)
台灣	50 (1.2)	75 (1.4)	37 (1.2)	40 (1.3)
香港	67 (1.3)	79 (1.0)	47 (1.2)	60 (1.4)
新加坡	71 (0.8)	75 (0.7)	60 (1.0)	68 (0.9)
韓國	未參加	未參加	33 (0.9)	38 (1.1)
日本	62 (1.4)	81 (0.9)	30 (1.1)	47 (1.1)

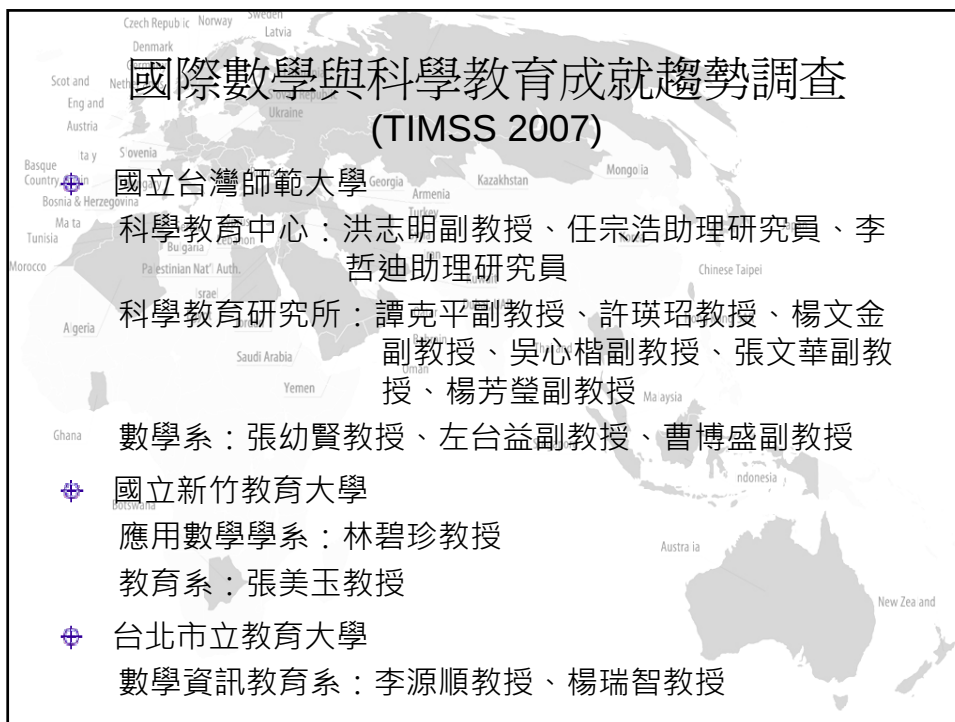
表六： TIMSS 2007 台灣、香港、新加坡、韓國與日本高數理學習自信
心人數百分比與國際平均

參加國家	四年級		八年級	
	高數學學習自信 心人數比例(標準 誤)	高科學學習自信 心人數比例(標準 誤)	高數學學習自信 心人數比例(標準 誤)	高科學學習自信 心人數比例(標準 誤)
國際平均	57 (0.2)	61 (0.2)	43 (0.2)	48 (0.2)
台灣	36 (1.0)	58 (1.2)	27 (1.1)	23 (1.0)
香港	46 (1.0)	52 (1.3)	30 (1.1)	33 (1.3)
新加坡	46 (1.2)	41 (0.9)	41 (1.0)	40 (1.0)
韓國	未參加	未參加	29 (0.8)	24 (1.0)
日本	45 (1.1)	53 (1.2)	17 (0.6)	20 (0.7)

表七：我國男女學生在數學及科學上的成就表現及差異

項目	四年級數學		四年級科學		八年級數學		八年級科學	
	2007 平均 成績(標準 誤)	和 2003 比較 進步成績(標 準誤)	2007 平 均成績 (標準 誤)	和 2003 比較進 步成績 (標準 誤)	2007 平 均成績 (標準 誤)	和 2003 比較進 步成績 (標準 誤)	2007 平 均成績 (標準 誤)	和 2003 比較進 步成績 (標準誤)
男生	577(2.0)	+13(2.9)*	558(2.4)	3(3.3)	598(5.3)	16(7.4)*	563(4.4)	-8(5.8)
女生	575(2.0)	+11(2.7)*	556(2.3)	8(3.1)	599(4.6)	10(6.8)	559(3.7)	-12(5.3)*
男女差 異	2(2.1)	--	2(2.5)	--	-1(4.2)	--	5(3.5)	--
*代表該進步(+)或退步(-)成績具統計上的顯著意義								





國際數學與科學教育成就趨勢調查 (TIMSS 2007)

國立台灣師範大學

科學教育中心：洪志明副教授、任宗浩助理研究員、李哲迪助理研究員

科學教育研究所：譚克平副教授、許瑛珢教授、楊文金副教授、吳心楷副教授、張文華副教授、楊芳瑩副教授

數學系：張幼賢教授、左台益副教授、曹博盛副教授

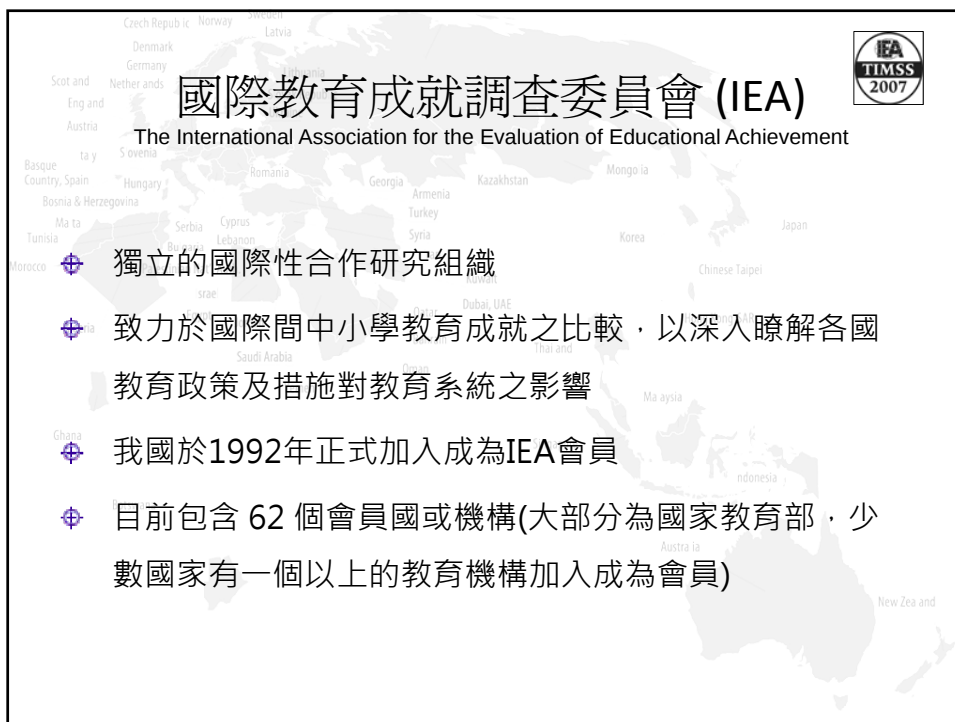
✚ **國立新竹教育大學**

應用數學學系：林碧珍教授

教育系：張美玉教授

✚ **台北市立教育大學**

數學資訊教育系：李源順教授、楊瑞智教授





國際教育成就調查委員會 (IEA)

The International Association for the Evaluation of Educational Achievement

✚ **獨立的國際性合作研究組織**

✚ 致力於國際間中小學教育成就之比較，以深入瞭解各國教育政策及措施對教育系統之影響

✚ 我國於1992年正式加入成為IEA會員

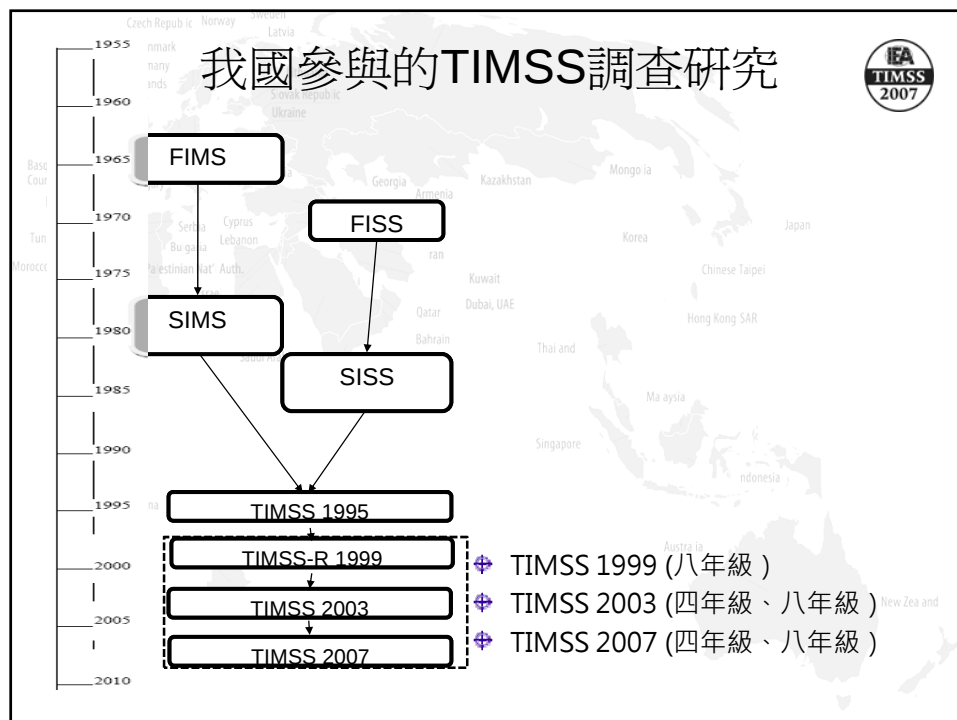
✚ 目前包含 62 個會員國或機構(大部分為國家教育部，少數國家有一個以上的教育機構加入成為會員)

我國參與IEA國際調查研究計畫





- ✚ 「國際數學與科學教育成就趨勢調查」(Trends in International Mathematics and Science Study · 簡稱 TIMSS)
- ✚ 「促進國際閱讀素養研究」(Progress in International Reading Literacy Study · 簡稱 PIRLS)
- ✚ 「數學師資培育跨國研究」(Teacher Education and Development Study in Mathematics · 簡稱 TED-M)
- ✚ 「國際公民教育與素養調查研究」(International Civic and Citizenship Education Study · 簡稱 ICCES)



參與TIMSS2007成就調查國家及地區



所有參與調查的國家：59國家/地區（67參與單位）

八年級：50國家 / 地區（57參與單位）	四年級：37國家 / 地區（44參與單位）
Algeria, Armenia, Australia, Bahrain, Basque Country (Spain), Bosnia and Herzegovina, Botswana, British Columbia (Canada), Bulgaria, Chinese Taipei , Colombia, Cyprus, Czech Republic, Dubai (UAE), Egypt, El Salvador, England, Georgia, Ghana, Hong Kong SAR, Hungary, Indonesia, <i>Iran, Islamic Rep. of</i> , Israel, Italy, Japan, Jordan, <i>Korea, Rep. of</i> , Kuwait, Lebanon, Lithuania, Malaysia, Malta, Massachusetts (US), Minnesota (US), Mongolia*, Morocco, Norway, Oman, Ontario (Canada), Palestinian Nat'l Auth., Qatar, Quebec (Canada), Romania, Russian Federation, Saudi Arabia, Scotland, Serbia, Singapore, Slovenia, Sweden, Syrian Arab Republic, Thailand, Tunisia, Turkey, Ukraine, United States	Alberta (Canada), Algeria, Armenia, Australia, Austria, British Columbia (Canada), Chinese Taipei , Colombia, Czech Republic, Denmark, Dubai (UAE), El Salvador, England, Georgia, Germany, Hong Kong SAR, Hungary, Iran, Islamic Rep. of, Italy, Japan, Kazakhstan, Kuwait, Latvia, Lithuania, Massachusetts (US), Minnesota (US), Mongolia*, Morocco, Netherlands, New Zealand, Norway, Ontario (Canada), Qatar, Quebec (Canada), Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, Sweden, Tunisia, Ukraine, United States, Yemen
*國際報告中並未包含Mongolia數據	

7

調查對象、期程與工具

由IEA監督抽樣過程，共抽測：

八年級生：150校，每校1班，共4046人

四年級生：150校，每校1班，共4131人

調查期程：

八年級：2007年5月21日~6月1日

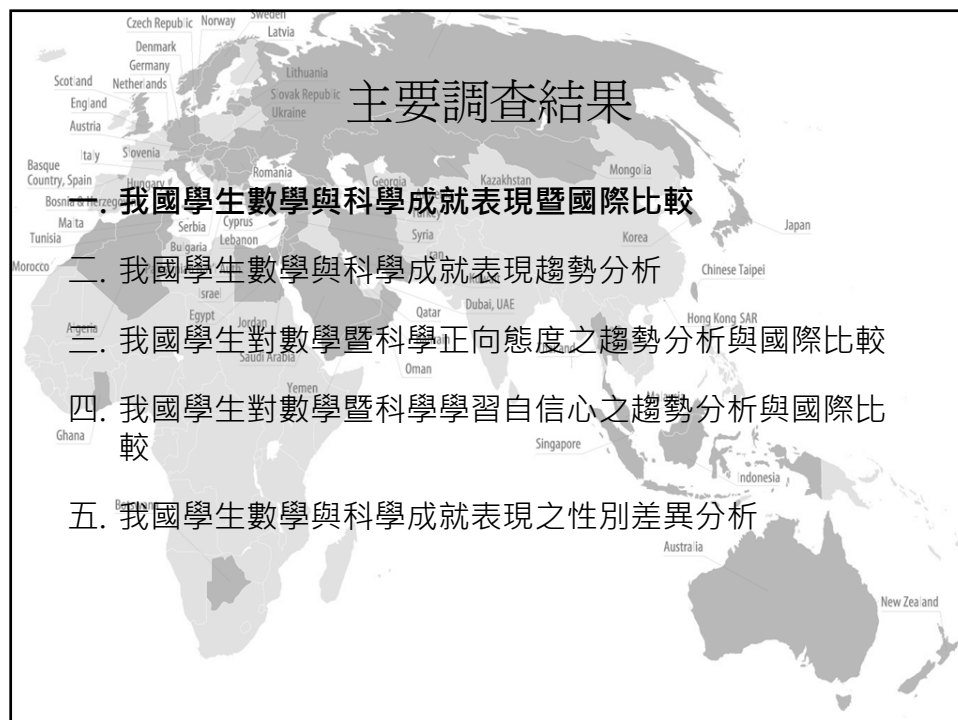
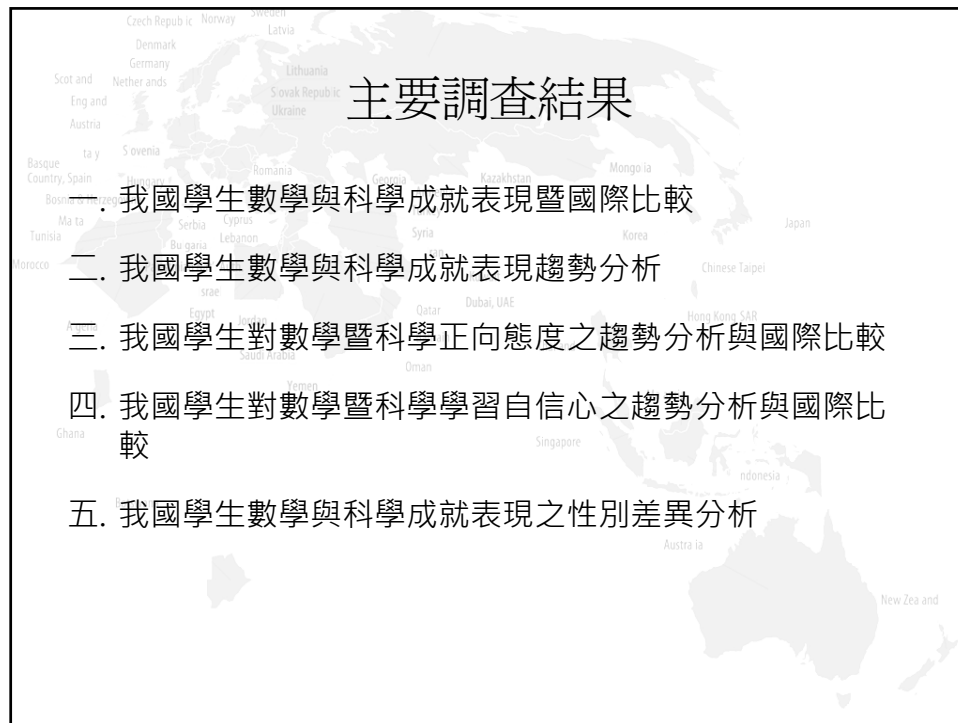
四年級：2007年5月28日~6月8日

調查工具：

學生：數學暨科學成就測驗題本、背景問卷

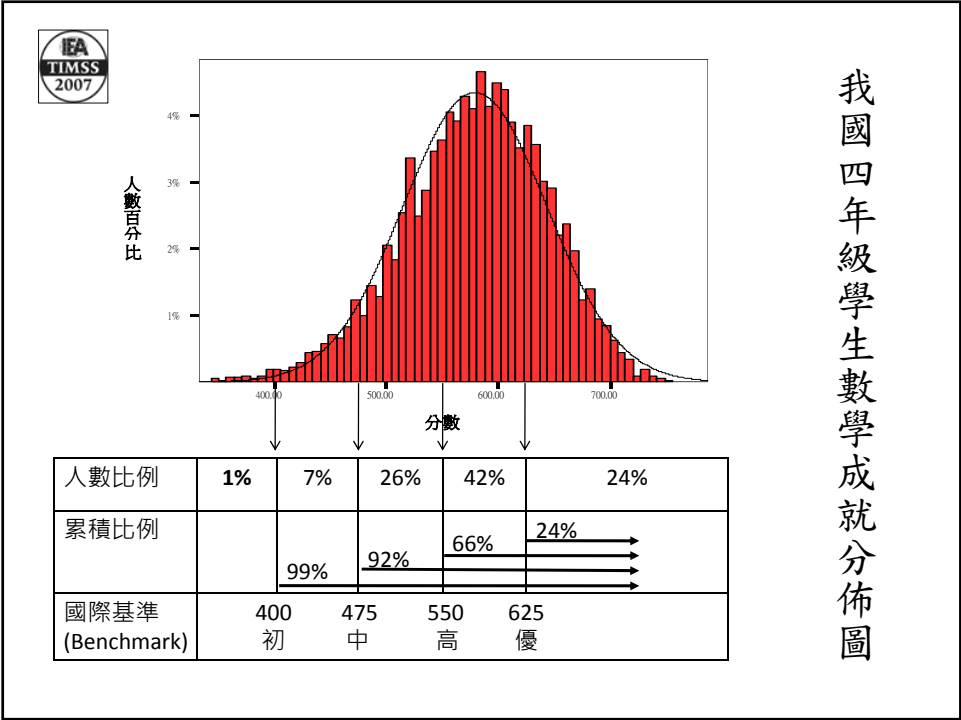
數理任課教師：教師問卷

校長(或代理人)：學校問卷

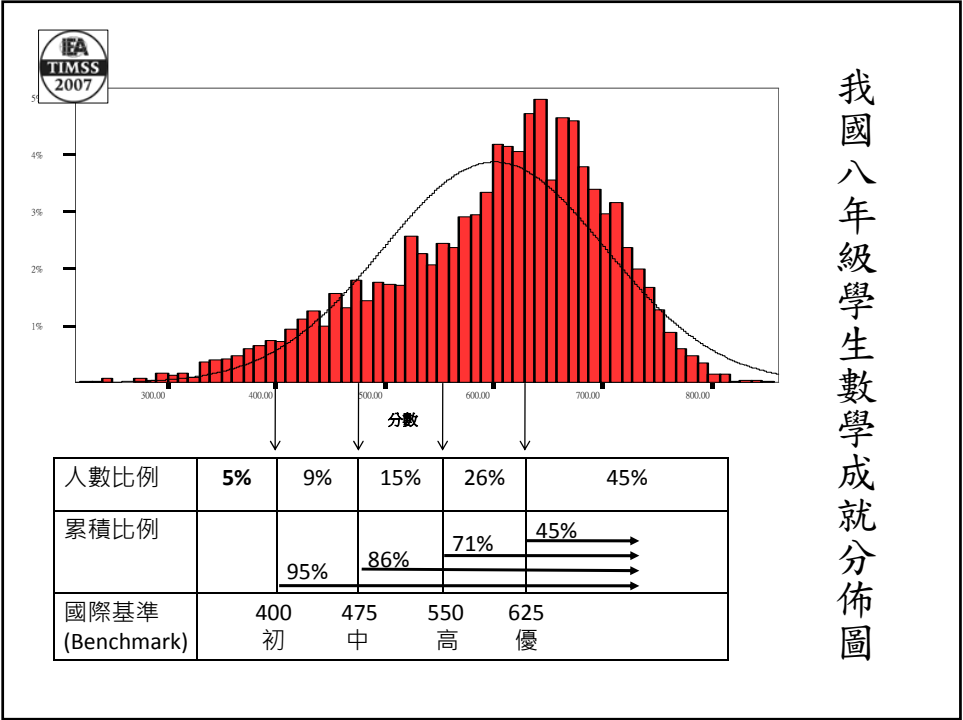




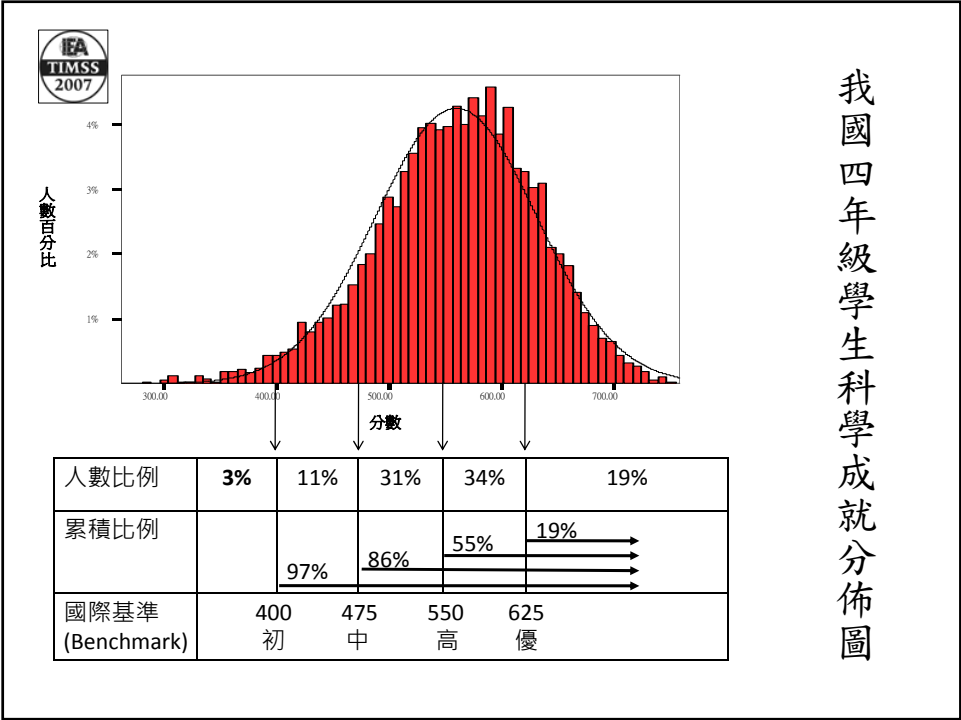
四年級數學成就排名				TIMSS2007 Mathematics 4 th Grade								
				1	2	3	4	5	6	7	8	9
參加國家或地區				香港	新加坡	台灣	日本	哈薩克	俄羅斯	英格蘭	拉托維亞	荷蘭
平均分數												
1	香港	607(3.6)		▲	▲	▲	▲	▲	▲	▲	▲	▲
2	新加坡	599(3.7)		▲	▲	▲	▲	▲	▲	▲	▲	▲
3	台灣	576(1.7)	▼	▼		▲	▲	▲	▲	▲	▲	▲
4	日本	568(2.1)	▼	▼	▼		▲	▲	▲	▲	▲	▲
5	哈薩克	549(7.1)	▼	▼	▼	▼						
6	俄羅斯	544(4.9)	▼	▼	▼	▼						
7	英格蘭	541(2.9)	▼	▼	▼	▼						
8	拉托維亞	537(2.3)	▼	▼	▼	▼						
9	荷蘭	535(2.1)	▼	▼	▼	▼						

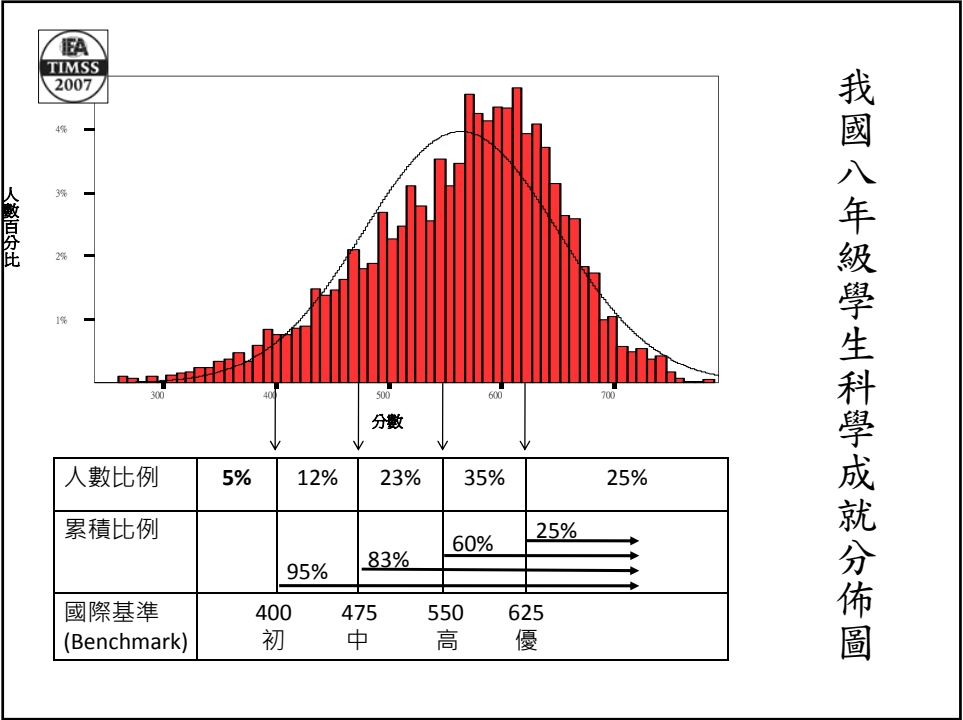
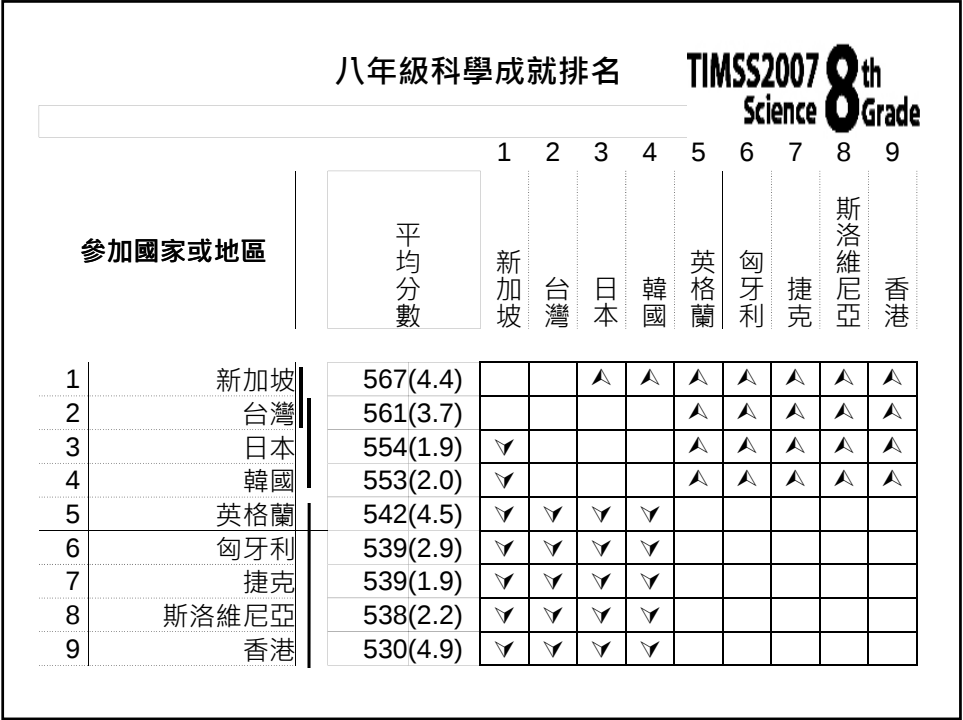


八年級數學成就排名						TIMSS2007 8 th Grade Mathematics					
			1	2	3	4	5	6	7	8	9
參加國家或地區		平均 分數	台灣	韓國	新加坡	香港	日本	匈牙利	英格蘭	俄羅斯	美國
1	台灣	598(4.5)				▲	▲	▲	▲	▲	▲
2	韓國	597(2.7)				▲	▲	▲	▲	▲	▲
3	新加坡	593(3.8)				▲	▲	▲	▲	▲	▲
4	香港	572(5.8)	▼	▼	▼			▲	▲	▲	▲
5	日本	570(2.4)	▼	▼	▼			▲	▲	▲	▲
6	匈牙利	517(3.5)	▼	▼	▼	▼	▼				
7	英格蘭	513(4.8)	▼	▼	▼	▼	▼				
8	俄羅斯	512(4.1)	▼	▼	▼	▼	▼				
9	美國	508(2.8)	▼	▼	▼	▼	▼				



四年級科學成就排名										TIMSS2007		4 th							
										Science		Grade							
										1	2	3	4	5	6	7	8	9	
參加國家或地區										平均分數	新加坡	台灣	香港	日本	俄羅斯	拉托維亞	英格蘭	美國	匈牙利
1	新加坡	587(4.1)		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
2	台灣	557(2.0)	▼			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
3	香港	554(3.5)	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
4	日本	548(2.1)	▼	▼						▲	▲	▲	▲	▲	▲	▲	▲	▲	
5	俄羅斯	546(4.8)	▼	▼						▲	▲	▲	▲	▲	▲	▲	▲	▲	
6	拉托維亞	542(2.3)	▼	▼	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	
7	英格蘭	542(2.9)	▼	▼	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	
8	美國	539(2.7)	▼	▼	▼	▼				▲	▲	▲	▲	▲	▲	▲	▲	▲	
9	匈牙利	536(3.3)	▼	▼	▼	▼	▼			▲	▲	▲	▲	▲	▲	▲	▲	▲	







四年級數學次主題成就排名									
數			幾何形狀與測量			資料呈現			
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績	
1	新加坡	611(4.3)	1	香港	599(3.1)	1	香港	585(2.7)	
2	香港	606(3.8)	2	新加坡	570(3.6)	2	新加坡	583(3.2)	
3	台灣	581(1.9)	3	日本	566(2.2)	3	日本	578(2.8)	
4	日本	561(2.2)	4	台灣	556(2.2)	4	台灣	567(2.0)	
5	哈薩克	556(6.6)	5	英格蘭	548(2.7)	5	英格蘭	547(2.5)	
6	俄羅斯	546(4.4)	6	丹麥	544(2.6)	6	荷蘭	543(2.3)	
7	拉托維亞	536(2.1)	7	哈薩克	542(7.4)	7	美國	543(2.4)	
8	荷蘭	535(2.2)	8	俄羅斯	538(5.1)	8	拉托維亞	536(3.0)	
9	立陶宛	533(2.3)	9	澳洲	536(3.1)	9	澳洲	534(3.1)	
10	英格蘭	531(3.2)	10	拉托維亞	532(2.6)	10	德國	534(3.1)	

八年級數學次主題成就排名

數			代數		
國家 / 地區		平均成績	國家 / 地區		平均成績
1	新加坡	597(3.5)	1	台灣	617(5.4)
2	韓國	583(2.4)	2	韓國	596(3.0)
3	台灣	577(4.2)	3	新加坡	579(3.7)
4	香港	567(5.6)	4	香港	565(5.6)
5	日本	551(2.3)	5	日本	559(2.5)
幾何			資料與機率		
國家 / 地區		平均成績	國家 / 地區		平均成績
1	台灣	592(4.6)	1	韓國	580(2.0)
2	韓國	587(2.3)	2	新加坡	574(3.9)
3	新加坡	578(3.4)	3	日本	573(2.2)
4	日本	573(2.2)	4	台灣	566(3.6)
5	香港	570(5.5)	5	香港	549(4.7)

四年級科學分科成就排名

生命科學			物質科學			地球科學		
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績
1	新加坡	582(4.1)	1	新加坡	585(3.9)	1	香港	560(3.2)
2	義大利	549(3.0)	2	日本	564(2.3)	2	新加坡	554(3.3)
3	匈牙利	548(2.8)	3	台灣	559(2.5)	3	台灣	553(1.9)
4	台灣	541(2.1)	4	香港	558(3.5)	4	英格蘭	538(2.9)
5	美國	540(2.5)	5	俄羅斯	547(4.6)	5	拉托維亞	536(2.2)
6	俄羅斯	539(4.1)	6	拉托維亞	544(2.4)	6	俄羅斯	536(4.3)
7	荷蘭	536(2.2)	7	英格蘭	543(2.7)	7	瑞典	535(2.7)
8	拉托維亞	535(2.1)	8	美國	534(2.3)	8	澳洲	534(3.2)
9	英格蘭	532(2.7)	9	斯洛維尼亞	530(1.6)	9	哈薩克	534(5.2)
10	香港	532(3.5)	10	匈牙利	529(3.3)	10	美國	533(2.6)

八年級科學分科成就排名

生物				化學			
國家 / 地區		平均成績		國家 / 地區		平均成績	
1	新加坡	564	(4.2)	1	台灣	573	(4.2)
2	日本	553	(1.9)	2	新加坡	560	(4.1)
3	台灣	549	(3.4)	3	日本	551	(1.9)
4	韓國	548	(1.9)	4	斯洛維尼亞	539	(2.5)
5	英格蘭	541	(4.4)	5	匈牙利	536	(3.5)

物理				地球科學			
國家 / 地區		平均成績		國家 / 地區		平均成績	
1	新加坡	575	(3.9)	1	台灣	545	(2.9)
2	韓國	571	(2.4)	2	斯洛維尼亞	542	(2.2)
3	日本	558	(1.9)	3	新加坡	541	(4.1)
4	台灣	554	(3.7)	4	韓國	538	(2.2)
5	英格蘭	545	(4.0)	5	捷克	534	(2.0)



四年級數學認知領域成就排名

知識(Knowing)			應用(Applying)			推理(Reasoning)		
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績
1	香港	599(3.4)	1	新加坡	620(4.0)	1	香港	589(3.5)
2	新加坡	590(3.7)	2	香港	617(3.5)	2	新加坡	578(3.8)
3	台灣	569(1.7)	3	台灣	584(1.7)	3	台灣	566(1.9)
4	日本	566(2.0)	4	日本	565(2.1)	4	日本	563(2.1)
5	哈薩克	547(7.2)	5	哈薩克	559(7.3)	5	俄羅斯	540(4.8)
6	俄羅斯	547(4.8)	6	英格蘭	544(3.6)	6	哈薩克	539(6.1)
7	英格蘭	540(3.1)	7	美國	541(2.6)	7	英格蘭	537(3.1)
8	拉托維亞	540(2.5)	8	俄羅斯	538(4.5)	8	拉托維亞	537(2.5)
9	荷蘭	540(2.0)	9	拉托維亞	530(2.2)	9	荷蘭	534(2.4)
10	立陶宛	539(2.4)	10	荷蘭	525(2.2)	10	德國	528(2.5)

八年級數學認知領域成就排名

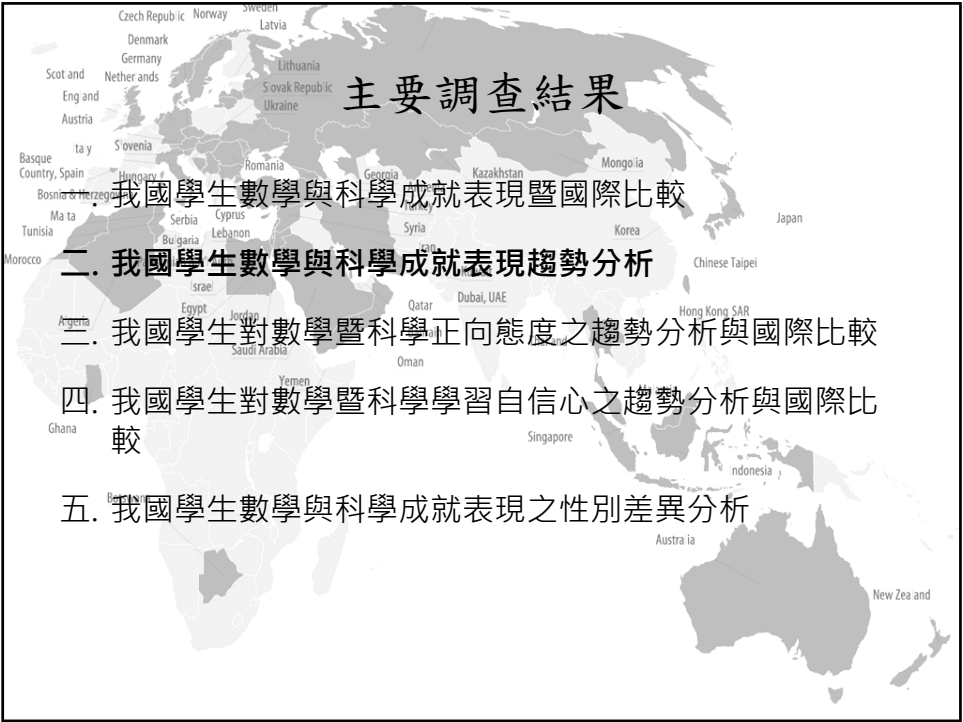
知識(Knowing)			應用(Applying)			推理(Reasoning)		
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績
1	韓國	595(2.8)	1	韓國	596(2.5)	1	台灣	591(4.1)
2	新加坡	593(3.6)	2	台灣	594(4.5)	2	韓國	579(2.3)
3	台灣	592(4.2)	3	新加坡	581(3.4)	3	新加坡	579(4.1)
4	香港	569(5.9)	4	香港	574(5.4)	4	日本	568(2.4)
5	日本	565(2.2)	5	日本	560(2.2)	5	香港	557(5.6)
6	英格蘭	514(4.9)	6	俄羅斯	521(3.9)	6	英格蘭	518(4.3)
7	匈牙利	513(3.1)	7	匈牙利	518(3.3)	7	匈牙利	513(3.2)
8	立陶宛	511(2.4)	8	美國	514(2.6)	8	美國	505(2.4)
9	俄羅斯	510(3.7)	9	立陶宛	508(2.5)	9	澳洲	502(3.3)
10	捷克	504(2.7)	10	亞美尼亞	507(3.1)	10	捷克	500(2.6)

四年級科學認知領域成就排名

知識(Knowing)			應用(Applying)			推理(Reasoning)		
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績
1	新加坡	579(3.7)	1	新加坡	587(4.1)	1	台灣	571(2.4)
2	台灣	556(2.1)	2	香港	546(3.2)	2	新加坡	568(3.7)
3	香港	549(3.0)	3	英格蘭	543(2.9)	3	日本	567(2.1)
4	俄羅斯	546(4.7)	4	俄羅斯	542(4.8)	4	香港	561(4.4)
5	日本	542(2.7)	5	美國	541(2.3)	5	拉托維亞	551(2.7)
6	義大利	539(3.1)	6	匈牙利	540(3.0)	6	俄羅斯	542(4.6)
7	英格蘭	536(2.7)	7	拉托維亞	540(2.2)	7	英格蘭	537(2.7)
8	哈薩克	536(4.9)	8	台灣	536(2.5)	8	美國	535(2.6)
9	拉托維亞	535(2.4)	9	哈薩克	534(5.8)	9	澳洲	530(3.4)
10	美國	533(2.8)	10	義大利	530(3.9)	10	匈牙利	529(3.7)

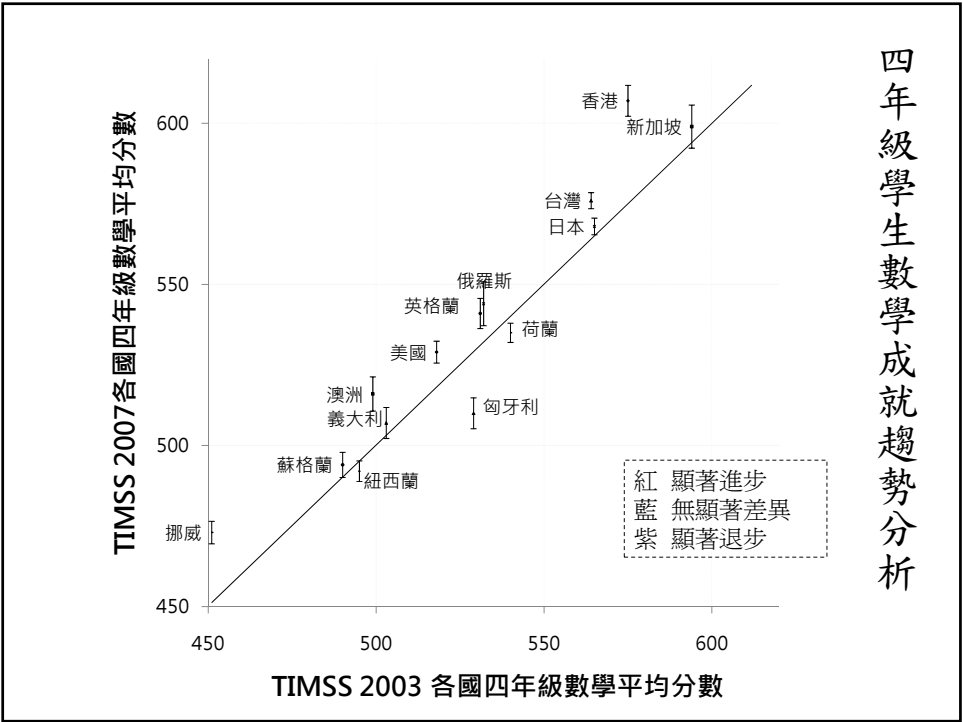
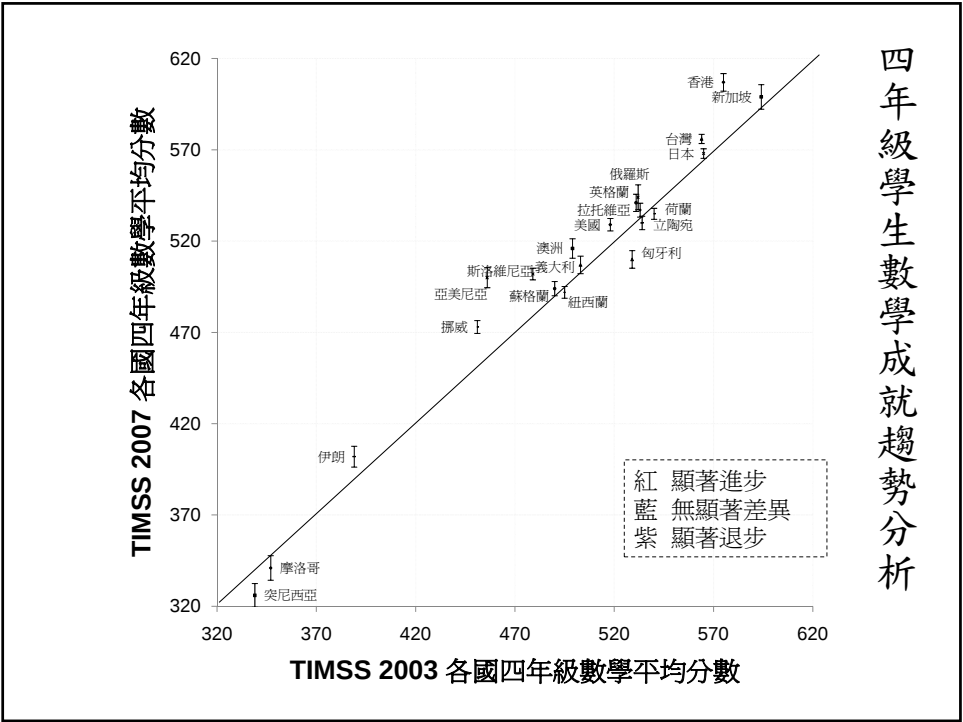
八年級科學認知領域成就排名

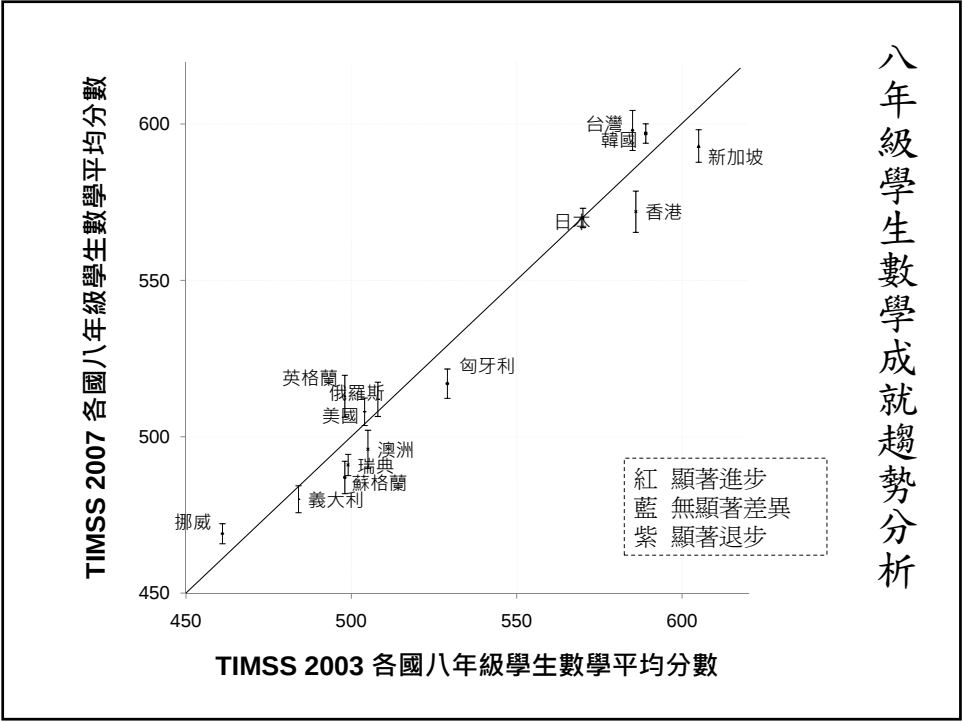
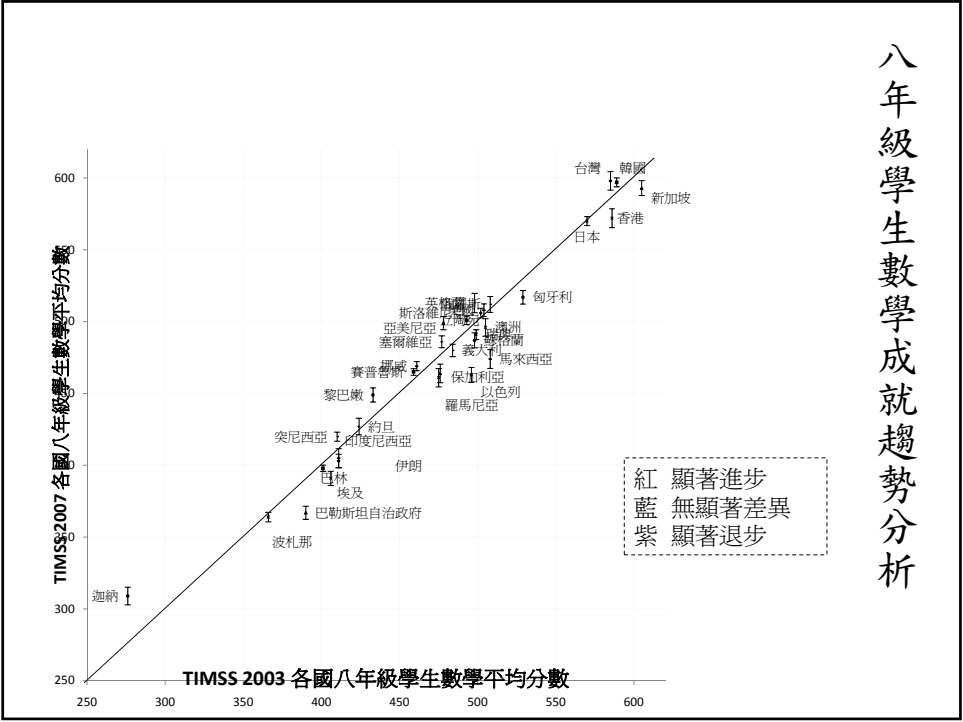
知識(Knowing)			應用(Applying)			推理(Reasoning)		
國家 / 地區		平均成績	國家 / 地區		平均成績	國家 / 地區		平均成績
1	新加坡	567(4.2)	1	台灣	565(3.5)	1	新加坡	564(4.1)
2	台灣	560(3.4)	2	新加坡	554(4.5)	2	日本	560(2.0)
3	日本	555(2.0)	3	韓國	543(2.0)	3	韓國	558(2.0)
4	匈牙利	549(3.0)	4	日本	534(2.2)	4	英格蘭	547(4.0)
5	韓國	547(2.0)	5	俄羅斯	534(4.3)	5	台灣	541(3.5)
6	捷克	539(1.9)	6	捷克	533(2.1)	6	斯洛維尼亞	538(2.2)
7	英格蘭	538(4.0)	7	斯洛維尼亞	533(2.0)	7	捷克	534(2.3)
8	斯洛維尼亞	533(2.2)	8	香港	532(4.5)	8	香港	533(5.0)
9	俄羅斯	527(3.8)	9	英格蘭	530(4.9)	9	澳洲	530(3.6)
10	香港	522(4.9)	10	匈牙利	524(3.0)	10	匈牙利	530(3.0)



亞洲五國數學成就趨勢分析比較

國家 / 地區	四年級數學						八年級數學					
	排名	平均成績	2003 ~ 2007 差異				排名	平均成績	2003 ~ 2007 差異			
台灣								XX				X
2007	3	576(1.7)				X	1	598(4.5)				X
2003	4	564(1.8)	12	(2.5)	▲		4	585(4.6)	13	(6.4)	▲	
韓國								XX				X
2007	--	----	----		--		2	597(2.7)				X
2003	--	----	----		--		2	589(2.2)	8	(3.1)	▲	
新加坡								XX				X
2007	2	599(3.7)					3	593(3.8)				X
2003	1	594(5.6)	5	(6.7)			1	605(3.6)	-13	(5.2)	▼	
香港								XX				X
2007	1	607(3.6)				X	4	572(5.8)				X
2003	2	575(3.2)	32	(4.8)	▲		3	586(3.3)	-14	(6.6)	▼	
日本								XX				X
2007		568(2.1)				X	5	570(2.4)				X
2003	3	565(1.6)	4	(2.6)	X		5	570(2.1)	0	(3.1)	X	

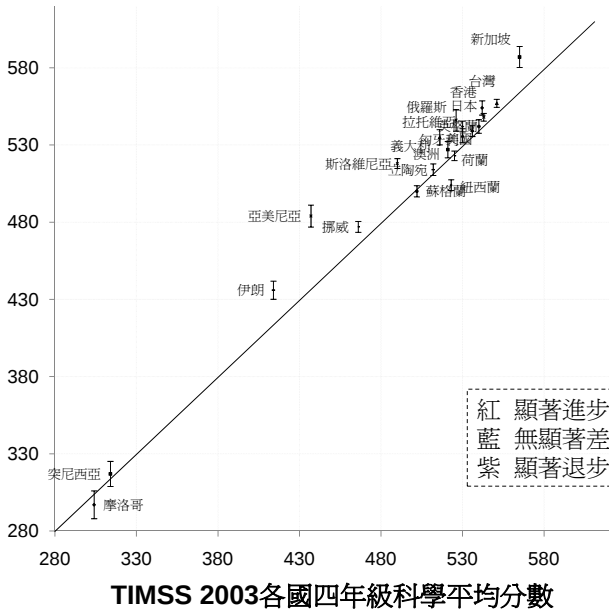




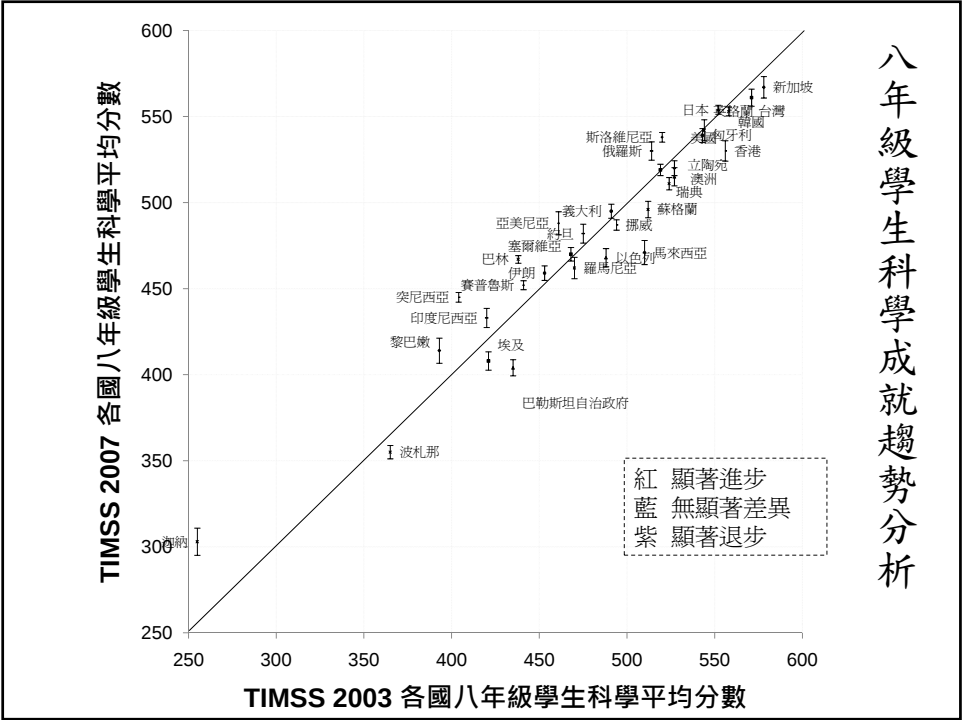
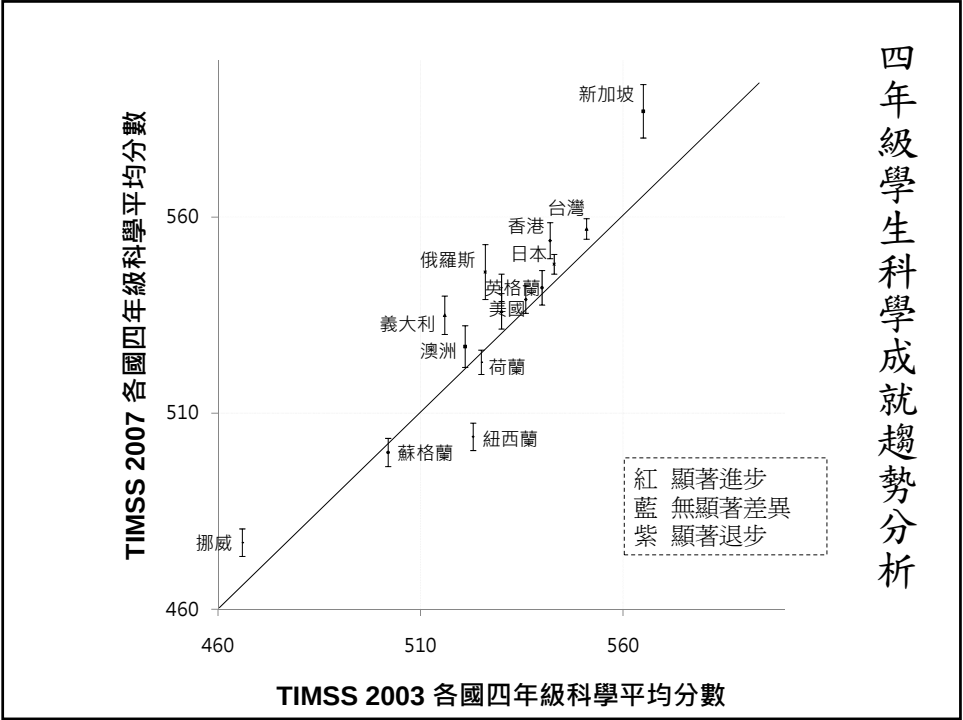
亞洲五國科學成就趨勢分析比較

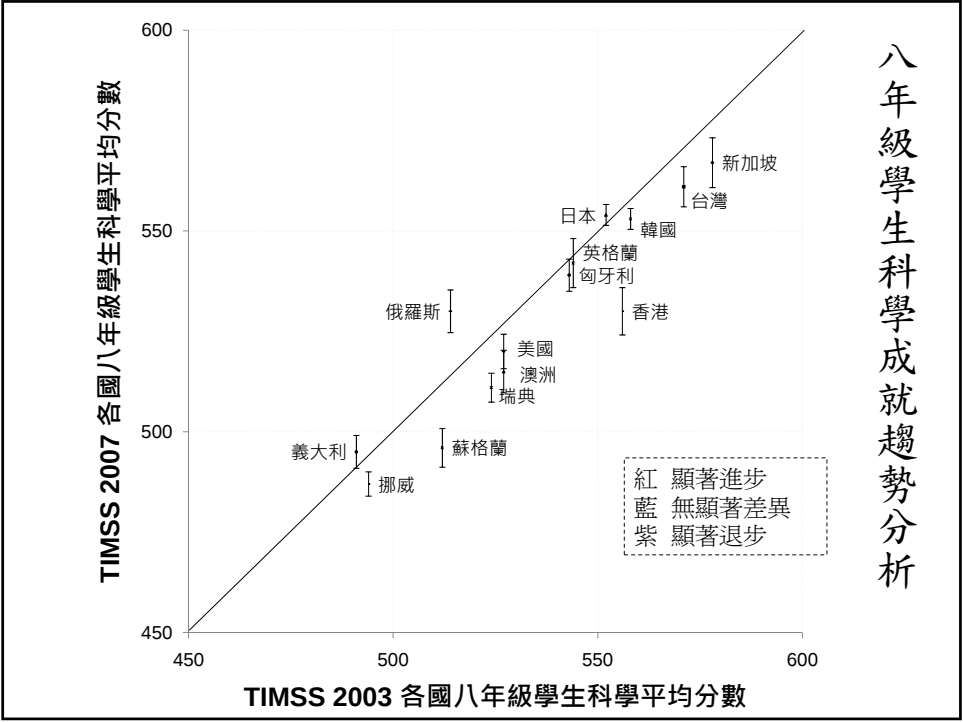
國家 / 地區	四年級科學					八年級科學				
	排名	平均成績	2003 ~ 2007差異			排名	平均成績	2003 ~ 2007差異		
台灣							XX			X
2007	2	557(2.0)			X	2	561(3.7)			
2003	2	551(1.7)	5(2.6)	▲		2	571(3.5)	-10(5.0)	▼	
韓國							XX			
2007	--	---	---	--		4	553(2.0)			
2003	--	---	---	--		3	558(1.6)	-5(2.6)	▼	
新加坡							XX			
2007	1	587(4.1)			X	1	567(4.4)			
2003	1	565(5.5)	22(6.8)	▲		1	578(4.3)	-11(6.2)		
香港							XX			
2007	3	554(3.5)			X	9	530(4.9)			
2003	4	542(3.1)	12(4.6)	▲		4	556(3.0)	-26(5.9)	▼	
日本							XX			
2007	4	548(2.1)			X	3	554(1.9)			X
2003	3	543(1.5)	4(2.5)	X		5	552(1.7)	2(2.6)	X	

TIMSS 2007 各國四年級科學平均分數



四年級學生科學成就趨勢分析





正向態度指標 (PA)

✚ 這個指標主要根據學生對於背景問卷中三項陳述的同意程度估算而得，此三項陳述分別為：

1. 我喜歡學習 數學/自然科學
2. 數學/自然科學 很無趣(反向題)
3. 我喜歡 數學/自然科學

試題的反應分為：

1. 非常同意
2. 有些同意
3. 不太同意
4. 非常不同意

* 如果三題答案皆為1或2的學生被歸為高度正向態度，三題答案皆為3或4的學生被歸為低度正向態度；其餘則歸為中度正向態度的學生。

四年級學生對數學的正向態度指標 (PATM) 的趨勢分析

國家/地區	平均 成績 (排序)	高 PATM			中 PATM			低 PATM		
		2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異
1 喬治亞	438(27)	90(0.9)	450(3.7)	◇◇	6(0.6)	415(9.0)	◇◇	4(0.5)	415(10.2)	◇◇
量尺平均	500(19.5)	72(0.2)	483(0.6)		14(0.1)	457(1.1)		14(0.1)	454(1.3)	
17 新加坡	599(2)	71(0.8)	610(3.5)	-15(1.1) ▼	14(0.6)	575(5.9)	5(0.8) ▲	15(0.6)	575(5.6)	11(0.7) ▲
19 德國	525(12)	70(0.9)	534(2.7)	◇◇	16(0.6)	520(3.7)	◇◇	14(0.7)	509(3.6)	◇◇
21 挪威	473(25)	68(1.2)	478(3.1)	-4(2.1)	15(0.6)	470(5.1)	1(1.1)	18(1.0)	462(3.7)	3(1.6) ▲
22 瑞典	503(18)	67(1.2)	505(2.8)	◇◇	16(0.7)	501(3.2)	◇◇	17(1.0)	497(4.1)	◇◇
23 香港	607(1)	67(1.3)	619(3.5)	-5(1.8) ▼	15(0.7)	588(4.2)	-2(1.2)	19(1.1)	579(5.1)	7(1.5) ▲
24 澳洲	516(14)	66(1.4)	525(3.6)	-7(1.7) ▼	16(0.8)	512(4.6)	3(1.0) ▲	18(1.1)	494(5.1)	4(1.2) ▲
25 美國	529(11)	66(0.8)	535(2.7)	-8(1.4) ▼	16(0.5)	526(3.0)	3(0.8) ▲	18(0.6)	517(2.5)	5(1.0) ▲
26 紐西蘭	492(23)	66(1.0)	499(2.6)	-5(1.6) ▼	18(0.8)	485(3.8)	3(1.2) ▲	17(0.8)	484(3.3)	3(1.2) ▲
31 奧地利	505(17)	62(1.0)	513(2.0)	-5(1.7) ▼	16(0.7)	499(4.1)	1(1.1)	22(0.9)	492(2.9)	4(1.3) ▲
32 日本	568(4)	62(1.4)	584(2.4)	-1(1.8)	21(0.8)	547(3.3)	-1(1.2)	17(1.0)	543(4.4)	3(1.2) ▲
33 蘇格蘭	494(22)	59(1.3)	497(2.7)	--	18(0.8)	496(3.5)	--	24(1.1)	490(3.9)	--
34 荷蘭	535(9)	56(1.4)	540(2.7)	-5(2.0) ▼	17(0.9)	531(3.6)	3(1.2) ▲	27(1.3)	528(3.4)	3(1.9)
35 丹麥	523(13)	55(1.8)	526(3.0)	◇◇	24(1.0)	521(3.4)	◇◇	21(1.4)	523(3.1)	◇◇
36 台灣	576(3)	50(1.2)	595(2.4)	◇◇	21(0.8)	563(3.2)	◇◇	29(0.9)	555(2.9)	◇◇

八年級學生對數學的正向態度指標 (PATM) 的趨勢分析

國家/地區	平均 成績 (排序)	高 PATM			中 PATM			低 PATM		
		2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異
1 阿爾及利亞	387(39)	83(0.7)	394(2.2)	◇◇	10(0.5)	364(3.9)	◇◇	7(0.5)	357(3.8)	◇◇
16 新加坡	593(3)	60(1.0)	615(3.6)	-7(1.5) ▼	20(0.6)	575(5.3)	1(1.0)	20(0.8)	545(5.4)	6(1.1) ▲
量尺平均	500(12.5)	54(0.2)	471(0.6)		21(0.1)	441(0.7)		26(0.1)	428(0.7)	
29 香港	572(4)	47(1.2)	603(5.5)	-9(1.6) ▼	22(0.9)	566(6.4)	-2(1.1)	31(1.2)	532(7.3)	11(1.4) ▲
33 美國	508(9)	41(0.8)	524(2.9)	-11(1.4) ▼	24(0.5)	511(3.3)	2(0.8) ▲	35(0.8)	490(3.3)	9(1.2) ▲
35 英格蘭	513(7)	40(1.4)	532(5.7)	-25(1.9) ▼	25(0.9)	515(6.1)	6(1.2) ▲	35(1.5)	495(4.9)	19(1.8) ▲
36 瑞典	491(15)	39(1.1)	517(2.9)	◇◇	24(0.6)	488(2.9)	◇◇	37(1.1)	470(2.7)	◇◇
39 台灣	598(1)	37(1.2)	657(3.7)	-8(1.6) ▼	18(0.6)	605(5.1)	-4(0.8) ▼	45(1.4)	547(4.6)	12(1.7) ▲
40 挪威	469(21)	37(1.1)	488(2.4)	◇◇	24(0.6)	474(2.6)	◇◇	39(1.1)	451(2.1)	◇◇
42 澳洲	496(14)	34(1.3)	521(6.2)	--	27(0.8)	498(3.7)	--	39(1.2)	476(4.1)	--
43 蘇格蘭	487(17)	33(1.0)	502(4.5)	◇◇	29(0.8)	490(4.1)	◇◇	38(1.0)	476(4.1)	◇◇
44 韓國	597(2)	33(0.9)	650(2.9)	3(1.1) ▲	23(0.6)	600(3.4)	-12(0.9) ▼	44(0.9)	558(3.1)	8(1.3) ▲
46 日本	570(5)	30(1.1)	609(3.7)	-1(1.5)	30(1.0)	567(3.0)	-4(1.2) ▼	40(1.2)	543(2.5)	5(1.6) ▲
48 斯洛伐尼亞	501(12)	25(1.1)	520(4.3)	--	22(0.7)	507(3.0)	--	53(1.1)	353(9.4)	--

四年級學生對科學的正向態度指標 (PATS) 的趨勢分析

國家/地區	平均 成績 (排序)	高 PATS			中 PATS			低 PATS		
		2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異
1 哈薩克	533(11)	90(1.1)	535(5.7)	◇◇	6(0.9)	517(12.8)	◇◇	4(0.5)	515(10.3)	◇◇
10 日本	548(4)	81(0.9)	553(2.1)	1(1.4)	12(0.6)	534(4.1)	-1(0.9)	7(0.5)	523(6.1)	1(0.8)
12 德國	528(12)	81(0.8)	536(2.5)	◇◇	11(0.6)	514(5.4)	◇◇	8(0.5)	501(4.7)	◇◇
14 香港	554(3)	79(1.0)	562(3.4)	-2(2.1)	11(0.6)	528(5.2)	-1(1.0)	10(0.8)	522(5.4)	3(1.6) ▲
16 義大利	535(10)	78(0.8)	541(3.3)	--	12(0.6)	522(4.1)	--	10(0.5)	516(5.2)	--
17 澳洲	527(13)	78(1.3)	534(3.6)	4(1.7) ▲	11(0.8)	513(5.8)	-1(1.0)	11(0.8)	505(5.1)	-2(1.1) ▼
18 俄羅斯	546(5)	78(1.0)	552(4.5)	◇◇	13(0.8)	540(7.2)	◇◇	9(0.5)	521(8.1)	◇◇
量尺平均	500(23.5)	77(0.2)	485(0.7)		13(0.1)	456(1.2)		11(0.1)	452(1.3)	
22 美國	539(8)	75(0.8)	545(2.5)	-1(1.3)	13(0.4)	529(4.1)	0(0.8)	12(0.6)	521(4.4)	1(0.9)
23 台灣	557(2)	75(1.4)	564(2.0)	◇◇	14(0.7)	539(4.2)	◇◇	11(0.9)	534(4.5)	◇◇
24 奧地利	526(15)	75(0.9)	530(2.6)	6(1.7) ▲	13(0.6)	518(4.1)	-4(1.1) ▼	12(0.7)	510(4.2)	-2(1.3)
25 新加坡	587(1)	75(0.7)	598(4.0)	-9(1.1) ▼	15(0.5)	557(6.0)	3(0.8) ▲	11(0.5)	553(5.8)	5(0.6) ▲
27 瑞典	525(16)	73(1.0)	531(3.2)	◇◇	14(0.7)	512(4.0)	◇◇	13(0.6)	509(4.1)	◇◇
28 挪威	477(25)	71(1.2)	484(3.4)	1(1.9)	14(0.8)	468(4.9)	-2(1.2)	15(0.9)	463(6.8)	1(1.4)
33 荷蘭	523(17)	66(1.5)	528(2.8)	3(2.2)	11(0.6)	514(4.1)	-4(1.1) ▼	23(1.3)	515(4.0)	1(1.8)
35 英格蘭	542(7)	59(1.2)	548(3.4)	-13(1.9) ▼	17(0.7)	538(4.7)	4(1.1) ▲	24(1.1)	533(4.1)	9(1.5) ▲
36 丹麥	517(19)	59(1.9)	525(3.2)	◇◇	20(1.1)	510(4.3)	◇◇	21(1.5)	507(4.1)	◇◇

八年級學生對科學的正向態度指標 (PATS) 的趨勢分析 *

國家/地區	平均 成績 (排序)	高 PATS			中 PATS			低 PATS		
		2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異	2007 百分比 (標準誤)	平均成績	1999-2007 差異
1 突尼西亞	445(35)	88(0.8)	447(2.1)	1(1.1)	7(0.5)	428(4.7)	1(0.7)	4(0.4)	434(7.1)	-2(0.7) ▼
15 新加坡	567(1)	68(0.9)	586(4.3)	-10(1.6) ▼	19(0.7)	535(6.1)	7(1.0) ▲	13(0.6)	517(6.8)	3(1.1) ▲
量尺平均	500(14.5)	65(0.2)	476(0.7)		19(0.1)	442(0.9)		16(0.2)	436(1.3)	
18 香港	530(9)	60(1.4)	549(4.8)	-4(1.9) ▼	22(0.8)	508(5.4)	4(1.1) ▲	19(1.1)	498(6.7)	0(1.5)
20 挪威	487(18)	59(1.3)	495(2.1)	◇◇	20(0.8)	483(3.7)	◇◇	21(1.0)	472(3.7)	◇◇
21 蘇格蘭	496(15)	56(1.3)	517(3.4)	◇◇	22(0.8)	482(4.5)	◇◇	22(1.1)	459(4.3)	◇◇
22 英格蘭	542(5)	55(1.3)	561(4.9)	-21(1.6) ▼	20(0.8)	532(4.9)	9(1.1) ▲	25(1.1)	510(5.0)	12(1.3) ▲
23 美國	520(11)	54(1.2)	533(2.9)	-6(1.5) ▼	22(0.5)	508(3.6)	4(0.8) ▲	24(0.9)	503(3.5)	2(1.1)
25 義大利	495(16)	47(1.1)	511(3.7)	-16(1.8) ▼	26(0.9)	488(3.6)	10(1.2) ▲	26(1.0)	475(3.5)	7(1.5) ▲
26 澳洲	515(13)	47(1.4)	535(4.7)	--	22(0.8)	504(4.2)	--	31(1.1)	494(4.3)	--
27 日本	554(3)	47(1.1)	574(2.2)	1(1.8)	28(0.9)	545(2.8)	5(1.1) ▲	25(1.1)	529(3.5)	-6(1.7) ▼
28 台灣	561(2)	40(1.3)	597(3.9)	-22(1.7) ▼	24(0.7)	552(4.2)	8(0.9) ▲	35(1.2)	527(3.8)	14(1.4) ▲
29 韓國	553(4)	38(1.1)	586(2.4)	8(1.5) ▲	27(0.7)	544(2.9)	-3(0.9) ▼	36(1.0)	526(2.6)	-5(1.5) ▼

本表僅針對29個國家進行排序，未包含20個採科學分科課程的國家

主要調查結果

- 一. 我國學生數學與科學成就表現暨國際比較
- 二. 我國學生數學與科學成就表現趨勢分析
- 三. 我國學生對數學暨科學正向態度之趨勢分析與國際比較
- 四. 我國學生對數學暨科學學習自信心之趨勢分析與國際比較
- 五. 我國學生數學與科學成就表現之性別差異分析

自信心指標 (SC)

✚ 這個指標主要根據學生對於背景問卷中四項陳述的同意程度估算而得，此四項陳述分別為：

1. 我在 數學/自然科學 方面通常表現不錯
2. 比起其他同學來說，數學/自然科學 對我來說是比較困難的(反向題)
3. 我對 數學/自然科學 不在行(反向題)
4. 我對關於 數學/自然科學 方面學得很快

學生的反應分為：

1. 非常同意
2. 有些同意
3. 不太同意
4. 非常不同意

* 如果三題答案皆為1或2的學生被歸為具高學習自信心，三題答案皆為3或4的學生被歸為具低學習自信心；其餘則歸為具中度學習自信心的學生。

四年級學生對學習數學的自信心指標 (SCM) 的趨勢分析

國家/地區	平均 成績 (排序)	高 SCM			中 SCM			低 SCM		
		2007 百分比 (標準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異
1 瑞典	503(18)	77 (0.9)	514 (2.4)	◇ ◇	19 (0.8)	467 (4.5)	◇ ◇	5 (0.4)	459 (5.2)	◇ ◇
2 奧地利	505(17)	70 (0.8)	524 (1.9)	◇ ◇	22 (0.8)	470 (2.6)	◇ ◇	8 (0.5)	445 (5.5)	◇ ◇
3 德國	525(12)	70 (0.9)	548 (2.1)	◇ ◇	21 (0.7)	493 (3.5)	◇ ◇	10 (0.5)	468 (4.5)	◇ ◇
4 丹麥	523(13)	70 (1.1)	540 (2.5)	◇ ◇	23 (1.0)	493 (3.5)	◇ ◇	7 (0.6)	469 (5.8)	◇ ◇
5 挪威	473(25)	69 (0.8)	490 (2.7)	5 (1.2) ▲	24 (0.8)	441 (3.8)	-3 (1.2) ▼	7 (0.4)	429 (7.1)	-1 (0.7)
8 美國	529(11)	67 (0.8)	551 (2.4)	13 (1.1) ▲	22 (0.6)	493 (2.8)	-17 (0.9) ▼	10 (0.4)	478 (3.1)	3 (0.5) ▲
9 蘇格蘭	494(22)	67 (1.1)	511 (2.4)	3 (1.4) ▲	24 (1.0)	472 (3.5)	-2 (1.3)	9 (0.6)	450 (5.3)	-1 (0.8)
10 荷蘭	535(9)	66 (1.0)	551 (2.5)	0 (1.4)	22 (0.9)	511 (3.1)	-1 (1.3)	12 (0.7)	489 (4.2)	1 (0.9)
14 英格蘭	541(7)	64 (1.0)	566 (3.0)	5 (1.5) ▲	26 (0.8)	507 (3.7)	-4 (1.3) ▼	10 (0.7)	483 (5.0)	-1 (0.9)
15 澳洲	516(14)	64 (1.3)	542 (2.8)	0 (1.6)	26 (0.9)	480 (3.8)	1 (1.2)	10 (0.8)	457 (6.7)	-1 (1.1)
量尺平均	500(19.5)	57 (0.2)	500 (0.6)		32 (0.2)	449 (0.8)		11 (0.1)	429 (1.2)	
24 紐西蘭	492(23)	52 (0.7)	527 (2.3)	13 (1.3) ▲	37 (0.7)	465 (2.6)	-19 (1.2) ▼	11 (0.5)	438 (4.8)	7 (0.6) ▲
28 新加坡	599(2)	46 (1.2)	639 (3.0)	-3 (2.0)	35 (0.8)	580 (3.8)	-1 (1.3)	19 (0.8)	544 (4.9)	3 (1.2) ▲
29 香港	607(1)	46 (1.0)	634 (3.7)	6 (1.5) ▲	38 (1.0)	588 (3.6)	-3 (1.3) ▼	16 (0.7)	574 (4.6)	-3 (1.1) ▼
32 日本	568(4)	45 (1.1)	602 (2.4)	6 (1.4) ▲	36 (0.9)	553 (2.9)	-5 (1.2) ▼	20 (0.7)	522 (3.1)	-1 (1.1)
35 台灣	576(3)	36 (1.0)	612 (2.1)	-5 (1.3) ▼	37 (0.8)	566 (2.7)	-2 (1.2)	27 (0.8)	542 (2.7)	7 (1.1) ▲
36 澳門	224(36)	35 (1.5)	261 (7.4)	◇ ◇	52 (1.5)	225 (5.8)	◇ ◇	13 (1.0)	210 (9.6)	◇ ◇

八年級學生對學習數學的自信心指標 (SCM) 的趨勢分析

國家/地區	平均 成績 (排序)	高 SCM			中 SCM			低 SCM		
		2007 百分比 (標準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異
1 以色列	463(24)	59(1.0)	495(4.1)	0(1.5)	29(0.9)	432(5.3)	-1(1.3)	12(0.7)	417(7.2)	1(1.0)
7 美國	508(9)	53(1.0)	537(2.5)	2(1.3)	28(0.7)	487(3.2)	-1(0.9)	19(0.7)	462(3.0)	-1(1.0)
8 英格蘭	513(7)	53(1.4)	543(4.9)	6(2.1) ▲	32(1.0)	494(4.7)	-2(1.6)	15(0.8)	457(5.5)	-4(1.3) ▼
11 挪威	469(21)	50(0.8)	505(2.1)	4(1.4) ▲	31(0.7)	450(2.1)	-1(1.1)	19(0.7)	415(2.2)	-2(1.1) ▼
12 瑞典	491(15)	49(1.0)	528(2.6)	1(1.6)	35(0.7)	468(2.4)	-1(1.2)	16(0.6)	438(3.6)	0(1.1)
6 蘇格蘭	487(17)	53(1.3)	515(4.0)	1(2.0)	33(1.0)	456(3.6)	1(1.4)	14(0.7)	442(4.6)	-1(1.2)
20 澳洲	496(14)	45(1.2)	539(4.8)	-5(2.1) ▼	35(0.8)	472(4.1)	5(1.3) ▲	19(0.9)	445(3.7)	0(1.5)
量尺平均	500(12.5)	43(0.2)	492(0.6)		37(0.1)	433(0.6)		20(0.1)	412(0.7)	
30 新加坡	593(3)	41(1.0)	638(3.3)	2(1.3)	34(0.9)	572(4.6)	0(1.1)	25(0.8)	547(4.7)	-2(1.0)
42 香港	572(4)	30(1.1)	622(5.1)	1(1.4)	40(1.0)	562(6.7)	2(1.2)	30(0.7)	539(5.8)	-2(1.1) ▼
43 韓國	597(2)	29(0.8)	668(2.6)	-2(1.1)	34(0.7)	606(3.1)	-2(1.0) ▼	38(0.8)	536(2.8)	4(1.2) ▲
46 台灣	598(1)	27(1.1)	674(3.7)	1(1.5)	27(0.7)	610(5.0)	-3(1.0) ▼	46(1.2)	547(4.4)	2(1.6)
48 日本	570(5)	17(0.6)	638(3.9)	0(0.9)	35(0.8)	586(2.9)	-3(1.1) ▼	48(0.9)	535(2.6)	2(1.2) ▲

四年級學生對學習科學的自信心指標 (SCS) 的趨勢分析

國家/地區	平均 成績 (排序)	高 SCS			中 SCS			低 SCS		
		2007 百分比 (標準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異
1 奧地利	526(15)	79 (0.9)	539 (2.4)	◇◇	16(0.7)	479(4.1)	◇◇	5(0.5)	477(7.2)	◇◇
2 德國	528(12)	76 (0.8)	544 (2.5)	◇◇	18(0.6)	491(4.7)	◇◇	5(0.4)	469(6.5)	◇◇
3 瑞典	525(16)	76 (0.9)	534 (3.0)	◇◇	20(0.7)	498(4.5)	◇◇	4(0.4)	484(8.1)	◇◇
7 美國	539(8)	69 (0.7)	556 (2.4)	14(1.2) ▲	22(0.5)	508(3.8)	-17(1.0) ▼	8(0.4)	493(4.7)	3(0.5) ▲
8 義大利	535(10)	69 (0.9)	548 (3.4)	0(1.4)	25(0.7)	514(3.6)	-1(1.2)	6(0.4)	496(6.6)	0(0.6)
10 丹麥	517(19)	68 (1.4)	531 (2.8)	◇◇	26(1.1)	494(4.3)	◇◇	7(0.6)	485(6.3)	◇◇
11 挪威	477(25)	67 (1.3)	492 (3.1)	3(1.7)	26(1.0)	454(4.8)	-3(1.4) ▼	7(0.5)	436(8.4)	0(0.7)
12 荷蘭	523(17)	67 (1.3)	535 (2.7)	-4(1.8) ▼	25(1.1)	504(3.8)	3(1.4) ▲	8(0.6)	490(5.5)	1(0.9)
17 俄羅斯	546(5)	63 (1.2)	563 (4.1)	0(1.8)	27(1.1)	523(6.9)	0(1.6)	10(0.7)	520(7.8)	0(1.1)
18 澳洲	527(13)	63 (1.0)	543 (3.0)	-3(1.6)	28(0.7)	509(4.4)	1(1.3)	9(0.7)	483(6.7)	2(0.8) ▲
19 蘇格蘭	500(23)	62 (1.2)	514 (2.6)	5(1.8) ▲	26(1.0)	485(3.9)	-4(1.5) ▼	11(0.8)	468(4.3)	-1(1.0)
量尺平均	500(23.5)	61 (0.2)	497 (0.7)		30(0.2)	453(0.9)		8(0.1)	437(1.5)	
23 台灣	557(2)	58 (1.2)	572 (2.3)	8(1.6) ▲	33(0.9)	538(2.9)	-4(1.2) ▼	9(0.7)	533(4.4)	-4(1.0) ▼
28 英格蘭	542(7)	55 (1.1)	561 (3.4)	2(1.6)	31(0.8)	524(3.6)	-2(1.2)	14(0.8)	512(4.8)	0(1.1)
29 日本	548(4)	53 (1.2)	562 (2.4)	8(1.6) ▲	35(1.0)	537(2.8)	-6(1.3) ▼	12(0.6)	521(4.2)	-2(0.9)
30 香港	554(3)	52 (1.3)	571 (3.4)	-8(1.9) ▼	38(1.0)	539(4.1)	6(1.5) ▲	11(0.7)	528(5.4)	2(0.8) ▲
36 新加坡	587(1)	41 (0.9)	621 (4.0)	9(1.3) ▲	38(0.7)	568(4.9)	-3(1.1) ▼	21(0.6)	556(5.0)	-6(1.0) ▼

八年級學生對學習科學的自信心指標 (SCS) 的趨勢分析 *

國家/地區	平均 成績 (排序)	高 SCS			中 SCS			低 SCS		
		2007 百分比 (標準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異	2007 百分比(標 準誤)	平均成績	2003-2007 差異
1 突尼西亞	445(35)	70(0.9)	457(2.2)	1(1.4)	26(0.8)	417(2.3)	0(1.2)	4(0.4)	417(6.2)	0(0.5)
8 挪威	487(18)	57(1.2)	507(2.0)	-3(1.7)	32(0.9)	467(2.8)	1(1.3)	11(0.7)	447(4.5)	2(1.0)
10 美國	520(11)	56(1.1)	543(3.1)	0(1.4)	29(0.7)	498(3.2)	-2(1.0)	15(0.7)	482(3.8)	2(0.9)
12 英格蘭	542(5)	53(1.5)	569(4.7)	1(2.1)	31(1.1)	517(5.5)	-1(1.7)	15(0.9)	504(4.7)	0(1.3)
13 義大利	495(16)	53(1.0)	517(3.6)	-4(1.5) ▼	33(0.9)	476(3.3)	1(1.3)	14(0.8)	460(4.6)	3(1.0) ▲
17 蘇格蘭	496(15)	52(1.4)	530(3.2)	-7(2.0) ▼	31(1.1)	468(3.6)	4(1.5) ▲	17(1.0)	447(4.5)	3(1.3) ▲
量尺平均	500(14.5)	48(0.2)	492(0.7)		38(0.2)	439(0.7)		13(0.1)	427(1.3)	
22 澳洲	515(13)	41(1.3)	549(4.9)	-7(2.0) ▼	39(1.0)	496(3.7)	4(1.5) ▲	20(1.0)	483(4.3)	3(1.3) ▲
23 新加坡	567(1)	40(1.0)	601(4.5)	-5(1.3) ▼	38(0.9)	544(5.4)	2(1.1)	21(0.7)	546(6.0)	3(0.9) ▲
24 香港	530(9)	33(1.3)	561(4.9)	1(1.7)	49(0.9)	516(5.1)	2(1.2)	18(1.0)	515(5.9)	-2(1.4)
27 韓國	553(4)	24(1.0)	603(2.5)	4(1.2) ▲	40(0.9)	556(2.4)	-2(1.1)	36(0.9)	516(2.5)	-2(1.3)
28 台灣	561(2)	23(1.0)	619(4.0)	-4(1.4) ▼	36(0.9)	552(4.2)	-2(1.2)	41(1.2)	536(3.3)	7(1.6) ▲
29 日本	554(3)	20(0.7)	601(2.8)	0(1.1)	44(1.0)	554(2.4)	-2(1.2)	36(1.1)	529(2.8)	2(1.5)

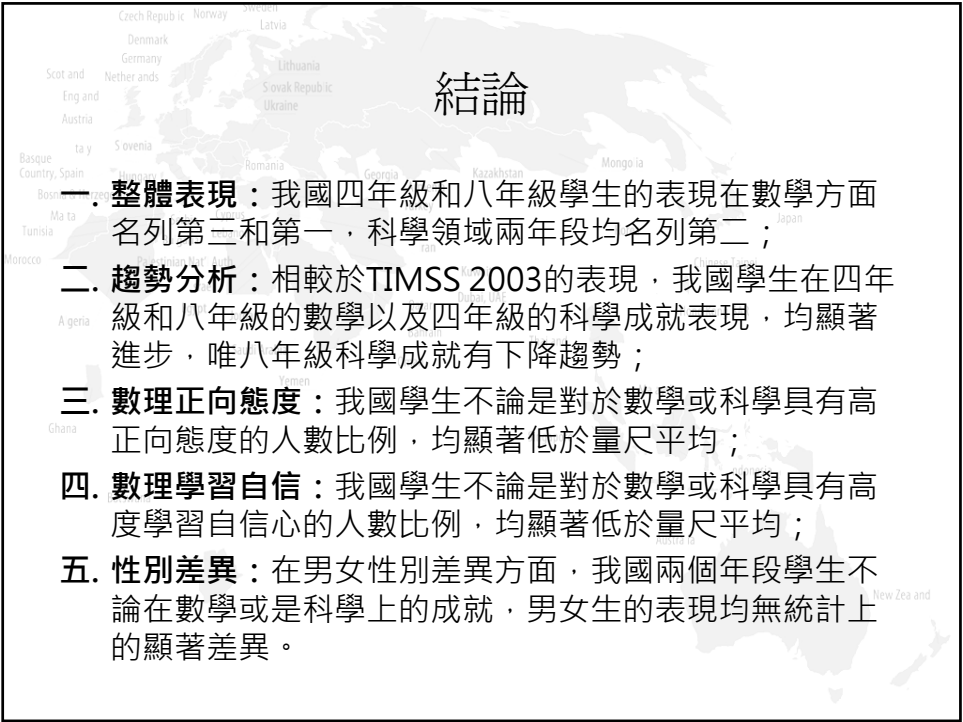
*本表僅針對29個國家進行排序，未包含20個採科學分科課程的國家

主要調查結果

- 一. 我國學生數學與科學成就表現暨國際比較
- 二. 我國學生數學與科學成就表現趨勢分析
- 三. 我國學生對數學暨科學正向態度之趨勢分析與國際比較
- 四. 我國學生對數學暨科學學習自信心之趨勢分析與國際比較
- 五. 我國學生數學與科學成就表現之性別差異分析

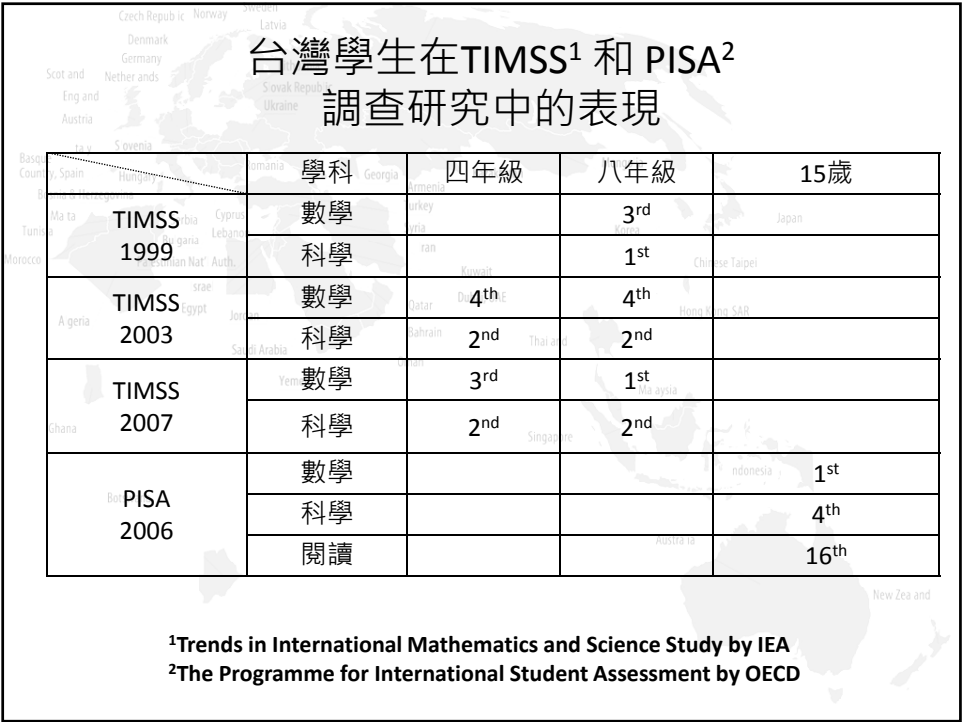






結論

- 一. 整體表現：我國四年級和八年級學生的表現在數學方面名列第三和第一，科學領域兩年段均名列第二；
- 二. 趨勢分析：相較於TIMSS 2003的表現，我國學生在四年級和八年級的數學以及四年級的科學成就表現，均顯著進步，唯八年級科學成就有下降趨勢；
- 三. 數理正向態度：我國學生不論是對於數學或科學具有高正向態度的人數比例，均顯著低於量尺平均；
- 四. 數理學習自信：我國學生不論是對於數學或科學具有高度學習自信心的人數比例，均顯著低於量尺平均；
- 五. 性別差異：在男女性別差異方面，我國兩個年段學生不論在數學或是科學上的成就，男女生的表現均無統計上的顯著差異。



台灣學生在TIMSS¹ 和 PISA² 調查研究中的表現

	學科	四年級	八年級	15歲
TIMSS 1999	數學		3 rd	
	科學		1 st	
TIMSS 2003	數學	4 th	4 th	
	科學	2 nd	2 nd	
TIMSS 2007	數學	3 rd	1 st	
	科學	2 nd	2 nd	
PISA 2006	數學			1 st
	科學			4 th
	閱讀			16 th

¹Trends in International Mathematics and Science Study by IEA
²The Programme for International Student Assessment by OECD



附錄二

TIMSS 2007 八年級科學可公布試題（中文版）



海洋中的生命

說明：第 **35**、**36**、**37** 題與海洋中的生命有關。在回答這些問題時，你可以使用在這個部份的任何資料。

淡水所佔地表上所有水的比例最接近下列哪一個百分比？

- Ⓐ 100%
- Ⓑ 90%
- Ⓒ 70%
- Ⓓ 3%

S032115

	溫度	溶解的鹽	水的體積	密度
純水	25°C	0 克	100 毫升	1.0 克/毫升
鹽水	25°C	10 克	100 毫升	?

上面未完成的表格為純水和鹽水一些數據，此鹽水的密度為何？

(請勾選一項)

- ☐ 1.0克/毫升
- ☐ 小於1.0克/毫升
- ☐ 大於1.0克/毫升

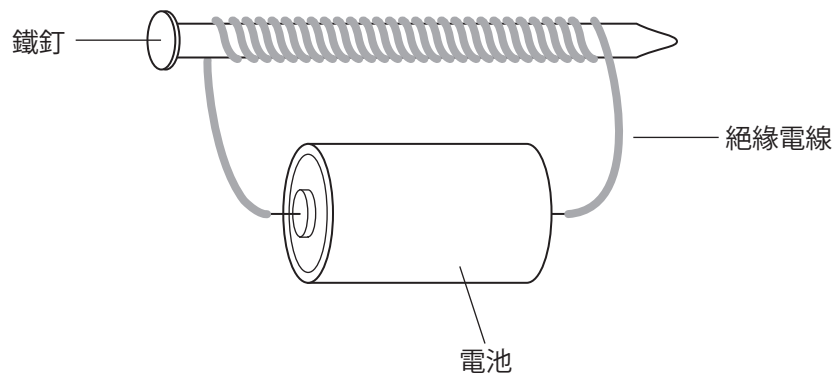
請解釋你的答案。

S032565

關於液體粒子和氣體粒子的比較，下列敘述何者正確？

- Ⓐ 液體粒子速率較慢且相距較遠
- Ⓑ 液體粒子速率較快且相距較遠
- Ⓒ 液體粒子速率較慢且相距較近
- Ⓓ 液體粒子速率較快且相距較近

S032403



上圖所示為一根繞著絕緣電線的鐵釘，當電流通過電線時，這根鐵釘會發生什麼事呢？

- Ⓐ 鐵釘會融化
- Ⓑ 電流會流經鐵釘
- Ⓒ 鐵釘會變成磁鐵
- Ⓓ 鐵釘不會發生任何事

S032273

岩石的風化（逐漸崩解）可能由物理作用或是化學作用所導致。請寫出一種物理作用和一種化學作用，並說明它們是如何造成岩石的風化。

S01_05

物理作用：

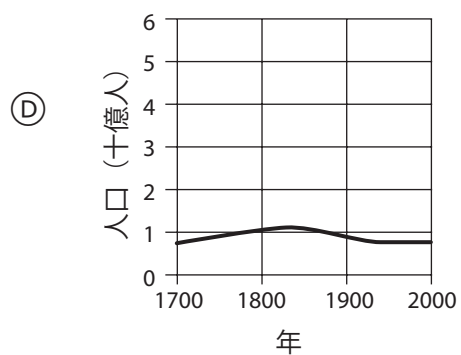
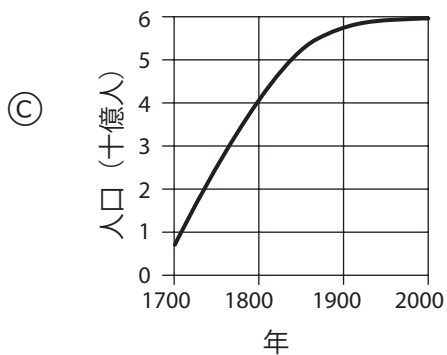
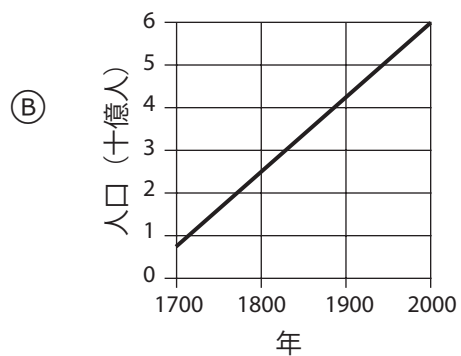
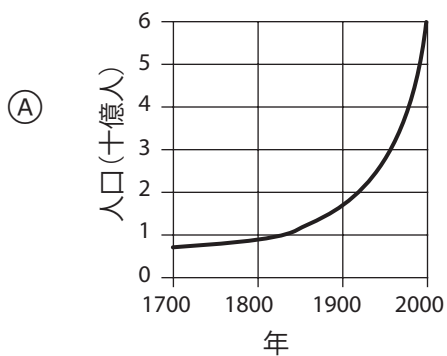
化學作用：

在一個鄉下地方，有許多的樹木，住在這裡的人們決定砍下這些樹木作為木材。

S01_06

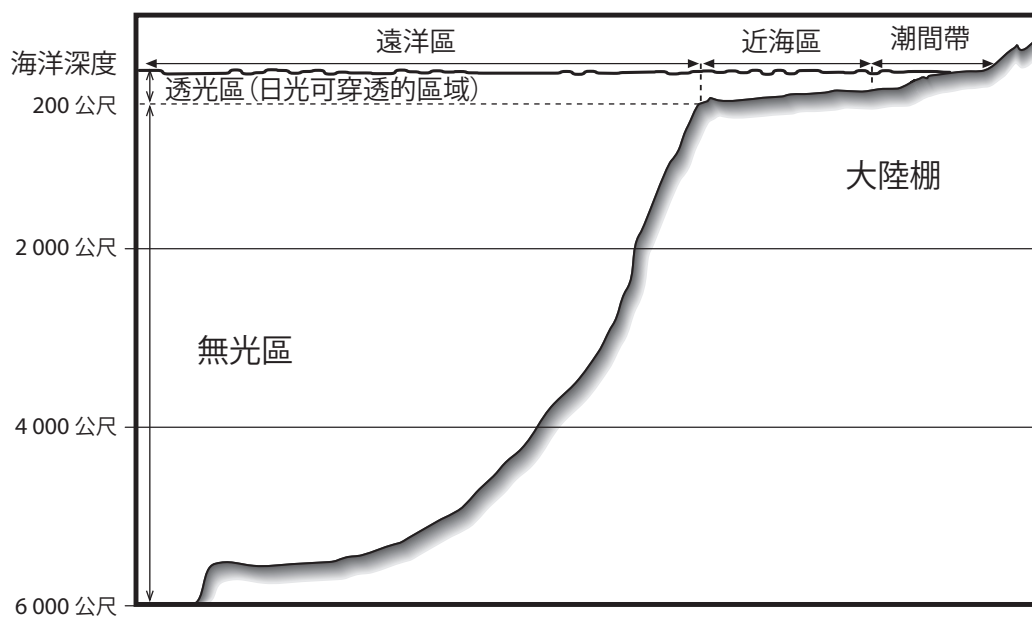
請提出一項這個決定可能產生的長期影響。

下面哪一個圖最能表示世界人口在最近 300 年內的變化？



S032620

下面是一個海洋的剖面示意圖。一些生物（植物和動物）生活在海洋的不同區域，這些生物互相依賴而生，且依賴陽光生存。



SP32693

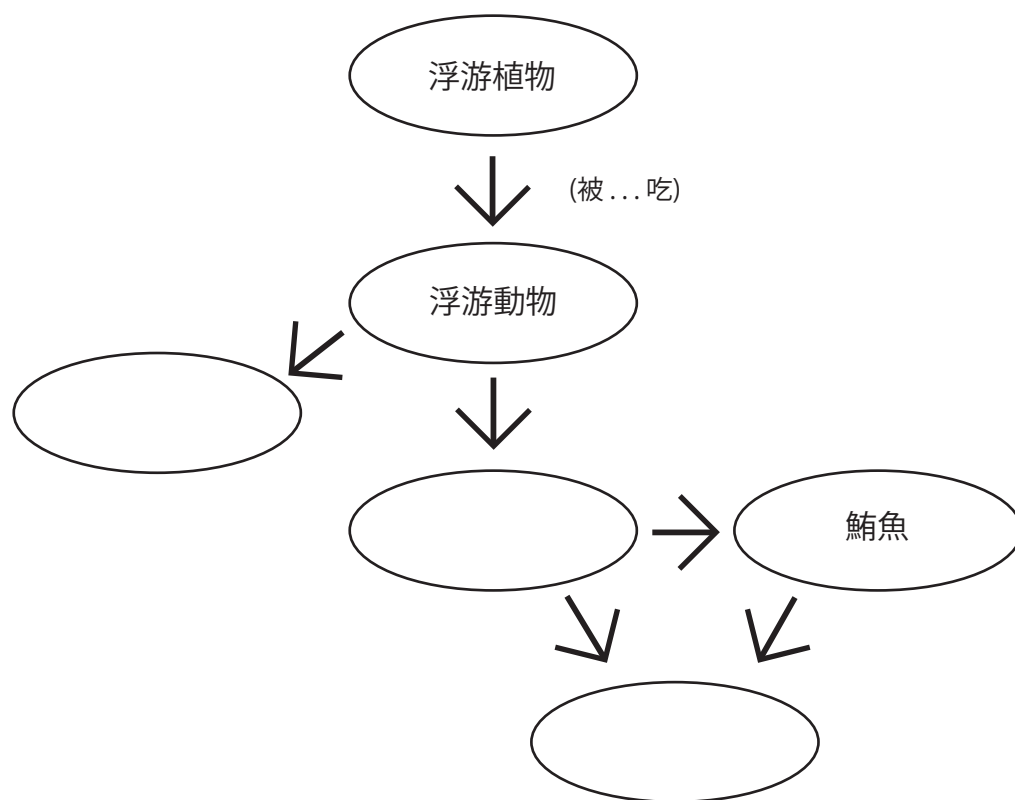
有關海洋中的生命之問題開始於下一頁 ➡

下表為在近海區生活的生物(動物和植物)。

生物種類	描述
浮游植物	可行光合作用的微小植物
浮游動物	以浮游植物為食的微小動物
鮪魚	以小型魚類為食的中型魚類
鯊魚	以浮游動物為食的小型魚類
鯨魚	以其他魚類為食的大型魚類
鯨魚	以浮游動物為食的大型哺乳動物

A. 請引用上表中的生物，完成以下的食物網，在每個圓圈中寫下一種生物的名稱。

該表中各種生物的資料可以幫助你作答。下圖中已經寫出三種生物，其中箭頭代表食物網中能量的流向。



繼續下一頁海洋中的生命之問題 ➡

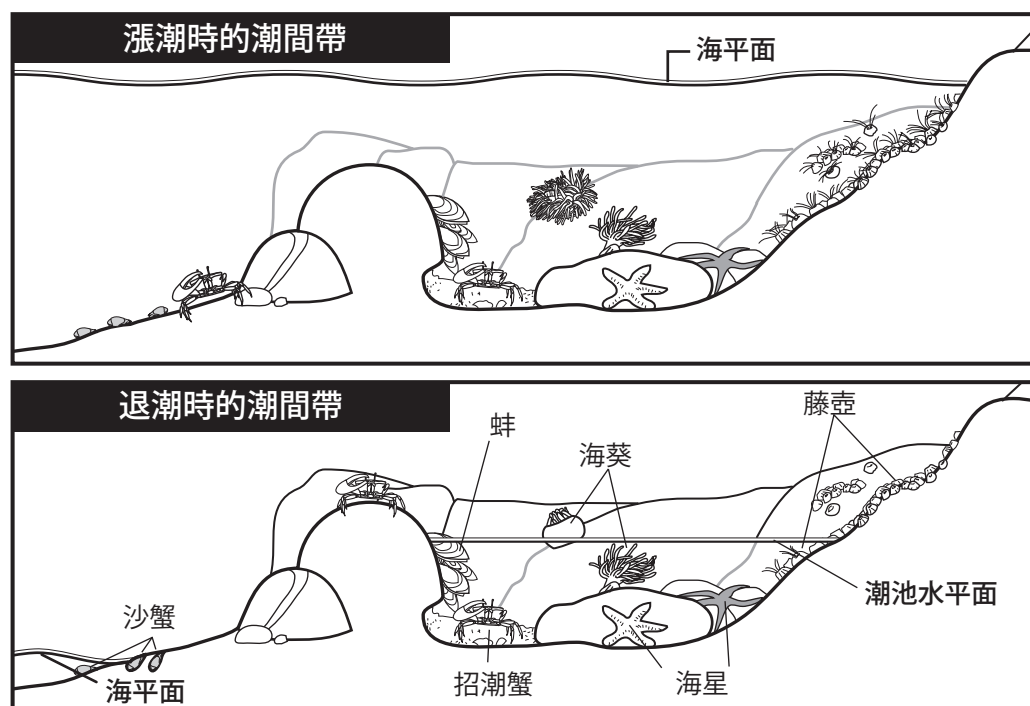
- B. 有一年由於過度捕撈，使鮭魚數量減少。這對鯊魚的族群數量最可能發生什麼變化並解釋你的答案。

S032693_2

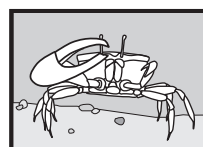
繼續下一頁海洋中的生命之問題 ➡

海洋中的生命（繼續）

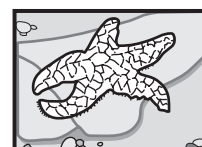
潮間帶是位於海洋沿岸的高潮線與低潮線之間的區域，下圖表示在漲潮時與退潮時潮間帶的剖面示意圖以及居住在潮間帶的生物。



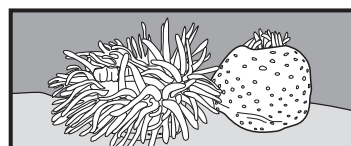
沙蟹



招潮蟹



海星



海葵



藤壺



蚌

繼續下一頁海洋中的生命之問題 ➡



居住在潮間帶的生物具有特殊的適應方式，使其能在潮汐的作用下生存。

從潮間帶選出一種生物，寫出一個幫助牠在退潮時維持生存的身體特徵或行為。請解釋你所寫的特徵如何幫助這個生物在低潮時生存。

生 物 名 稱： _____

特 徵 或 行 爲： _____

解釋：

繼續下一頁海洋中的生命之問題 ➡

海洋中的生命（繼續）

在深海中所發現的生物種類（植物和動物）比在靠近海面處的生物種類少。居住在海底的生物必須適應海底的環境條件。

請寫出兩個使得大多數動物與植物難以在海底生存的條件。

1.

2.

S01_10

S032697

海洋中的生命部分結束



在生命體中，大而複雜的分子被分解成為小而簡單的分子。
這個過程叫做什麼作用？

- Ⓐ 排泄
- Ⓑ 吸收
- Ⓒ 消化
- Ⓓ 循環

S042009

小凱得了流行性感冒。他和兩位朋友玩遊戲。其中一位朋友感冒了，但是另一位朋友沒有感冒。

小凱這一位朋友**沒有**得流行性感冒，可能的原因是什麼？

S042313

哪一種食物所含的蛋白質百分比最高？

- Ⓐ 米
- Ⓑ 棗子
- Ⓒ 胡蘿蔔
- Ⓓ 雞肉

S042059

居住在氣候炎熱地區的人與居住在氣候寒冷地區的人, 平均體溫比較起來如何?

(勾選一項。)

- ☐ 在炎熱氣候較高
- ☐ 在炎熱氣候較低
- ☐ 在兩種氣候相同

請解釋你的答案。

S042011

下列哪一個在授精後立即形成?

- Ⓐ 卵
- Ⓑ 精子
- Ⓒ 受精卵
- Ⓓ 胚胎

S042028

某種動物有鱗並且只用肺進行氣體交換。

這種動物最可能被分為哪一類？

- Ⓐ 魚類
- Ⓑ 爬蟲類
- Ⓒ 哺乳類
- Ⓓ 兩生類

S042001

如蘋果這類物體的顏色跟哪個光波的顏色一樣？

- Ⓐ 在該物體中行進的光波
- Ⓑ 該物體吸收的光波
- Ⓒ 該物體反射的光波
- Ⓓ 從該物體旁經過折射的光波

S042276

大振幅的聲波跟小振幅的聲波相比，下列的敘述哪一項正確？

- Ⓐ 大振幅的聲波能量較小，聽起來較小聲。
- Ⓑ 大振幅的聲波能量較大，聽起來較大聲。
- Ⓒ 兩者能量相同，大振幅的聲波聽起來較小聲。
- Ⓓ 兩者能量相同，大振幅的聲波聽起來較大聲。

S042279

孟哲從腳踏車上摔下來，弄灑了他所帶著的一袋鹽。他從地上把鹽和砂及樹葉一起收集起來，把混合物放進塑膠袋裏。



在下表中，描述孟哲可從鹽、砂、以及樹葉的混合物中分離出鹽的步驟，並請說明使用每個步驟的一個理由。
第一步驟已幫你完成。

步驟	步驟的描述	使用該步驟的原因
1.	把混合物通過篩網。	這樣可以去除樹葉。
2.		
3.		
4.		

使用磅秤來測量物質 A 和 B 的質量，如圖 1 所示。把物質 B 放入大燒杯中形成了物質 C，空燒杯再放回磅秤上，如圖 2 所示。

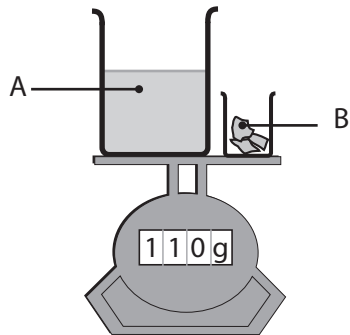


圖 1

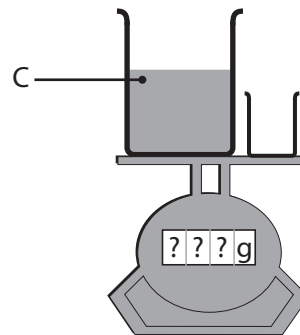


圖 2

圖 1 中的讀數顯示質量為 110 克。

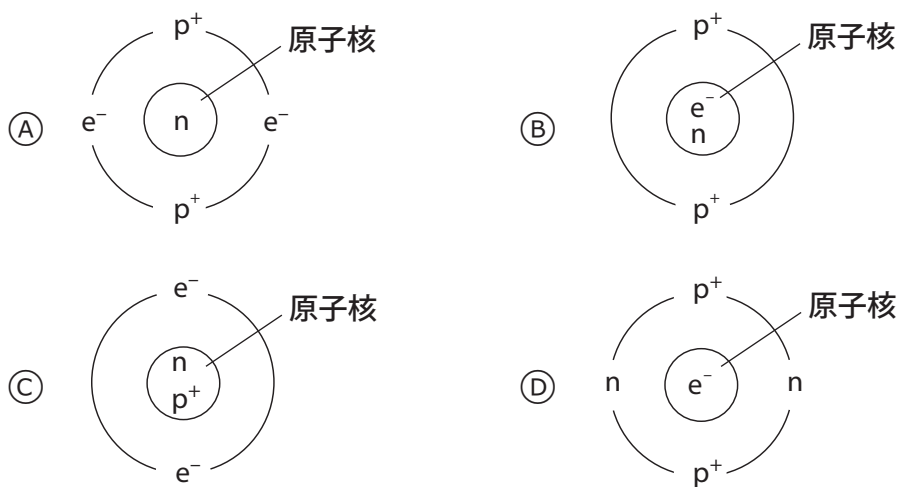
圖 2 中所示的讀數為何？

(勾選一項。)

- ☐ 大於 110 克
- ☐ 110 公克
- ☐ 小於 110 克

請解釋你的答案。

下列哪一種模型顯示在一個原子中，質子 (p^+)、電子 (e^-) 以及中子 (n) 正確的位置？



S042071

梵霖拿了一杯牛奶並用藍色石蕊試紙測試它。石蕊試紙仍呈藍色。兩天後，梵霖再用藍色石蕊試紙測試這一杯牛奶，藍色石蕊試紙變為粉紅色。



這杯牛奶發生了哪一種變化？

(勾選一項。)

☐ 化學變化

☐ 物理變化

請解釋你的答案。

S042101

請解釋土壤是如何形成的。

S02_13

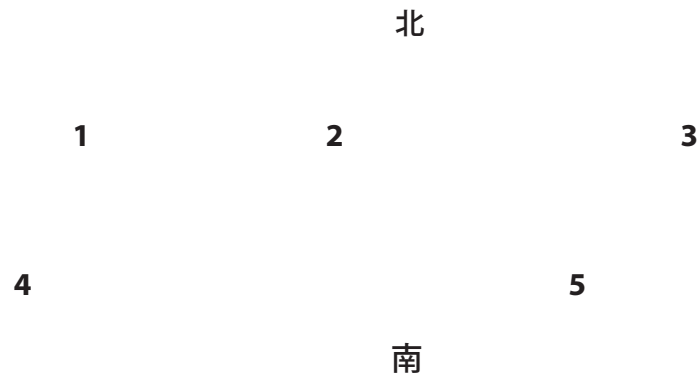


鄉鎮 1、2、3 的位置在鄉鎮 4 和 5 的北方，他們都建在平原上。

S02_14

星期一，鄉鎮 1 晴天而鄉鎮 2 和 3 卻在下雨。

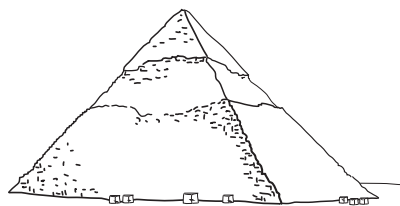
風是從所有鄉鎮的北方吹來的。



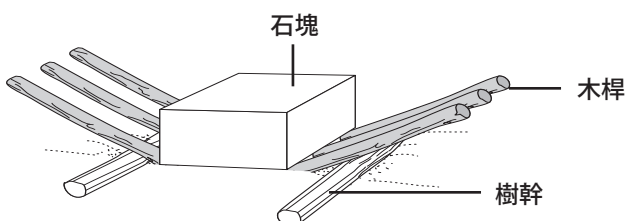
如果風繼續從北方吹來，星期二，鄉鎮 4 和 5 最可能是什麼樣的天氣？

- Ⓐ 兩個都下雨
- Ⓑ 兩個都晴天
- Ⓒ 4 是晴天，5 是雨天
- Ⓓ 4 是雨天，5 是晴天

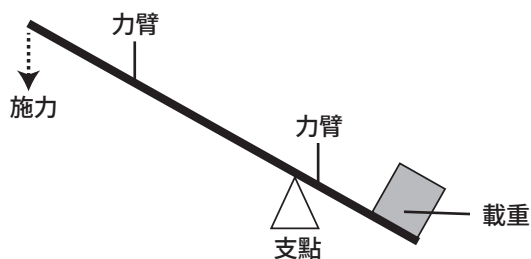
阿德和小瓊正學習到發現於埃及的古夫金字塔。



他們想知道古埃及人如何抬起石塊來建造金字塔。他們在網路上作了點研究並發現下圖。



阿德不確定他是否瞭解這張圖，所以小瓊畫了張圖來幫助他瞭解石塊是如何被抬起來的。她的圖如下所示。

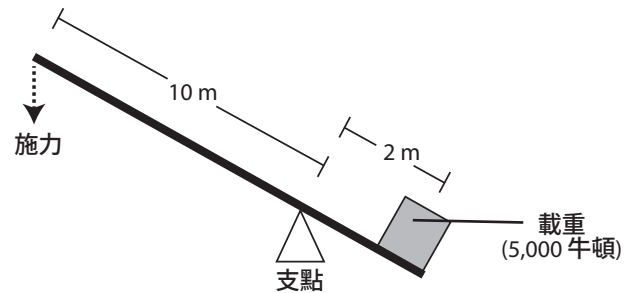


A. 請把埃及人槓桿的各個部份對應到小瓊畫的槓桿圖上。

小瓊的圖	埃及人的槓桿
施力	工人向下的拉力
載重	
支點	
力臂	

- B. 阿德和小瓊讀到六個人可一起抬起重30,000牛頓的石塊。每個人需要能夠抬起這個重量的六分之一（5,000牛頓）。他們決定算出每個人必須對木桿所施的力。

阿傑把每個力臂的長度加到小瓊的圖上，如下所示。



他在教科書上查到：

$$\frac{\text{載重}}{\text{施力}} = \frac{\text{施力到支點的長度}}{\text{載重到支點的長度}}$$

每個人必須施多少力才能抬起石塊？

_____ 牛頓

S042244_2

請解釋為什麼回收家用物質如塑膠類、瓶罐和紙類是很重要的。

S042153

下列哪一種氣體會使金屬罐生鏽？

- Ⓐ 氫氣
- Ⓑ 氧氣
- Ⓒ 氮氣
- Ⓓ 氨氣

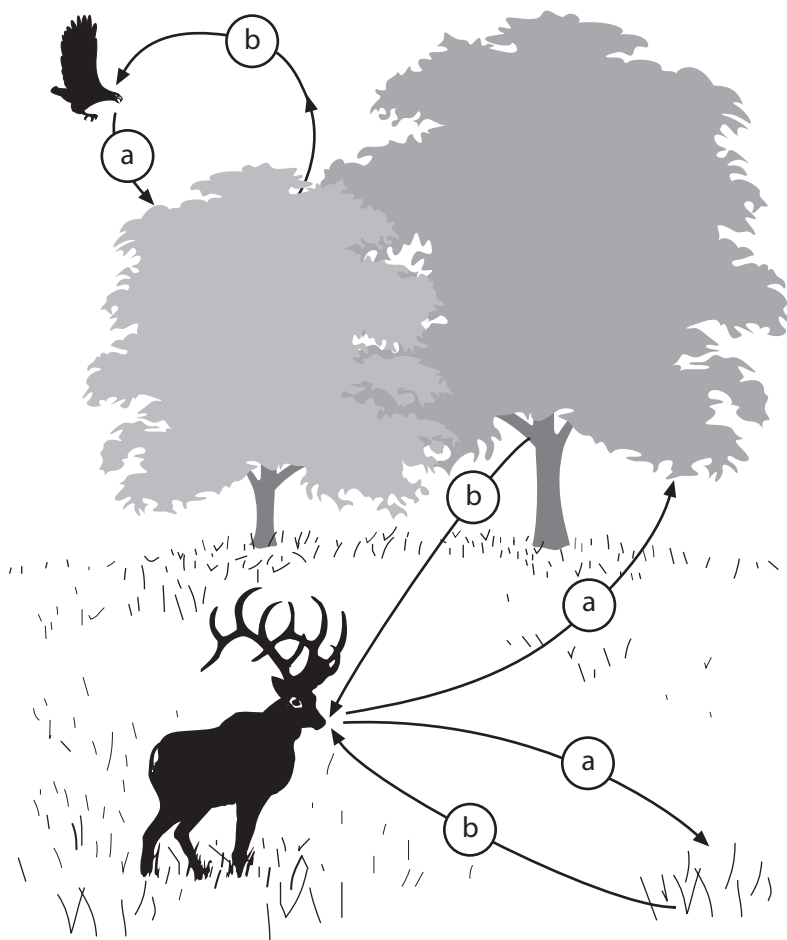
S022183

小莉把一對電極放入一個盛有溶液的燒杯中，並把電極連接到電池上。小莉報告中部分寫到：“在其中一個電極，有氣泡釋出”。此種敘述是屬於下列何者？

- Ⓐ 一個觀察
- Ⓑ 一種預測
- Ⓒ 一個結論
- Ⓓ 一個理論
- Ⓔ 一種假設

S022276

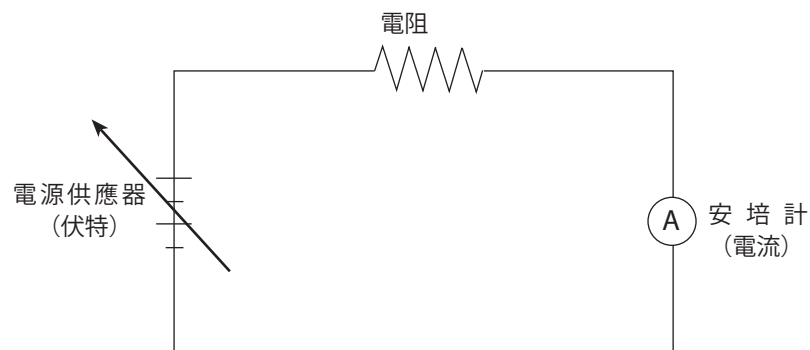
下圖代表生物間相互倚賴的一個例子，白天生物體會依圖上的箭頭方向所表示用掉或放掉 (a) 或 (b)。



請試由下列選出 (a) 與 (b) 的正確答案。

- Ⓐ (a) 是二氧化碳和 (b) 是氮
- Ⓑ (a) 是氧和 (b) 是二氧化碳
- Ⓒ (a) 是二氧化碳和 (b) 是水蒸氣
- Ⓓ (a) 是二氧化碳和 (b) 是氧

一些學生們利用安培計 (A) 測量不同電壓下線路的電流。



下表顯示部分的測量結果，請完成此表格。

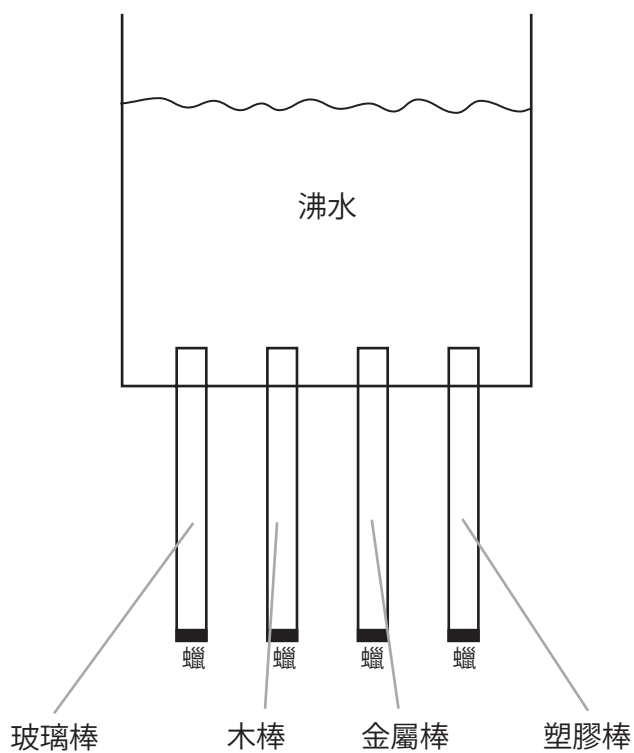
電壓 (伏特)	電流 (毫安培)
2	15
4	30
	60

S022022

人們在地球上的山谷裡高聲喊叫，可聽到四周圍山壁所反射的回聲。在月球上同樣的山谷裡喊叫，卻聽不到回聲，這是因為.....

- Ⓐ 月球上的重力太小
- Ⓑ 月球上的溫度太低
- Ⓒ 月球上沒有空氣傳播聲音
- Ⓓ 月球上山壁不能反射聲音

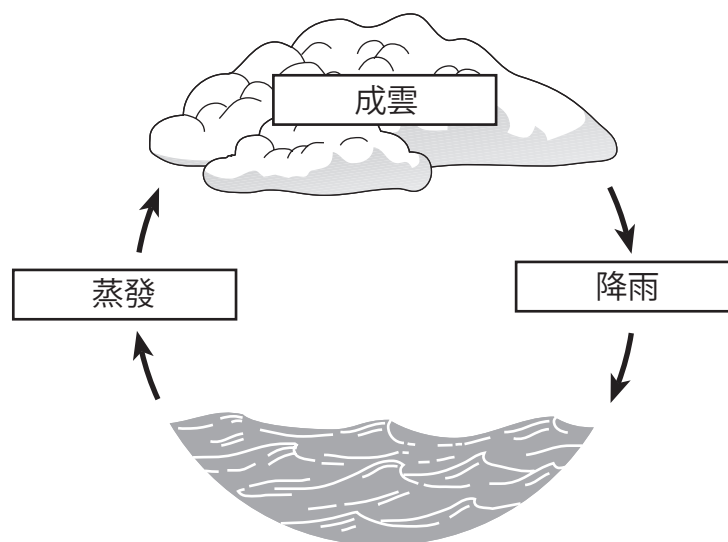
S022019



上圖表示四支大小完全相同但材質不同的細棒，緊緊地塞住一容器底部。每支細棒的下端均塗有等量的蠟，然後容器內倒入沸水。哪一支細棒上的蠟將最先融化呢？

- Ⓐ 玻璃棒
- Ⓑ 木棒
- Ⓒ 金屬棒
- Ⓓ 塑膠棒

下圖表示地球上的水循環。



水循環的能量來源是什麼？

- Ⓐ 月亮
- Ⓑ 太陽
- Ⓒ 潮汐
- Ⓓ 風

S022294

下列何種動物，已經在地球上存活最久？

- Ⓐ 人類
- Ⓑ 鳥類
- Ⓒ 魚類
- Ⓓ 爬蟲類

S022106

中央市的發電廠使用燃煤來產生電力，供全市利用。當煤在燃燒的時候，其中的硫便會和氧氣反應，生成二氧化硫，然後以氣體的形式散逸出去。

這個過程是怎麼造成酸雨的呢？

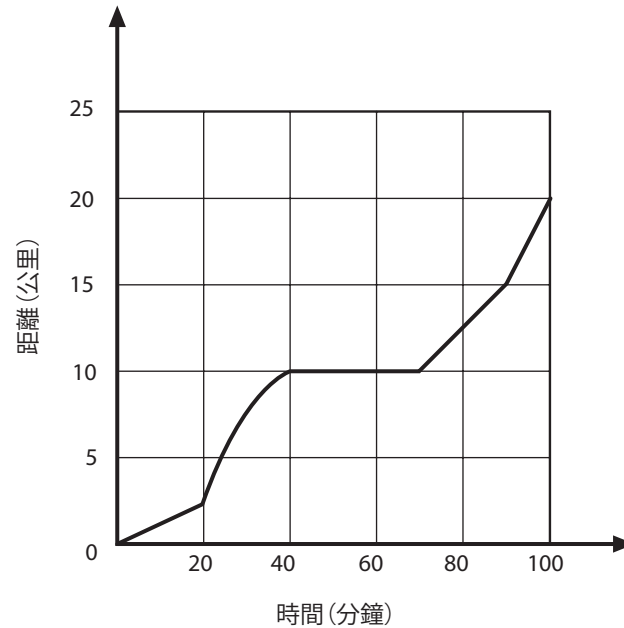
S022244

生物體的架構，由最不複雜到最複雜的順序為何？

- Ⓐ 細胞、組織、器官、個體
- Ⓑ 細胞、器官、組織、個體
- Ⓒ 組織、細胞、器官、個體
- Ⓓ 組織、器官、細胞、個體

S022150

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。



瑪莉花多久時間修補車胎呢？

- Ⓐ 20 分鐘
- Ⓑ 30 分鐘
- Ⓒ 40 分鐘
- Ⓓ 70 分鐘

當你運動時心跳會加快，請解釋其原因。

S03_12

小明將一壺水放在火爐上加熱，他在水開始沸騰時就測量水的溫度，結果溫度計顯示為 100°C 。他將火開大，讓水繼續沸騰 5 分鐘，然後再次測量水溫。

S03_13

請問溫度計所顯示的溫度會高於、低於或等於 100°C ?

答：_____

解釋你的答案。

小華取兩個盛有等量水的相同小容器，其中一個溶入一湯匙食鹽，再將這兩個容器放進冷凍庫裡。小華每隔 5 分鐘就去觀察一次，一直到其中之一的溶液結凍為止。

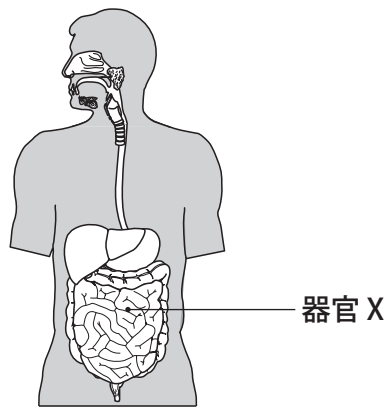
試問小華能從他的實驗中發現什麼？

S022268

能傳達訊息的細胞是

- Ⓐ 皮膚細胞
- Ⓑ 神經細胞
- Ⓒ 血球細胞
- Ⓓ 腎臟細胞

S042013



器官 X 是什麼?

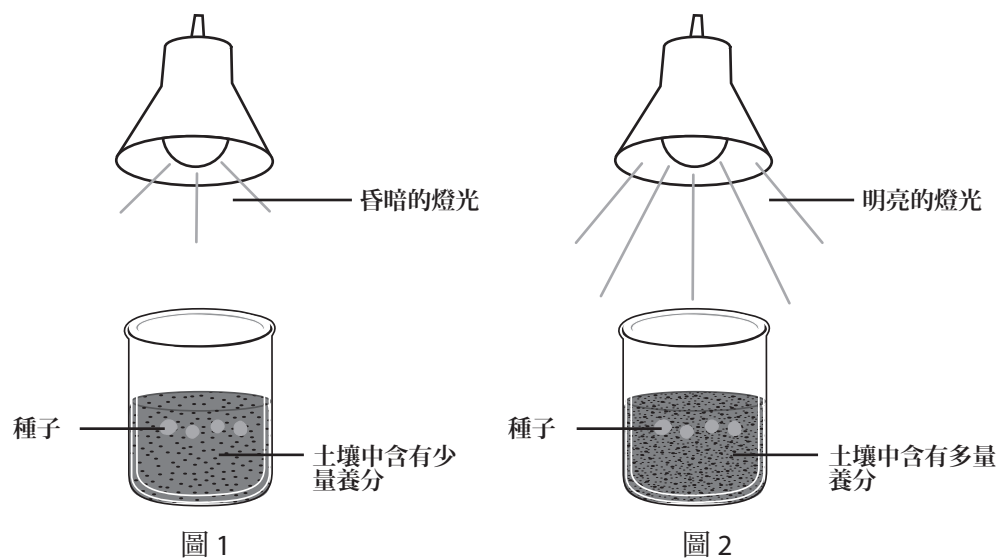
- Ⓐ 肝
- Ⓑ 胃
- Ⓒ 小腸
- Ⓓ 大腸

S042006

小德有一包基因相同的豆子種子。

這些豆子可以長成高莖的豆類植物。

他把四顆豆子種子種在圖1的容器中，把另外四顆種子種在圖2的容器中。
他每天為這些種子澆水。



關於這些豆類植物長的高度，可以提出什麼預測？

請解釋你的答案。

一個可以容納50萬人的城市,現在預計十年後人口數將增加到100萬人。
請描述這個城市因為人口數增加,所需要面對的兩個環境問題。

1.

2.

S042052

下列哪一項疾病是由病毒引起?

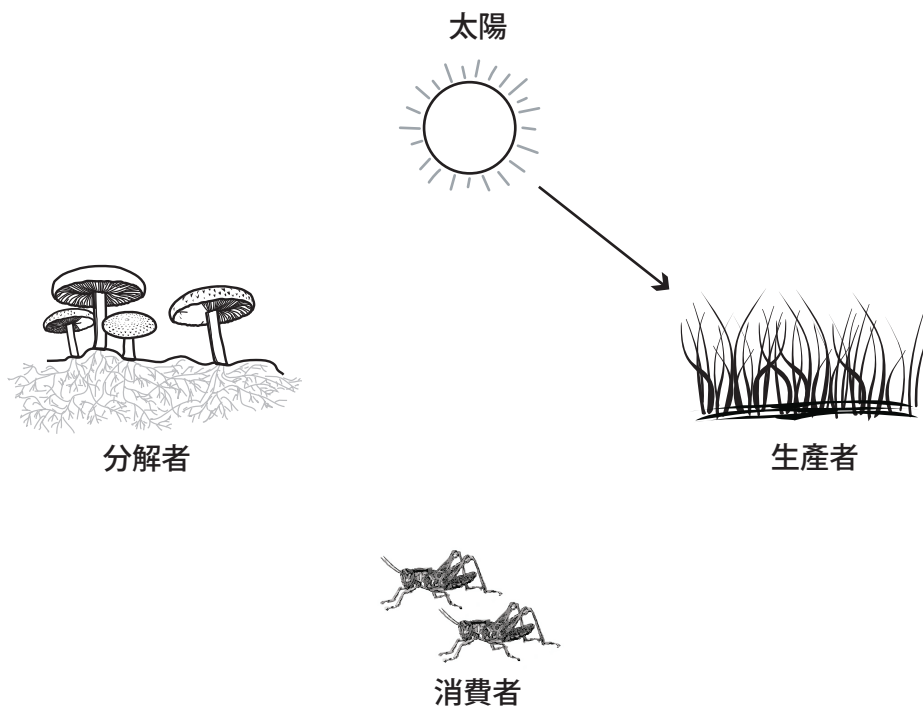
- Ⓐ 潰瘍
- Ⓑ 瘧疾
- Ⓒ 肺結核
- Ⓓ 流行性感冒

S042054

請在下圖中，用箭頭標示出能量在生產者、消費者，和分解者之間流動的方向。



圖中的箭頭表示能量由太陽流到生產者的流動方向。



S042043

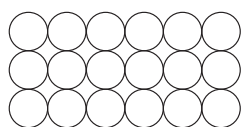
爲了提供電力，住家用電是採並聯電路而非串聯電路。

家中使用並聯電路有什麼優點？

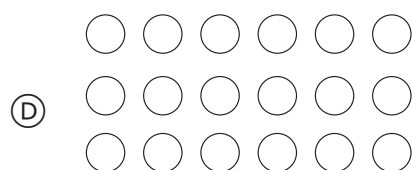
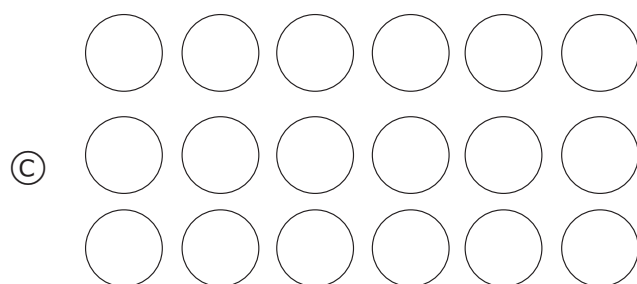
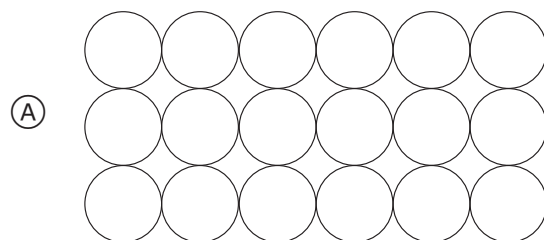


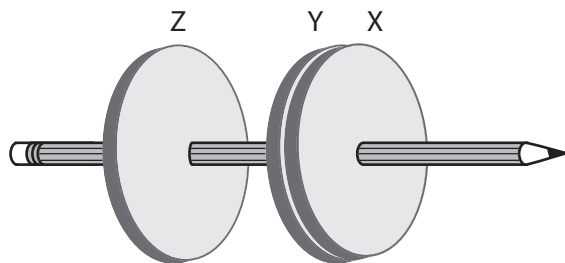
S042196

金屬加熱前，其中粒子的排列情形如圖所示：



哪一張圖表示金屬加熱後其粒子的排列情形？





此圖顯示三個磁鐵穿過一枝鉛筆而放在一起後發生的現象。
磁鐵 X 和 Y 會移動到它們碰在為止，但磁鐵 Y 和 Z 則保持分開。

1. 請解釋為什麼磁鐵 X 和 Y 碰在一起。

2. 請解釋為什麼磁鐵 Y 和 Z 保持分開。

S042292

哪一種物質是燃燒所必需的？

- Ⓐ 臭氧
- Ⓑ 氧氣
- Ⓒ 氫氣
- Ⓓ 二氧化碳

S042109

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 1 表示每組如何測得汽水罐的質量。

表1:質量

組別	方法	質量 (g)
A	我們用天平來測汽水罐的質量。	389.30
B	我們把汽水罐打開把汽水倒光。我們用天秤來測汽水罐的質量。	13.85
C	我們把汽水罐打開把汽水倒光。我們用天秤來測汽水罐的質量。	13.85
D	我們用天平來測汽水罐的質量。	389.30

A. 請解釋為什麼 A、D 兩組與 B、C 兩組得到不同的質量結果。

B. 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。罐子沈下去。 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，罐子沈到燒杯底，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。

C. 下表表示各組所測得的質量、體積以及密度結果。

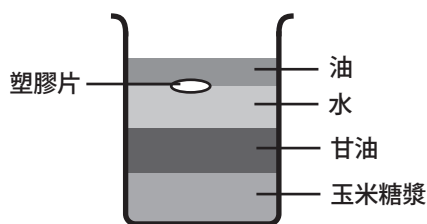
組別	A	B	C	D
質量 (g)	389.30	13.85	13.85	389.30
體積 (mL)	376.00	376.00	5.00	371.00
密度 (g/mL)	1.04	0.04	2.77	1.05

根據所用的方法，哪一組能得到製造汽水罐的金屬之密度？

- Ⓐ A 組
- Ⓑ B 組
- Ⓒ C 組
- Ⓓ D 組

S042232_3

小華把玉米糖漿倒入空的燒杯底部。他小心地加入一層甘油、一層水和一層油，如圖所示。然後他把一塊塑膠片丟入燒杯裡。



哪一個敘述是正確的？

- Ⓐ 油的密度比玉米糖漿大。
- Ⓑ 塑膠片的密度比油小。
- Ⓒ 甘油的密度比油大。
- Ⓓ 玉米糖漿的密度比水小。

S042294

請描述地下水會被污染的一種成因？

S04_13

S042149

請解釋為什麼種樹可以減少土壤的侵蝕。

S04_14

S042155

何者屬於非再生資源？

S04_15

- Ⓐ 石油
- Ⓑ 砂
- Ⓒ 木材
- Ⓓ 氧氣

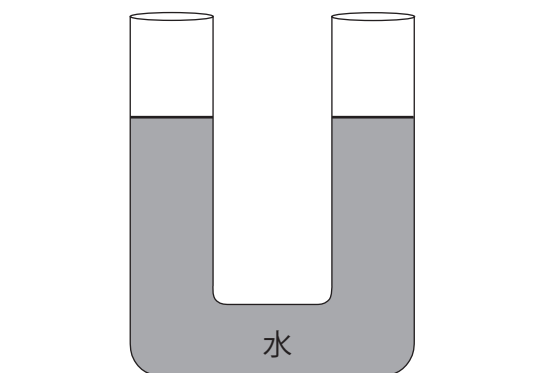
S042150

地球上所謂的“一年”，是指什麼意思？

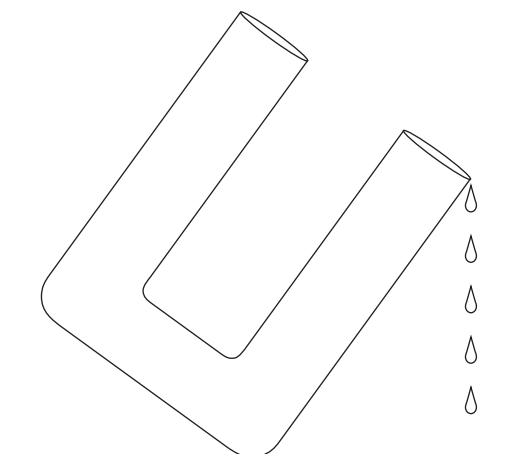
- Ⓐ 地球繞地軸自轉一圈
- Ⓑ 月亮繞地球公轉一圈
- Ⓒ 太陽繞地球公轉一圈
- Ⓓ 地球繞太陽公轉一圈

S022290

有一個開口 U 形管，裡面裝有水，如下圖所示。



若將 U 形管傾斜，使水恰可滴出管外，在下圖中畫線顯示此時的水面在何處呢？



S022292

加色的酒精被用在一些玻璃溫度計中，當其在不同溫度時，玻璃溫度計內加色的酒精柱會上升或下降。下列對玻璃溫度計內有色酒精柱高度改變的解釋，何者最佳？

- Ⓐ 當加熱時，玻璃收縮
- Ⓑ 當加熱時，酒精收縮
- Ⓒ 當加熱時，玻璃要比酒精容易膨脹
- Ⓓ 當加熱時，酒精要比玻璃容易膨脹

糖是由許多分子組成，當糖溶在水裡時，這些分子變得如何？

- Ⓐ 它們不再存在
- Ⓑ 它們存在於溶液中
- Ⓒ 它們蒸發掉了
- Ⓓ 它們和水結合，形成一些新的元素

下列哪一項**不是**化學變化？

- Ⓐ 冰融化
- Ⓑ 銀腐蝕
- Ⓒ 火柴燃燒
- Ⓓ 青菜腐爛

光線由太陽射到地球需時 8 分鐘，但是光線以相同的速率進行，從月亮到地球只需 1.5 秒，為什麼？



S022078

葉綠素在植物體內的主要功能是什麼？

- Ⓐ 吸收日光能
- Ⓑ 分解二氧化碳
- Ⓒ 使植物的葉子對昆蟲有毒
- Ⓓ 保護植物免於疾病

S022126

將裝有 300 公克水的製冰盒放入冰箱中的冷凍庫製冰。
水冷凍成冰的質量有多少？

(請擇一打勾)

- ☐ 大於 300 公克
☐ 300 公克
☐ 小於 300 公克

請解釋你的答案。

S022281

下列何種特徵是哺乳動物所獨有的？

- Ⓐ 辨識顏色的眼睛
Ⓑ 分泌乳汁的腺體
Ⓒ 吸收氧氣的皮膚
Ⓓ 有鱗片保護的身體

S032385

下列哪一項是判定兩個人是否有血緣關係的最佳方式？

- Ⓐ 比較他們的血型
- Ⓑ 比較他們的筆跡
- Ⓒ 比較他們的基因
- Ⓓ 比較他們的指紋

S032035

爲了發電和灌溉的目的，政府決定在城鎮附近的一條河川興建水庫。寫出這個水庫可能對野生物（動物或植物）的一種影響。

S032519

下表顯示不同的材料已被分類成兩組。

第 1 組	第 2 組
空氣	銅
冰	銅
木材	金

下列何者可用來將這些材料分類成第 1 組和第 2 組？

- Ⓐ 在水中的溶解度
- Ⓑ 可壓縮性
- Ⓒ 物理狀態
- Ⓓ 導電性

S032683

膽囊可以儲存膽汁，而膽汁可以幫助消化脂肪。請問當一個人的膽囊移除之後，應該要避免攝食哪一類的食物？

- Ⓐ 水果
- Ⓑ 五穀
- Ⓒ 乳酪
- Ⓓ 蔬菜

S032258

科技在很多方面可以應用來保護環境，例如新的塑膠原料所製成的垃圾袋，掩埋時較容易分解。

請描述科技如何解決下列的環境問題。

A. 海洋上的漏油污染：

B. 大氣中二氧化碳增加所引發的全球暖化現象：

心臟、靜脈、動脈、微血管構成了那一個系統？

- Ⓐ 生殖
- Ⓑ 肌肉
- Ⓒ 排泄
- Ⓓ 循環

S032606

請寫出一個植物細胞有而動物細胞沒有的構造。

S032015

綠色植物行光合作用時會產生養分和氧氣。葉綠素是行光合作用的因素之一。

請寫出行光合作用所需的其他兩項因素。

1.

2.

S032310

下表中列出三種純物質(X,Y 和 Z)的一些性質。其中一種是鐵，一種是水，而另一種是氧。



純物質	熔點／凝固點 (°C)	沸點(°C)	電的良導體
X	-218	-183	不是
Y	1 535	2 750	是
Z	0	100	不是

在下列空格中寫上**鐵**、**水**或**氧**等語詞，以表明 X、Y、Z分別是什麼物質。

物質 X 是： _____

物質 Y 是： _____

物質 Z 是： _____

S032680

下列那一個是酸性溶液的例子？

- Ⓐ 漂白劑
- Ⓑ 醋
- Ⓒ 糖水
- Ⓓ 鹽水

S032672

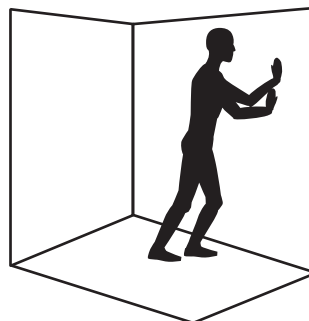
當一個物體沿著作用力的方向移動時，這就是對它作了功。有一個人作了下圖所示的各種活動，在那一個圖中的人有作功？

Ⓐ



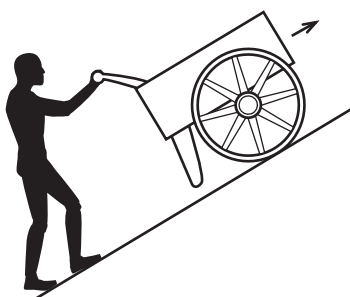
舉著一個重物

Ⓑ



推一面牆

Ⓒ



推車上斜坡

Ⓓ



閱讀一本書

S032392

下面那一項物質的性質在熱膨脹時是不變的？

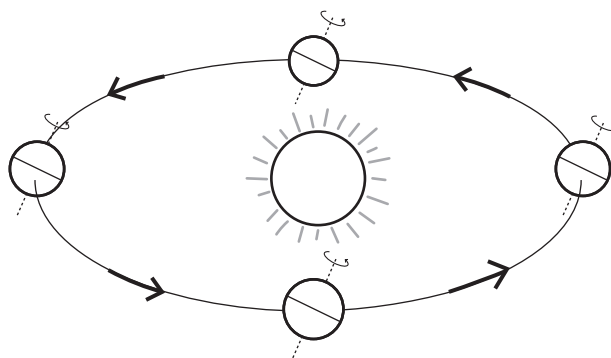
- Ⓐ 質量
- Ⓑ 體積
- Ⓒ 形狀
- Ⓓ 粒子間的距離

S032425

當你撥動吉他的弦時會聽到聲音。如果用更大的力撥同一根弦，會發生什麼事呢？

- Ⓐ 聲音大小相同，音調會高一點
- Ⓑ 音調相同，聲音會變大
- Ⓒ 音調會變高，聲音也會變大
- Ⓓ 音調與聲音都和原來一樣

S032257



上圖為地球繞太陽公轉軌道和地球自轉軸傾斜的示意圖，下列哪一個現象是因為地球自轉軸傾斜所造成的？

- Ⓐ 季節
- Ⓑ 晝夜
- Ⓒ 年
- Ⓓ 時區

S032663

地球上淡水最多的地方主要在何處？

- Ⓐ 海洋
- Ⓑ 河流
- Ⓒ 湖泊
- Ⓓ 極區冰帽

S032660

一塊土地平均分成十等分，在各份土地上種植稻米，並施加不同份量的肥料。以下的表格呈現每份土地的稻米收成量與所施加的肥料量。

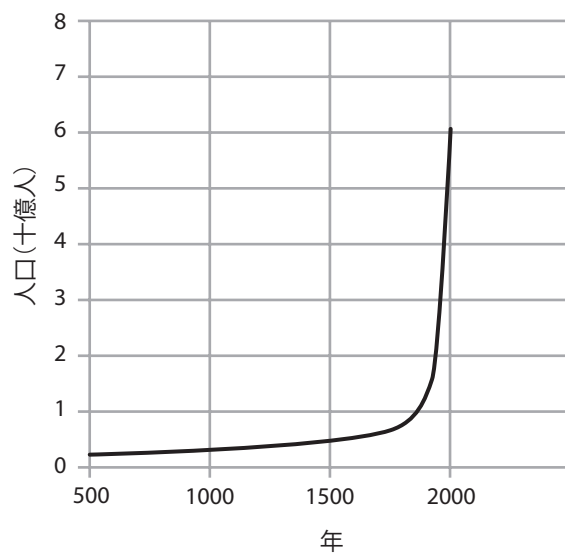


	土地編號									
	1	2	3	4	5	6	7	8	9	10
所施加的肥料量 (每塊土地的氮 肥量)	0	30	50	60	70	80	100	120	140	160
稻米收成量 (每 塊土地稻米收成 的公斤數)	7.1	8.3	14.2	25.4	26.2	26.2	26.2	26.1	17.6	14.4

根據這個表的數據，請解釋施加的肥料量對稻米收成的影響。

S032555

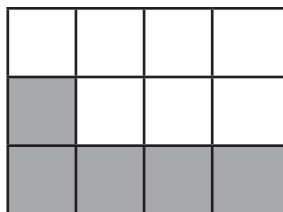
下圖顯示過去 1500 年來世界人口的成長。



請提出一個理由來解釋「為什麼在西元 1800 年到西元 2000 年之間人口快速增加？」

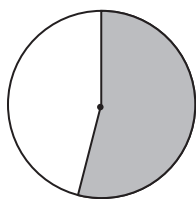
附錄三

TIMSS 2007 八年級數學可公布試題（中文版）

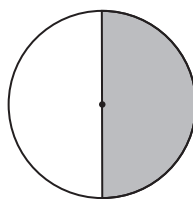


下列哪一個圓中，陰影部分面積占整個圓的比例，和上面長方形中陰影部份面積所占比例大約相同？

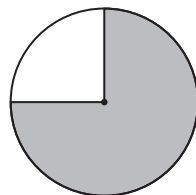
Ⓐ



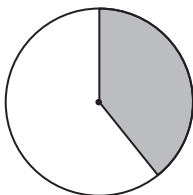
Ⓑ



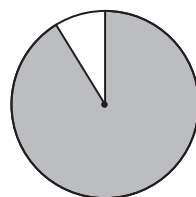
Ⓒ



Ⓓ



Ⓔ



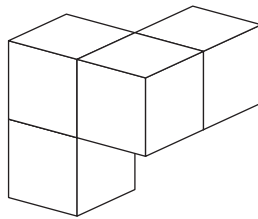
園丁爲了在草坪上播種，將 4.45 公斤的牧草種子和 2.735 公斤的苜蓿種子混合。他現在共有多少公斤的混合種子？

M01_02

答：_____

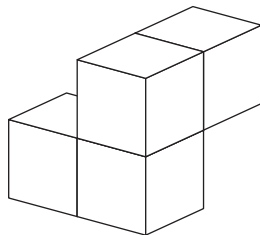
這個東西將被轉動到另一個方向。

M01_03

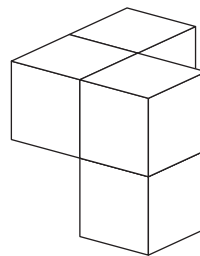


下列哪一個圖形表示它轉動後的樣子？

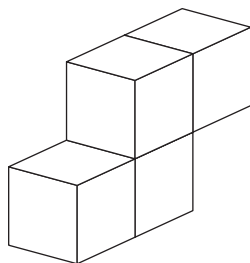
Ⓐ



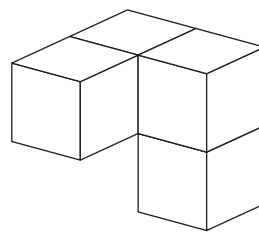
Ⓑ



Ⓒ



Ⓓ



$\frac{x}{3} > 8$ 可以化簡成下列哪一個不等式？

- Ⓐ $x < 5$
- Ⓑ $x < 24$
- Ⓒ $x > \frac{8}{3}$
- Ⓓ $x > 5$
- Ⓔ $x > 24$

M022050

有一個面積為 100 平方公尺的正方形，它的周長是多少？



答：_____

M022055

有一年某一家公司賣出了 1426 公噸的肥料，第二年這家公司少賣了 15%。下列哪一個數字最接近這家公司第二年所賣出肥料的公噸數？

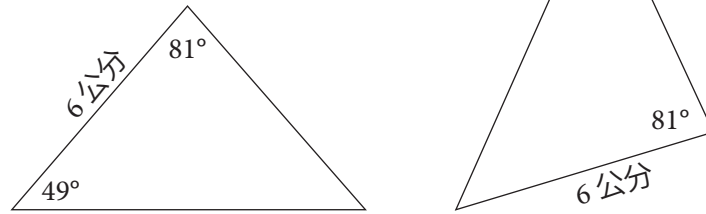
- Ⓐ 200
- Ⓑ 300
- Ⓒ 1200
- Ⓓ 1600
- Ⓔ 1700

M022057

一個碗裡裝了 36 個同樣大小的各色珠子，有藍色、綠色、紅色、也有黃色的。不看珠子，從碗中任意選取一個，拿到藍色珠子的機率是 $\frac{4}{9}$ 。那麼碗裡的藍色珠子有多少個？

- Ⓐ 4
- Ⓑ 8
- Ⓒ 16
- Ⓓ 18
- Ⓔ 20

M022257



上面兩個三角形是完全相等的全等圖形，其中有一些角和邊的大小已經知道了（如圖所示）。那麼 x 的值是多少？

- Ⓐ 49
- Ⓑ 50
- Ⓒ 60
- Ⓓ 70
- Ⓔ 81

$$\frac{2}{5} + \frac{5}{4} + \frac{9}{8} =$$

Ⓐ $\frac{16}{17}$

Ⓑ $\frac{41}{40}$

Ⓒ $\frac{81}{40}$

Ⓓ $\frac{111}{40}$

M022066

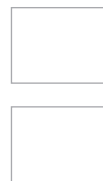
君君做了一個表，用來記錄燒杯裡的水從 95°C 冷卻到 70°C 所花費的時間。他以 5°C 為間隔來測量水冷卻所需的時間。

下降溫度區間讀數	冷卻所需的時間
95°C – 90°C	2 分 10 秒
90°C – 85°C	3 分 19 秒
85°C – 80°C	4 分 48 秒
80°C – 75°C	6 分 55 秒
75°C – 70°C	9 分 43 秒

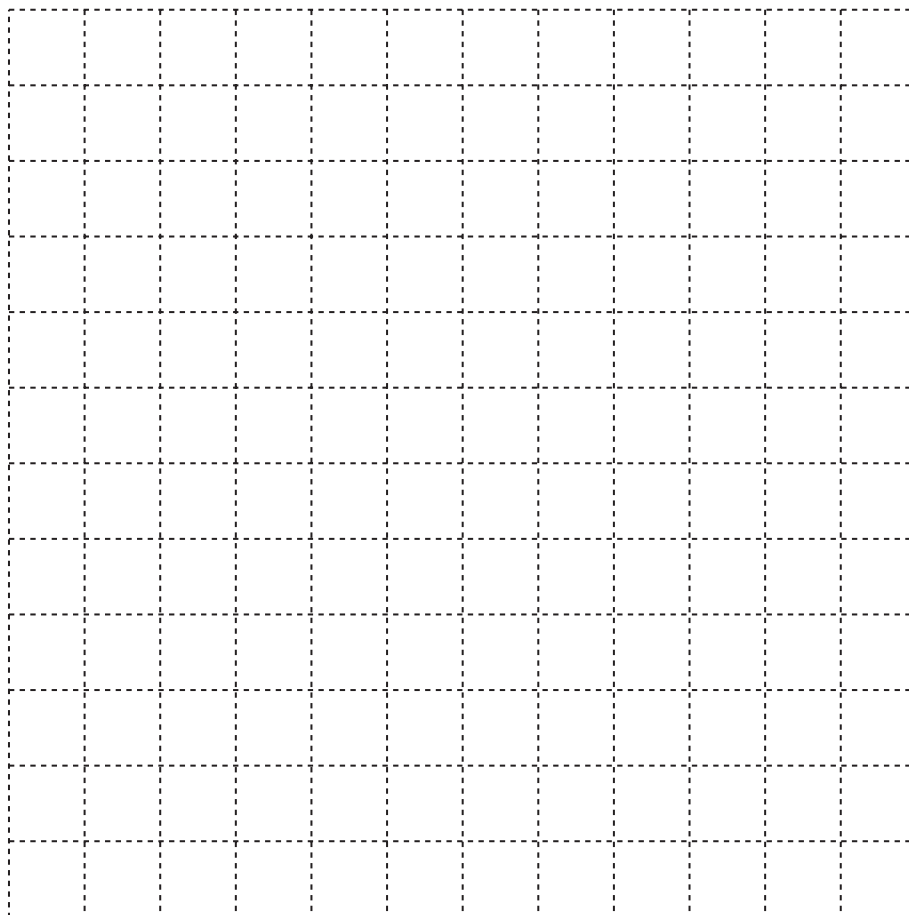
請估計一下：全部的水從 95°C 冷卻到 70°C 所需花費的時間最接近多少分鐘？並請解釋你是如何得到這個估計值的。

估計：_____

解釋：

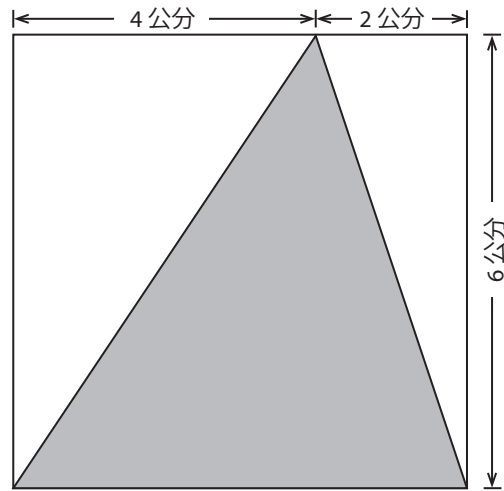


- A. 利用下面的方格紙畫出一個長方形，使它的長是上圖長方形長的四分之三，而它的寬則是上圖長方形寬的二又二分之一倍。同時也請在新的長方形圖上，以公分為單位標出長度和寬度。方格紙中每一個小方格都是邊長為 1 公分的正方形。



- B. 原來的長方形和新的長方形面積的比是多少？

下圖的正方形中有一個陰影的三角形。



這個陰影三角形的面積是多少？

答：_____

哪一組數是從**最大**排到**最小**？

- Ⓐ 10,011; 10,110; 11,001; 11,100
- Ⓑ 10,110; 10,011; 11,100; 11,001
- Ⓒ 11,001; 11,100; 10,110; 10,011
- Ⓓ 11,100; 11,001; 10,110; 10,011

M042003

3.4×10^2 的值為何？

- Ⓐ 3.4
- Ⓑ 34
- Ⓒ 340
- Ⓓ 3400

M042079

請在每一個格子中填入 + 或 -，使這個算式得到最大可能的總和。

$$-5 \square -6 \square 3 \square -9$$

M042018

某個班級有 30 位學生，其中男生與女生的比例是 2 : 3。
請問這個班級中有多少位男生？

- Ⓐ 6
- Ⓑ 12
- Ⓒ 18
- Ⓓ 20

M042055

一件外套平常賣 60 元。在減價 30% 的時候，阿倫買了這件外套。
請問阿倫省了多少錢？

- Ⓐ 18 元
- Ⓑ 24 元
- Ⓒ 30 元
- Ⓓ 42 元

M042039

下列哪一個與 $4x - x + 7y - 2y$ 相等？

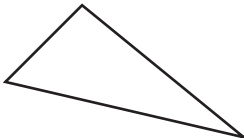
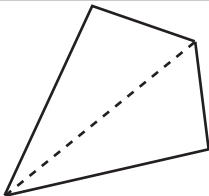
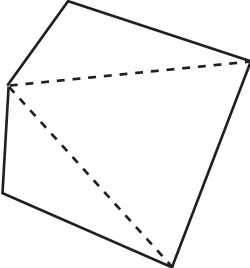
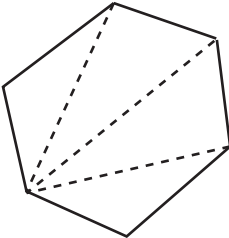
- Ⓐ 9
- Ⓑ $9xy$
- Ⓒ $4 + 5y$
- Ⓓ $3x + 5y$

M042199

內角

小傑在探索多邊形的性質。他做了一個表如下所示，想看看是否能發現邊和角度之間的關係。

A. 填空以完成這個表格。

多邊形	邊數	三角形的個數	內角和
	3	1	$1 \times 180^\circ$
	—	—	— $\times 180^\circ$
	—	—	— $\times 180^\circ$
	—	—	— $\times 180^\circ$

B. 請將正確的數填進格子中。

10 邊形的內角和 = $\times 180^\circ$

C. 小傑看出一個規律，並寫成一個以 n 表示的式子，這個式子對於任何多邊形都成立。請完成他所寫的式子。

M02_07

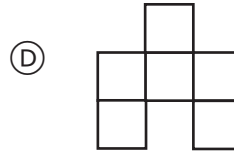
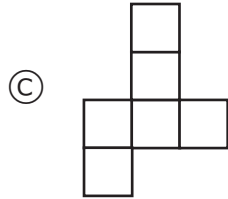
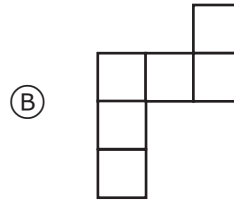
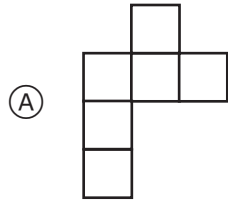
一個 n 邊形的內角總和 = _____ $\times 180^\circ$

小明知道一枝原子筆比一枝鉛筆貴 1 元。他的朋友花 17 元買了 2 枝原子筆和 3 枝鉛筆。請問小明買 1 枝原子筆和 2 枝鉛筆要付多少錢？

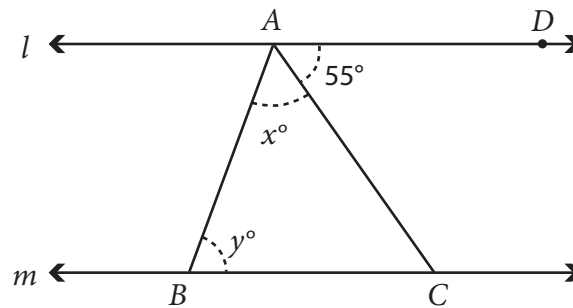
M02_08

請寫出你的算法。

下列哪一個展開圖可以折疊成一個正立方體？



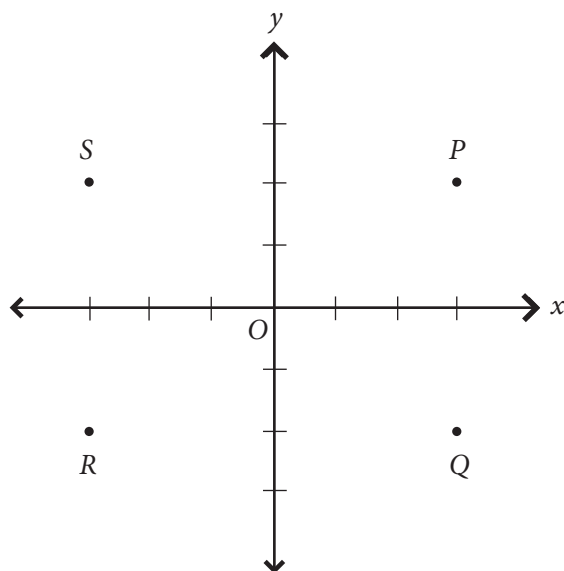
M042265



如圖，直線 l 平行直線 m ，角 DAC 的角度為 55° ，請問 $x + y$ 的值為何？

- Ⓐ 55
- Ⓑ 110
- Ⓒ 125
- Ⓓ 135

M042137



下列哪一個是代表圖中的點 $(3, -2)$ ？

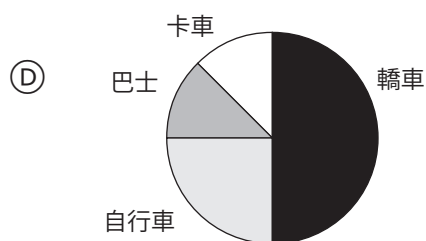
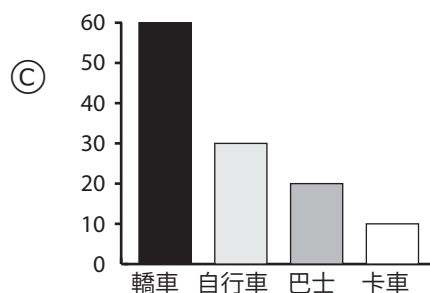
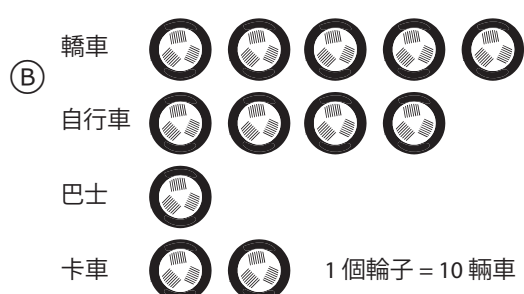
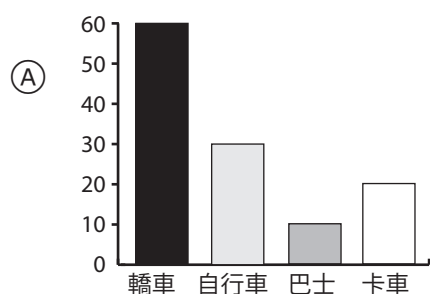
- Ⓐ P
- Ⓑ Q
- Ⓒ R
- Ⓓ S

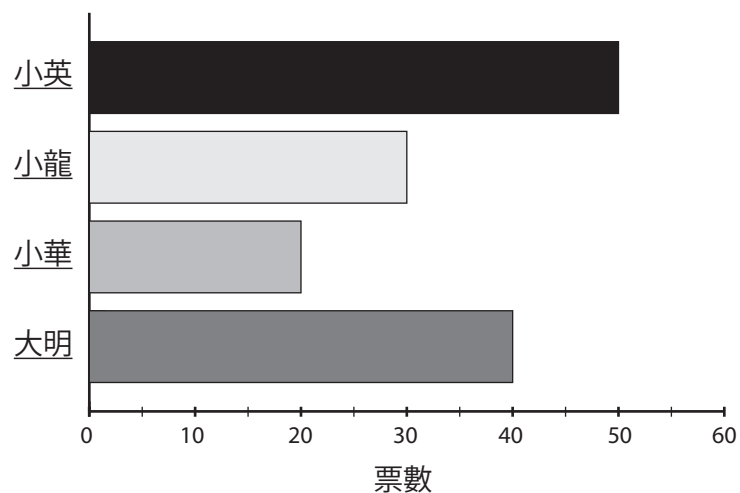
有四位學生觀察經過學校的交通狀況持續 1 小時。

下表顯示他們所看到的結果：

車輛類別	數量
轎車	60
自行車	30
巴士	10
卡車	20

每位學生畫一個圖來呈現這結果。哪一個圖是正確地呈現這結果呢？

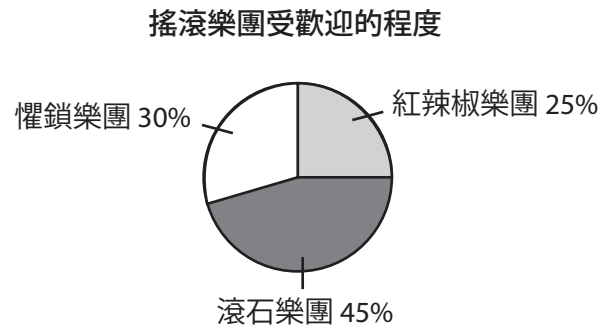




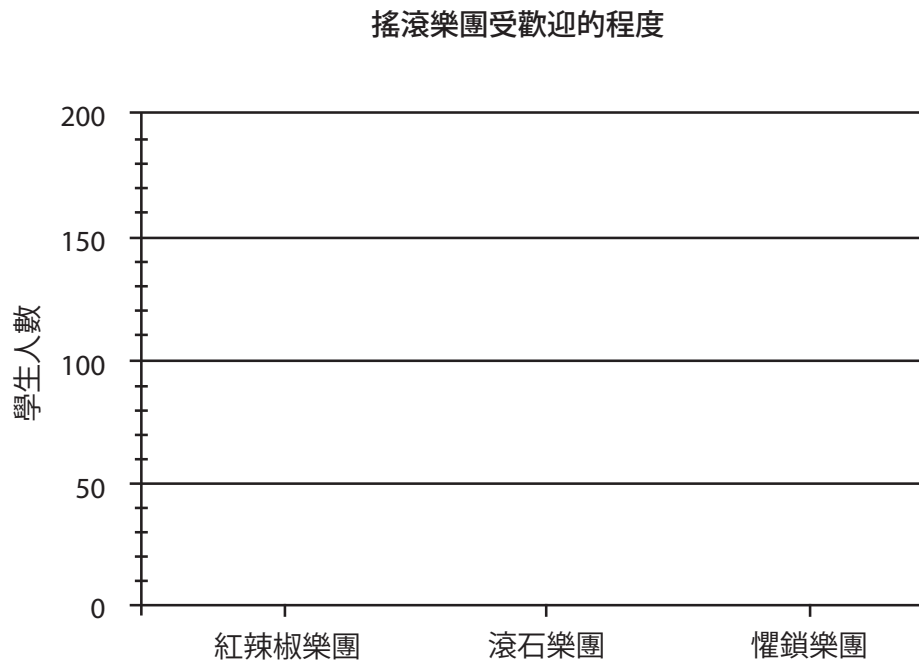
小英、小龍、小華和太明在賣校園音樂會的票。這個圖顯示出他們每一個人所賣的票數。在他們當中，有兩個人所賣的票數加起來與小英一樣多，請問是哪兩位？

答案：_____ 和 _____

對 200 個學生做調查的結果以圓形百分圖表示如下。



根據圓形百分圖中每個區塊學生的人數作出一個長條圖。

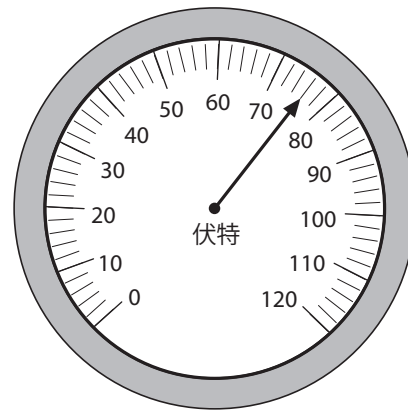


一位汽車推銷員在報紙上登了以下的廣告：「新車與中古車待售，價格不一，平均價格是 5000 元。」根據這則廣告，請問下列哪一個敘述一定正確？

- Ⓐ 大部分的汽車要賣 4000 元到 6000 元。
- Ⓑ 半數的汽車售價低於 5000 元，另外一半的汽車售價超過 5000 元。
- Ⓒ 至少有一輛車的售價是 5000 元。
- Ⓓ 有一些汽車的售價低於 5000 元。

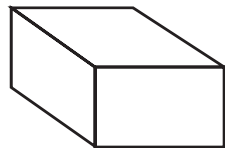
這個電壓計上表示的讀數是多少伏特？

- (A) 73
- (B) 74
- (C) 76
- (D) 78

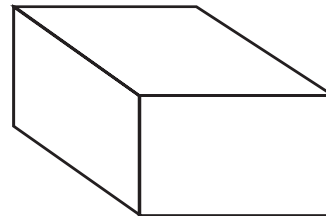


M022097

從小盒子中裝了 20 張編號從 1 到 20 的票券，大盒子中裝了 100 張編號從 1 到 100 的票券。



20 張票券



100 張票券

你可以任意從其中一個盒子裏拿出一張票券來，問哪一個盒子中拿到 17 號票券的機率比較大？

- (A) 裝 20 張票券的盒子
- (B) 裝 100 張票券的盒子
- (C) 兩個盒子的機率相同
- (D) 無法得知

M022101

下列各數中，哪一個最小？

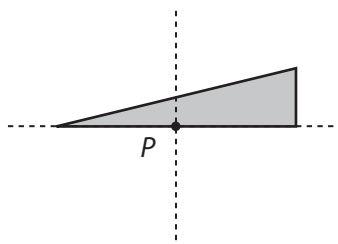
Ⓐ $\frac{1}{2}$

Ⓑ $\frac{5}{8}$

Ⓒ $\frac{5}{6}$

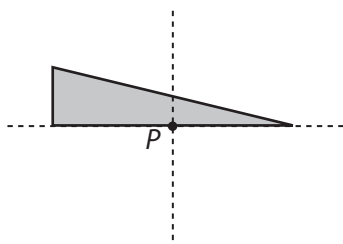
Ⓓ $\frac{5}{12}$

將平面上陰影區域繞著 P 點旋轉半圈。

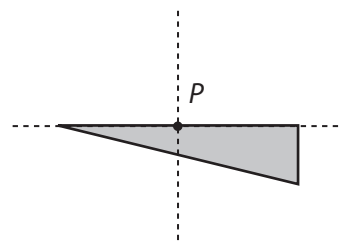


下列哪一個圖是旋轉半圈以後的結果？

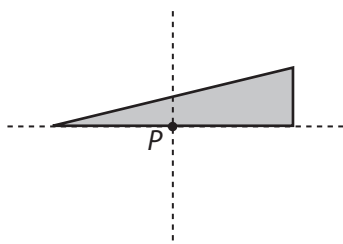
Ⓐ



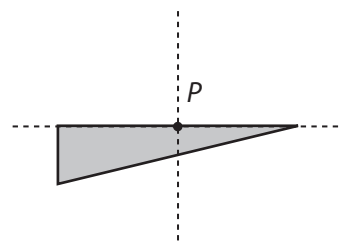
Ⓑ



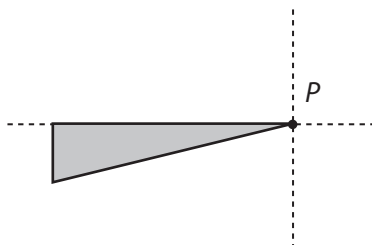
Ⓒ



Ⓓ



Ⓔ



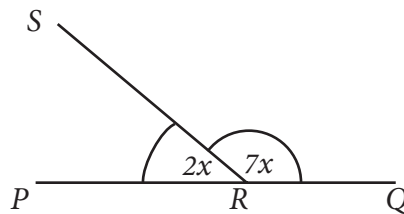
公車上有 36 位乘客，其中兒童人數和大人人數的比為 5 : 4 。公車上的兒童有多少人呢？

M03_05

答： _____

圖中， PQ 是一條直線。

M03_06



$\angle PRS$ 的度數是多少？

- Ⓐ 10°
- Ⓑ 20°
- Ⓒ 40°
- Ⓓ 70°
- Ⓔ 140°

乘乘看： $0.402 \times 0.53 =$

M03_07



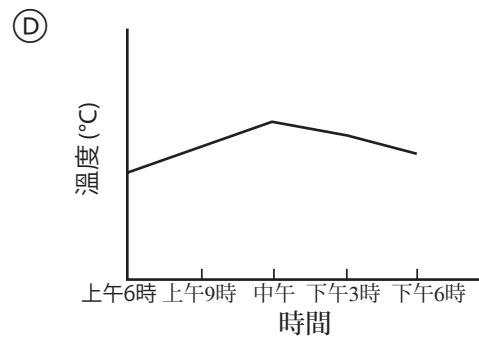
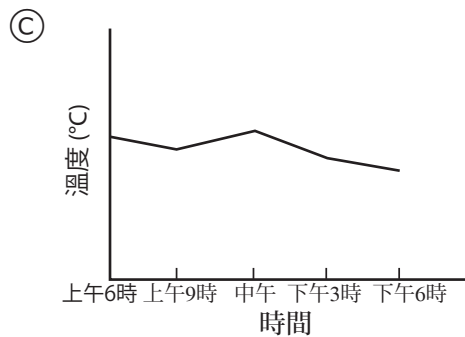
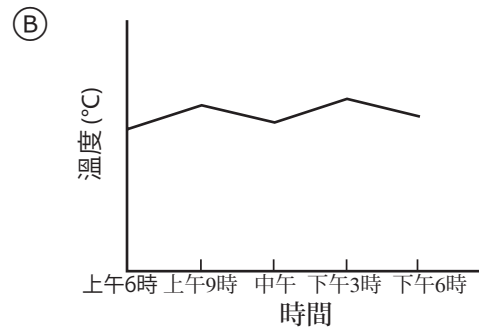
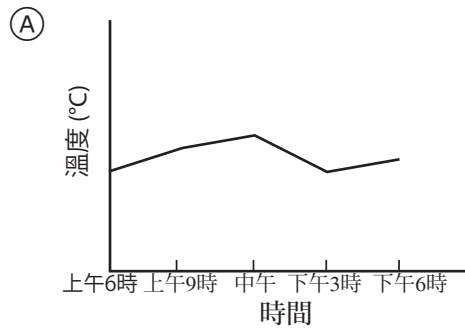
答： _____

這是某一天不同時間的氣溫紀錄表。

M03_08

時間	上午 6 時	上午 9 時	中午	下午 3 時	下午 6 時
溫度 °C	12	17	14	18	15

根據這個紀錄表畫出一個沒有標出溫度刻度的圖，下列哪一個是這個紀錄表的圖？





有一場音樂會的門票售價有 10 zeds，15 zeds，和 30 zeds 三種。在所售出的 900 張門票中，其中 $\frac{1}{5}$ 是每張 30 zeds， $\frac{2}{3}$ 是每張 15 zeds。請問所售出每張 10 zeds 的票數，佔全部售票數的幾分之幾？

答：_____

M032307

宜靜想做很多小紅莓口味的麵包，它的份量是原本食譜的 1.5 倍。若原來食譜需要 $\frac{3}{4}$ 杯的糖，則宜靜實際上需要多少杯的糖？

- Ⓐ $\frac{3}{8}$
- Ⓑ $1\frac{1}{8}$
- Ⓒ $1\frac{1}{4}$
- Ⓓ $1\frac{3}{8}$

M032523

學校舉行校外教學時，每 12 位學生就有一個老師隨行。若有 108 位學生參加校外教學，則需有多少老師隨行？

- Ⓐ 7
- Ⓑ 8
- Ⓒ 9
- Ⓓ 10

M032701

某輛公車一直保持相同速率前進，使得所走距離和所花時間成比例。如果該公車 5 小時走了 120 公里，則該公車 8 小時可走多少公里？

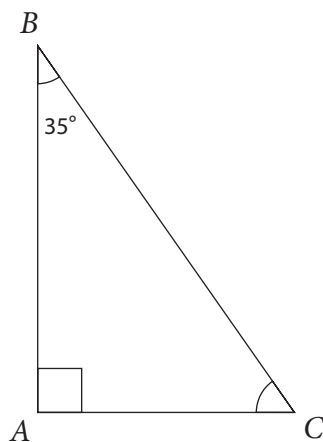
- Ⓐ 168
- Ⓑ 192
- Ⓒ 200
- Ⓓ 245

M032704

什麼數除以 -6 ，結果得到 12？

- Ⓐ -72
- Ⓑ -2
- Ⓒ 2
- Ⓓ 72

M032525



上圖三角形中，角 C 的度數是多少？

- Ⓐ 45°
- Ⓑ 55°
- Ⓒ 65°
- Ⓓ 145°

M032579

利用下面所給的線段 AO ，畫一條直線 BC 通過 O 點，使得角 AOB 是銳角且角 AOC 是鈍角，並標出點 B 與 C 。



M032691

下列哪一個數是一千零二萬零三十？

- Ⓐ 102,030
- Ⓑ 10,020,030
- Ⓒ 10,200,030
- Ⓓ 102,000,030

M042001

下列哪一個式子是將 1080 寫成質因數的連乘積？

- Ⓐ $1080 = 8 \times 27 \times 5$
- Ⓑ $1080 = 2 \times 4 \times 3 \times 9 \times 5$
- Ⓒ $1080 = 2 \times 2 \times 2 \times 3 \times 3 \times 3 \times 5$
- Ⓓ $1080 = 2^2 \times 3^2 \times 6 \times 5$

M042022

若 $a = 3$ 且 $b = -1$ ，

則 $2a + 3(2 - b)$ 的值為何？

- Ⓐ 15
- Ⓑ 14
- Ⓒ 13
- Ⓓ 9

M042082



第一根管子長 x 公尺，且第二根管子的長是第一根管子的 y 倍。
請問第二根管子有多長？

- Ⓐ xy 公尺
- Ⓑ $x+y$ 公尺
- Ⓒ $\frac{x}{y}$ 公尺
- Ⓓ $\frac{y}{x}$ 公尺

三項全能賽

三項全能運動是一種比賽，在此競賽中，選手要先游泳，再騎自行車，然後跑步，每項競賽都要完成各別規定好的距離。第一位完成整個路程的人就是優勝者，小西、小白與小蘇同時參加這個三項全能運動競賽。

這段路程包括游泳 1 公里，接著騎自行車 40 公里，然後再跑步 15 公里。

- A. 小白是最快的游泳選手，她游完 1 公里花 25 分鐘，小西比小白多花 10 分鐘，而小蘇又比小西多花 5 分鐘。

利用這些資訊去完成有關游泳比賽的表格：

游泳	<u>小西</u>	<u>小白</u>	<u>小蘇</u>
所花時間(分鐘)		25	

- B. 小西是最快的自行車選手，在 40 公里的賽程中，她平均每小時騎 30 公里。小白比小西多花 10 分鐘，而小蘇則比小西多花 15 分鐘。

利用這些資訊去完成有關自行車比賽的表格：

自行車	<u>小西</u>	<u>小白</u>	<u>小蘇</u>
所花時間(分鐘)			

- C. 小蘇是跑最快的人，她以每小時 7.5 公里的速度跑完 15 公里。
小白比小蘇多花 10 分鐘，而小西則比小白多花 5 分鐘。

利用這些資訊去完成有關賽跑的表格：

賽跑	小西	小白	小蘇
所花時間(分鐘)			

- D. 完成下表以顯示出每個人比完三項全能運動所需的總時間。

三項全能	小西	小白	小蘇
所花時間(分鐘)			

誰贏了三項全能運動比賽？

答：_____

在甲城市這個地方，運送一件貨物的總運費是由方程式 $y = 4x + 30$ 決定。
這裡的 x 代表重量，以克為單位， y 是費用，以元為單位。如果你有 150 元，你能運送多少克？

- Ⓐ 630
Ⓑ 150
Ⓒ 120
Ⓓ 30

下列哪一個等於 $2(x + y) - (2x - y)$ ？

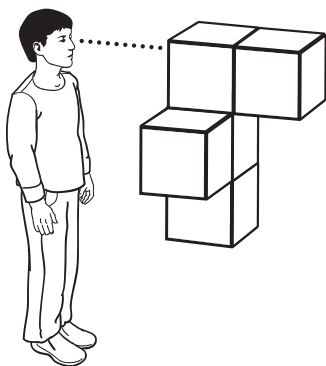
- Ⓐ $3y$
- Ⓑ y
- Ⓒ $4x + 3y$
- Ⓓ $4x + 2y$

M042239

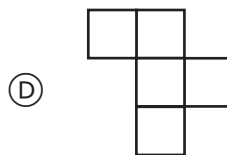
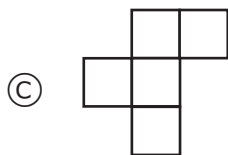
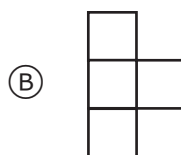
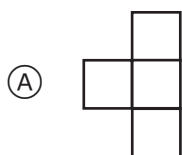
下列哪一個點在直線 $y = x + 2$ 上？

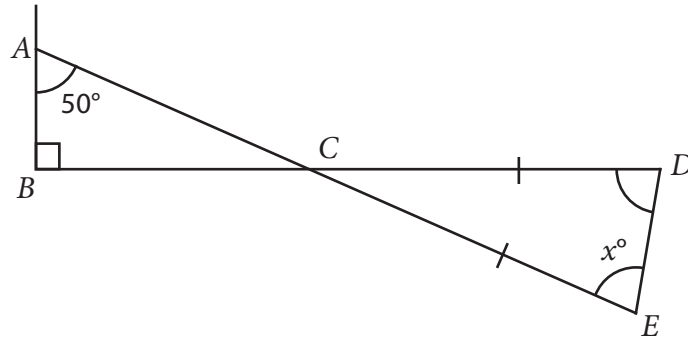
- Ⓐ $(0, -2)$
- Ⓑ $(2, -4)$
- Ⓒ $(4, 6)$
- Ⓓ $(6, 4)$

M042238



有一個立體物件是由 5 個小正立方體組成。下列哪一個形狀是圖中那個人所看到的？

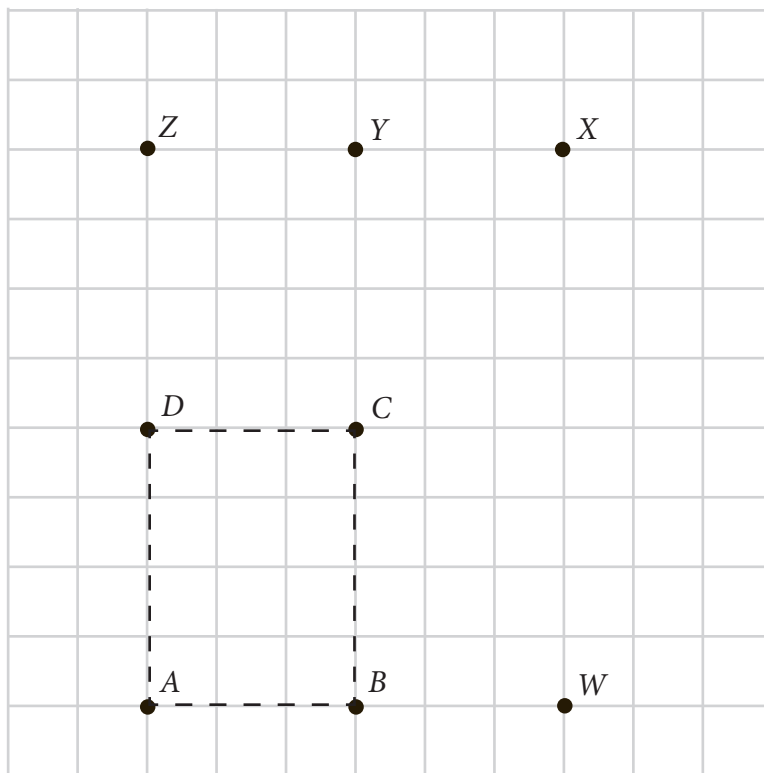




在這個圖中，若 $CD = CE$ ，
則 x 的值為何？

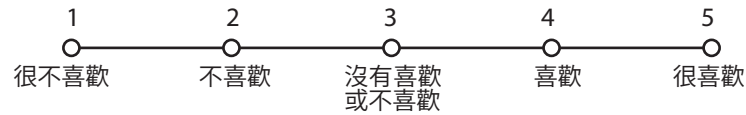
- Ⓐ 40
- Ⓑ 50
- Ⓒ 60
- Ⓓ 70

利用所標示的點畫出一個三角形，使它的面積是長方形 $ABCD$ 面積的兩倍。



學科受歡迎的程度

有 10 位學生想找出到底是數學還是歷史在他們這一群人中比較受歡迎。他們用下面這個量尺來對兩科做評分。



下表是評分結果：

學生的評分

學生	數學評分	歷史評分
亞倫	1	2
麗莎	4	4
小安	5	4
約翰	2	2
康諾	4	2
喬婕	3	3
布萊特	2	1
科尼	1	1
小珍	5	3
小杰	3	2
合計	30	24

A. 請計算每一科的平均得分。

數學科平均得分 = _____

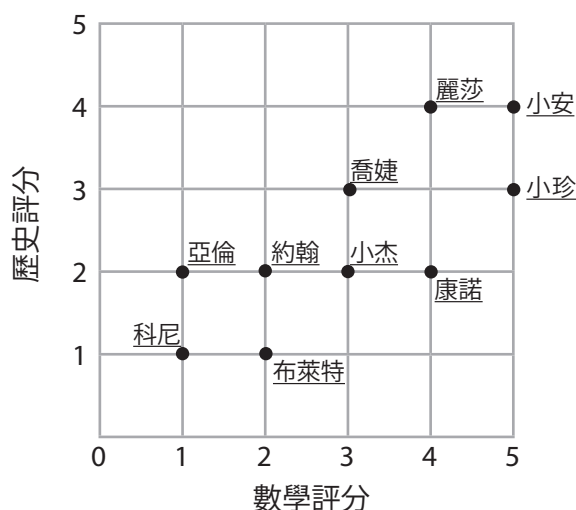
歷史科平均得分 = _____

根據這些得分，哪一個科目在這群學生之中比較受歡迎？

比較受歡迎的科目： _____

B. 學生的評分畫在下面的圖中。例如，亞倫的名字在他的評分旁邊(數學 1，歷史 2)。

M04_12



請在下列敘述後方的空格中寫下對或錯：

在這一群之中的所有學生都喜歡數學甚於喜歡歷史。 _____

有接近一半的學生給兩個科目相同的評分。 _____

有兩個學生沒有喜歡或不喜歡任何一個科目。 _____

M042303_2

菲菲有一個袋子裡面有 16 顆彈珠：8 顆紅的和 8 顆黑的。她從袋子中拿出 2 顆彈珠且不將彈珠放回。這兩顆彈珠都是黑色的，然後她再從袋子中拿出第三顆彈珠，你認為第三顆彈珠比較可能是哪一種顏色？

M04_13

- Ⓐ 紅色比黑色更有可能。
- Ⓑ 黑色比紅色更有可能。
- Ⓒ 紅色或黑色同樣可能。
- Ⓓ 你無法確定紅色或黑色哪一種比較有可能。

M042222



班級旅行

說明：第 5 題、第 6 題、第 7 題及第 8 題是有關一個班級的旅行計畫。

班級	男生	女生
1	12	9
2	14	11
3	16	12
4	18	15

上表列出四個班級的男女生人數。哪兩個班男女生的人數比是相同的？

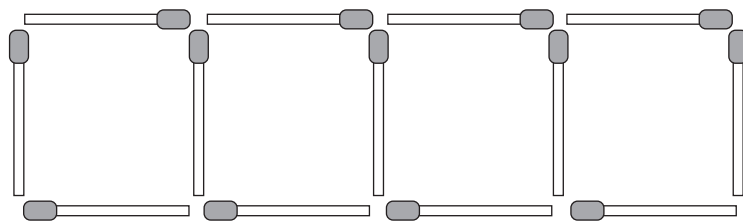
- Ⓐ 1 和 2
- Ⓑ 1 和 3
- Ⓒ 2 和 3
- Ⓓ 2 和 4

M032142

$$2a^2 \times 3a =$$

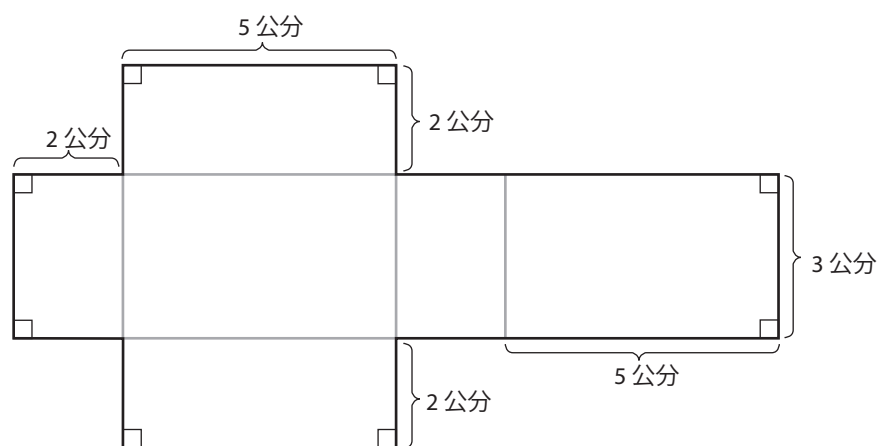
- Ⓐ $5a^2$
- Ⓑ $5a^3$
- Ⓒ $6a^2$
- Ⓓ $6a^3$

M032198



如上圖所示，用 13 根火柴棒排出一列 4 個正方形。按照這種方式，用 73 根火柴棒可以排出幾個正方形？寫出你的算法。

答：_____

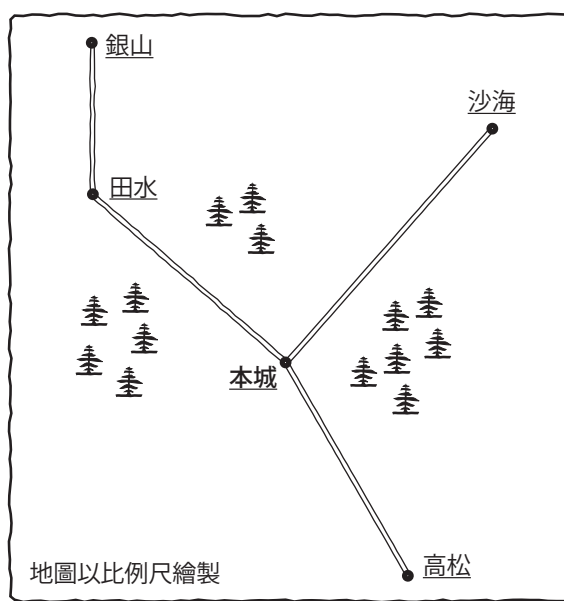


若將上圖中的圖形摺疊起來，會變成一個長方體的盒子。請問這個盒子的體積是多少？

答：_____ 立方公分

大雄和靜香計劃班級一天旅行

他們計畫從學校所在地本城到田水、銀山、沙海或高松這些城鎮當中的一個。



因為老師要求他們必須當天往返，所以他們旅行的地點，離開本城最遠不能超過 80 公里。若從本城到沙海是 80 公里，請使用上面的地圖來完成下列的表格，請在空格中填入「是」或「否」。

M05_05

	田水	銀山	沙海	高松
是否有符合少於或等於 80 公里的限制			是	

繼續班級旅行的題目 ➡

這次旅遊花費最高只有 500 zeds。已知班上有 30 位學生。
底下是到各城鎮的車費：

往田水或沙海學生票

來回票：25 zeds

25 人以上的團體少算 $\frac{1}{3}$

往銀山或高松學生票

來回票：20 zeds

15 人以上的團體打九折

請問他們負擔得起去那一個景點的旅遊？且寫出你的理由。

老師說明這次旅遊計畫的三個條件如下：

1. 我們必須在早上 9 點或 9 點以後離開本城；
2. 我們必須在傍晚 5 點之前回到本城；
3. 我們必須在某一景點城鎮停留至少 3 個小時。

大雄和靜香使用公車時刻表來找出他們是否能夠符合老師的要求。他們開始將資訊填入下面表格中，但他們尚未完成。

A. 利用下一頁的公車時刻表來完成田水那列的表格位置。

B. 利用下一頁的公車時刻表來完成沙海那列的表格位置。

	最佳公車時段					老師的條件		
目的地	離開本城的時間	到達目的地的時間	離開目的地回本城的時間	回到本城的時間	在目的地停留時間	在上午 9 點或 9 點以後離開	停留至少三個小時	傍晚 5 點之前回來
田水	上午 9:00	上午 11:15						
銀山	上午 9:15	上午 12:20	下午 2:30	下午 5:35	2 小時 10 分	是	否	否
沙海	上午 9:25							
高松	上午 9:10	上午 11:15	下午 2:40	下午 4:45	3 小時 25 分	是	是	是

田水的公車時刻表

從本城到田水的 公車時刻表	
離開本城	到達田水
上午 8:00	上午 10:15
上午 9:00	上午 11:15
上午 10:00	下午 12:15
上午 11:00	下午 1:15
中午 12:00	下午 2:15
下午 1:00	下午 3:15
下午 2:00	下午 4:15
下午 3:00	下午 5:15
下午 4:00	下午 6:15

從田水到本城的 公車時刻表	
離開田水	到達本城
上午 8:30	上午 10:45
上午 9:30	上午 11:45
上午 10:30	下午 12:45
上午 11:30	下午 1:45
下午 12:30	下午 2:45
下午 1:30	下午 3:45
下午 2:30	下午 4:45
下午 3:30	下午 5:45
下午 4:30	下午 6:45

沙海的公車時刻表

從本城到沙海的 公車時刻表	
離開本城	到達沙海
上午 8:25	上午 10:40
上午 9:25	上午 11:40
上午 10:25	下午 12:40
上午 11:25	下午 1:40
下午 12:25	下午 2:40
下午 1:25	下午 3:40
下午 2:25	下午 4:40
下午 3:25	下午 5:40
下午 4:25	下午 6:40

從沙海到本城的 公車時刻表	
離開沙海	到達本城
上午 8:35	上午 10:50
上午 9:35	上午 11:50
上午 10:35	下午 12:50
上午 11:35	下午 1:50
下午 12:35	下午 2:50
下午 1:35	下午 3:50
下午 2:35	下午 4:50
下午 3:35	下午 5:50
下午 4:35	下午 6:50

繼續下一頁班級旅行的題目 ➡

C. 那些地點符合老師所列時間限制的三個條件？

答： _____

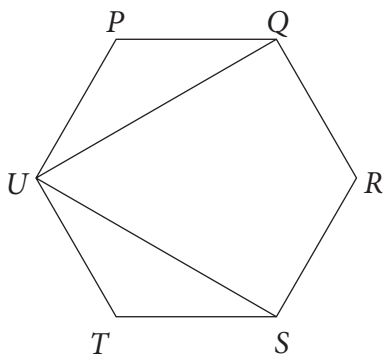
M032753_3

繼續下一頁班級旅行的題目 

考量整個旅遊的距離、老師對於時間的限制和這次旅遊的車費，能去哪一個地點旅遊？

答： _____





$PQRSTU$ 是一個正六邊形。請問角 QUS 是幾度？

- (A) 30°
- (B) 60°
- (C) 90°
- (D) 120°

M032205

下表所示是 x 與 y 之間的關係。

x	1	2	3	4	5
y	1	3	5	7	9

下列哪一個方程式可以表示這個關係？

- (A) $y = x + 4$
- (B) $y = x + 1$
- (C) $y = 2x - 1$
- (D) $y = 3x - 2$

M032163

參加旅遊的小孩人數比 55 人多，但比 65 人少。若小孩分成 7 人一組，恰好全部都分完組，但若 8 人一組，則無法恰好分完，請問共有多少小孩參加此次旅遊？

M07_01

答： _____

哪一個是求出 $\frac{1}{5} - \frac{1}{3}$ 的正確步驟？

M07_02

Ⓐ $\frac{1}{5} - \frac{1}{3} = \frac{1-1}{5-3}$

Ⓑ $\frac{1}{5} - \frac{1}{3} = \frac{1}{5-3}$

Ⓒ $\frac{1}{5} - \frac{1}{3} = \frac{5-3}{5 \times 3}$

Ⓓ $\frac{1}{5} - \frac{1}{3} = \frac{3-5}{5 \times 3}$

有一金和銀的合金，是以金 1 公克、銀 4 公克的比例混合而成。在 40 公克的這種合金中，銀有多少公克？

- Ⓐ 8
- Ⓑ 10
- Ⓒ 30
- Ⓓ 32

M032160

有一數列為 2, 5, 11, 23, ...

該數列的首項為 2。下列哪一個規則可以說明該數列數字變化的規律？

- Ⓐ 前一項加 1 後，乘以 2。
- Ⓑ 前一項乘以 2 後，加 1。
- Ⓒ 前一項乘以 3 後，減 1。
- Ⓓ 前一項減 1 後，乘以 3。

M032273

$$3(2x - 1) + 2x = 21$$

x 的值為何？

Ⓐ -3

Ⓑ $-\frac{11}{4}$

Ⓒ $\frac{11}{4}$

Ⓓ 3

M032540

小明比小英多三件夾克。若 n 是小明的夾克數，則小英的夾克數是多少？
(以 n 表示)

Ⓐ $n - 3$

Ⓑ $n + 3$

Ⓒ $3 - n$

Ⓓ $3n$

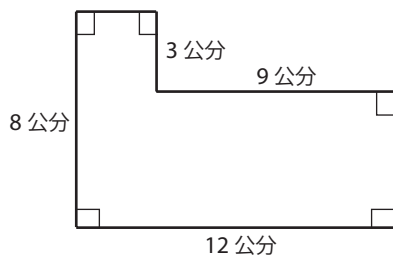
M032698

在一個半徑 10 公尺的圓形池塘中，平均每平方公尺有 2 隻青蛙，則池塘中約有幾隻青蛙？

π 約 3.14

- Ⓐ 120
- Ⓑ 300
- Ⓒ 600
- Ⓓ 2400

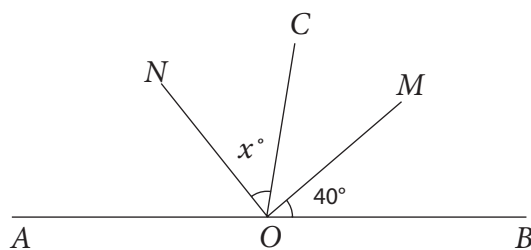
M032097



上圖的面積是多少平方公分？

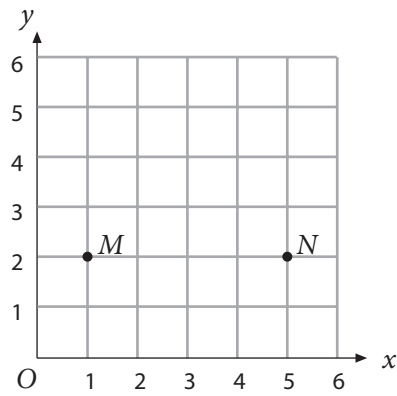
- Ⓐ 66
- Ⓑ 69
- Ⓒ 81
- Ⓓ 96

M032575



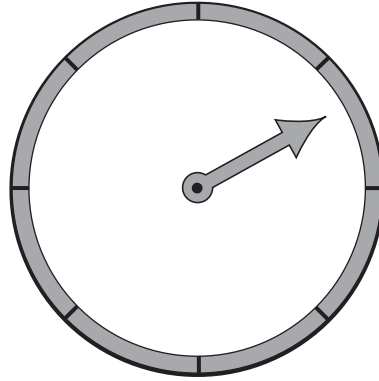
上圖中， A 、 O 、 B 三點共線。 OM 平分角 BOC ， ON 平分角 AOC 。求 x 的值是多少？

答：_____



點 M 和 N 的位置如上圖所示。小華想找出點 P ，使得 MNP 為等腰三角形。則下列何者可能是點 P 的位置？

- Ⓐ (3,5)
- Ⓑ (3,2)
- Ⓒ (1,5)
- Ⓓ (5,1)



小國的轉盤有橘色、紫色和綠色三個不同顏色的扇形區域。小國將指針旋轉 1000 次。下表為指針停在每一個扇形區域的次數：

顏色	停的次數
橘色	510
紫色	243
綠色	247

根據上表，請將上面的轉盤以線劃分成三個扇形區域，並分別標示出橘色、紫色及綠色所在區域。

M032688

在瑞德蘭這個地方，一件外套原價為 120 zeds。在大拍賣時，這件外套賣 84 zeds。請問這件外套的售價下降了百分之多少？

- (A) 25
- (B) 30
- (C) 35
- (D) 36

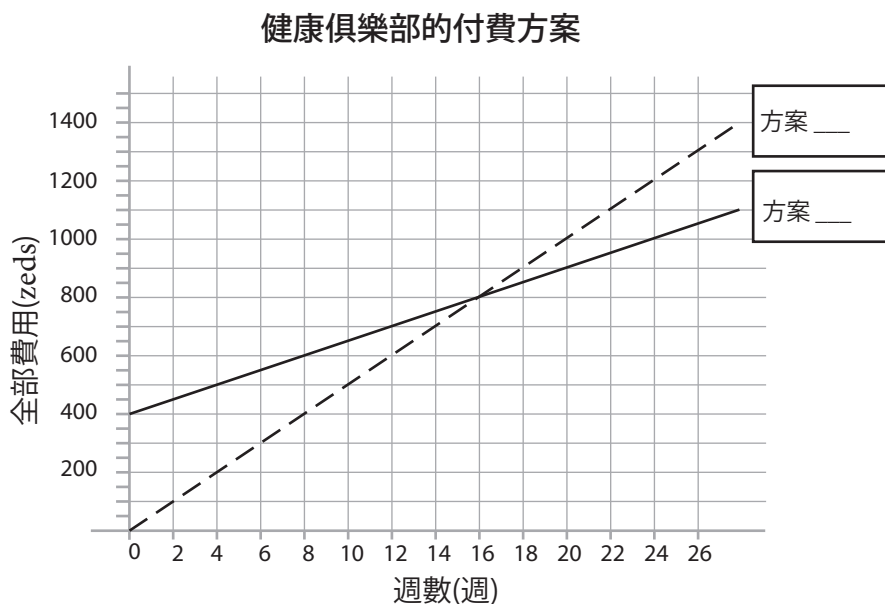
M032529

某健康俱樂部提供兩種不同的付費方案。

方案 A 是先付 400 zeds，然後每週再付費 25 zeds。

方案 B 是不須先付費，但每週付費 50 zeds。

下圖是比較方案 A 和方案 B 的費用。



A. 標示出方案 A 的費用所代表的直線和方案 B 的費用所代表的直線。

B. 在第幾個週時，使用方案 A 及方案 B 會繳付同樣金額？

C. 在第 24 週時，兩方案所花的全部費用相差多少？

附錄四

TIMSS 2007 四年級科學可公布試題（中文版）



光和顏色

說明:第 **23**、**24**、**25** 題是關於光和顏色的問題。回答這些問題時，你可以使用這一小節中有關於光和顏色的任何一張圖片。

小美和安安走進一家服飾店，買了一件橘色的襯衫。在回家的路上，他們打開袋子，把這件橘色的新襯衫給一位朋友看。當他們看到襯衫竟是紅色而不是橘色時，感到很驚訝。



在服飾店



在回家的路上

小美認為店員忙中有錯，拿錯了襯衫，安安認為新襯衫的顏色看起來會不一樣的原因是由於陽光和服飾店的燈光不同。他們決定調查一下，到底誰的想法才正確。

光和顏色的問題由此頁開始。 ➡



研究新襯衫

小美和安安買了一個燈和四個顏色的燈泡——一個白色燈泡、一個紅色燈泡、一個黃色燈泡和一個綠色燈泡。他們拿出新買的襯衫，在不同顏色的燈泡下觀察有何變化。下面的圖表示他們所看到的結果。

在不同顏色的燈泡下看到的新襯衫的顏色：



甲、請寫下小美和安安調查新襯衫的結果。

乙、小美和安安是否真的被給錯了襯衫？

(請勾選一項)

☐ 是

☐ 否

請用前面他們調查的結果，來寫你的理由。

丙、服飾店中的燈泡是什麼顏色？

答：_____

研究一件白襯衫

小美和安安好奇其他顏色的襯衫在不同光線下看起來會怎樣，接著他們拿了一件白襯衫，並且在每個燈泡下觀察它。

下圖表示他們所看到的情形。

在不同的燈泡下看到的白襯衫的顏色



甲、請寫下他們調查白襯衫的結果。

乙、若在藍色燈泡下，你想白襯衫看起來會是什麼顏色？

答：_____

接著小美和安安又拿出另一件襯衫。它在白色燈泡下看起來是藍色的。



白色燈泡

藍色燈泡



?

看起來是藍色

當在藍色燈泡下看這件襯衫時，你認為它看起來會是什麼顏色？

答：_____

請用小美和安安研究的結果，寫出你的理由。



下列哪一種動物的牙齒與人類的牙齒最像？

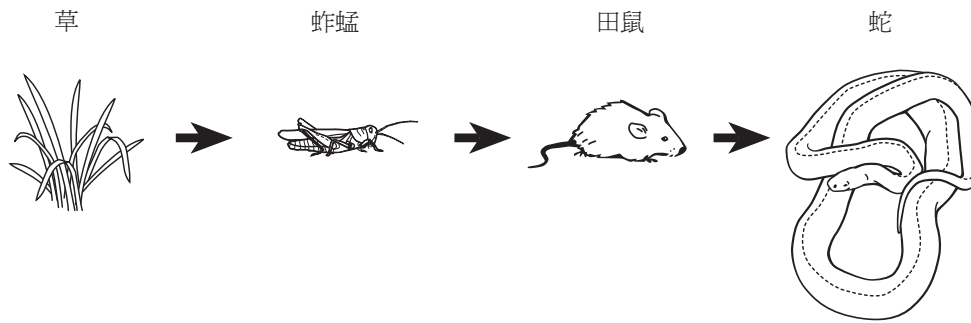
- ① 鹿
- ② 獅子
- ③ 猴子
- ④ 狗

S031193

有些人直髮，有些人捲髮，是什麼因素決定人們生下來是捲髮還是直髮？

- ① 他們父母的髮質
- ② 他們兄弟姊妹的髮質
- ③ 他們自己頭髮的顏色
- ④ 他們自己皮膚的顏色

S031264



看上圖有關下列食物鏈的敘述，哪一個才是正確的？

- ① 田鼠吃蚱蜢和草
- ② 蚱蜢吃草與田鼠
- ③ 蛇吃田鼠
- ④ 蛇吃草

S031347

文森告訴他的朋友小琳，他能從水果中獲得維持健康的所有營養成分。小琳則認為文森也要吃一些其他種類食物。

他們誰對呢？

（請在 ☐ 中勾選一個）

☐ 文森

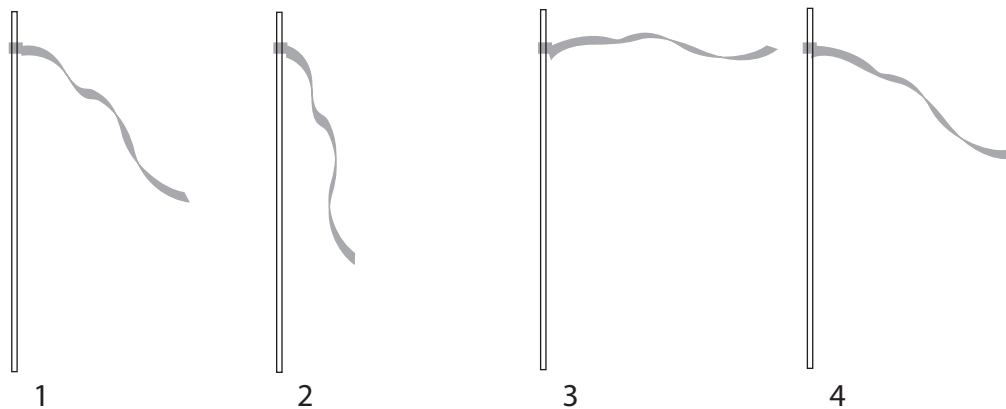
☐ 小琳

請說明你的答案。

☐

S031346

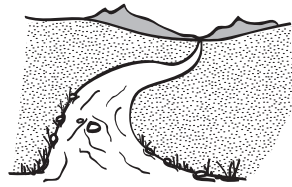
如下圖所示，將一條絲帶綁在竹竿上以測量風力的大小。



依風力由**強**到**弱**的次序，將每個圖下面的號碼寫在橫線上。

答：____, ____ , ____ , ____

請看以下這四張圖。



河流



樹木



種子



火

哪些圖代表**無生命**的東西？

- ① 樹木和火
- ② 火和河流
- ③ 河流和種子
- ④ 種子和樹木

S041007

有一種六隻腳的動物。

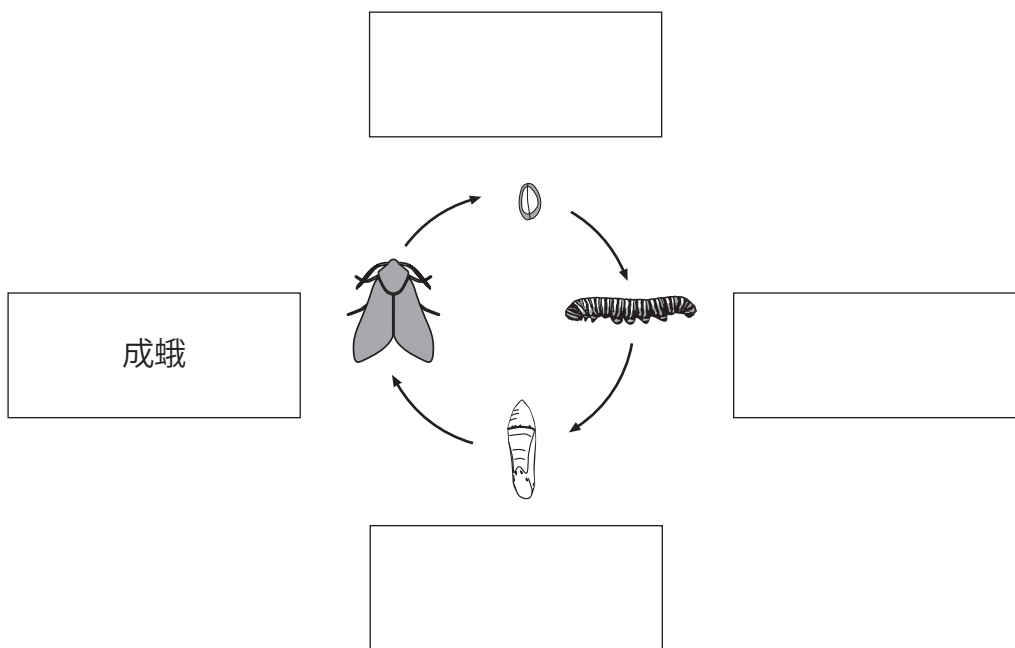
請問牠最有可能是下列哪一種動物？

- ① 蜘蛛
- ② 蒼蠅
- ③ 蜥蜴
- ④ 蜈蚣

S041164

下圖為蛾的成長週期。

請在下面方格內填入每個階段的名稱，其中一個階段已經填寫完成。



S041018

請問植物從哪裡得到能量來製造食物？

- ① 空氣
- ② 泥土
- ③ 水
- ④ 太陽光

S041160

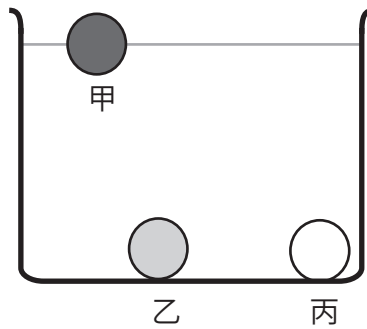
抽菸對身體有害，它對身體的哪一個器官造成的傷害最大？

- ① 肺
- ② 腎
- ③ 肝
- ④ 胃

S041042

大明有三件形狀和大小都相同的物體。

大明把這三件物體放置在一杯裝有水的燒杯裡，他觀察到甲浮在水面上而乙跟丙卻下沉。

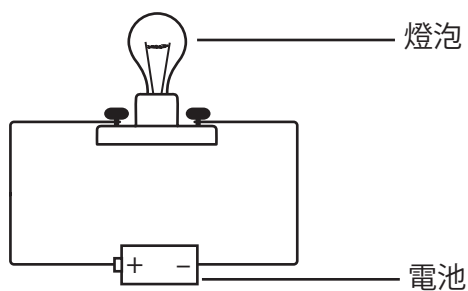


他將物體甲的重量和乙及丙的重量做比較，他能怎麼說物體甲的重量？

- ① 甲比乙或丙輕
- ② 甲比乙或丙重
- ③ 甲比乙輕可是比丙重
- ④ 甲比乙重可是比丙輕

S041079

在下面電路圖中，燈炮並不會亮。



其中可能的原因是燈炮破掉了。

請再舉出一個燈炮不會亮的原因。

S041073

小華有鐵屑和沙形成的混合物，她想把它們分離。

請問她應該怎麼做？

- ① 搖晃混合物，鐵屑會浮到表面
- ② 加入水到混合物中，沙子會在水中溶解
- ③ 將混合物倒入篩子，沙子會留在篩子裡
- ④ 將磁鐵靠近混合物，磁鐵會吸住鐵屑

S041217

下列哪一項**只能**靠電力才能運轉？

- ① 帆船
- ② 機車
- ③ 風扇
- ④ 蒸氣引擎

S041196

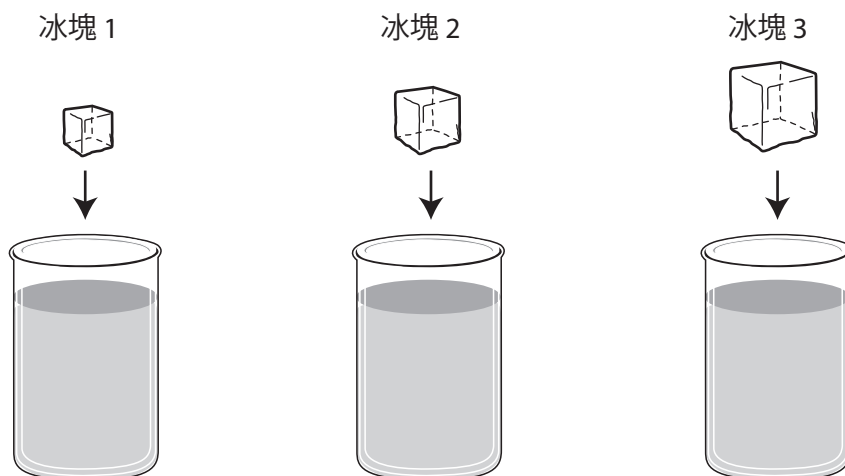
在一個天氣晴朗的日子裡，小美的老師把一個裝水的盤子放置在窗台上。
當小美晚上再去看這個盤子時，他發現裡面的水已經沒有了。



請解釋為什麼水不見了。

S041211

文姐有三個不同大小的冰塊，她把冰塊放進各裝有相同水量的燒杯中，如圖所示：



當這三塊冰塊都被放進水裡時，會產生什麼現象？

- ① 冰塊 1、2、3 都會下沉
- ② 冰塊 1、2、3 都會浮在水面上
- ③ 冰塊 1 會浮在水面上，而冰塊 2、3 卻下沉
- ④ 冰塊 1、2 會浮在水面上，而冰塊 3 卻下沉

S041051

地球表面絕大部分是被什麼所覆蓋？

- ① 沙
- ② 樹
- ③ 水
- ④ 山

S041089

下表為太陽系中的四個星球與太陽之間的距離。

	地球	火星	水星	土星
離太陽大約的 距離 (百萬公里)	150	230	58	1,400

甲. 請問下列哪一個行星離太陽最近？

- ① 地球
- ② 火星
- ③ 水星
- ④ 土星

乙. 請問這四個行星，哪一個表面的平均溫度是**最低**的？

下列哪一組動物都是哺乳類？

- ① 鴨子、老鷹、鸚鵡
- ② 老鼠、猴子、蝙蝠
- ③ 蝴蝶、螞蟥、蚊子
- ④ 鱷魚、蛇、烏龜



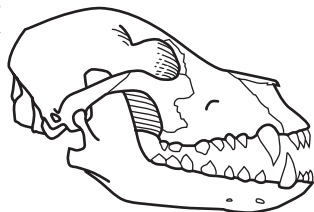
橡樹和向日葵種在一起，能不能長出一棵會開向日葵花的橡樹？

（在下面 ☐ 中勾選一個）

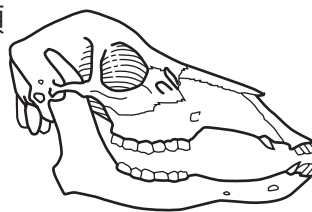
- ☐ 能
- ☐ 不能

解釋你的答案。

1 號骨頭



2 號骨頭



上面的圖代表兩個動物的頭骨——1號頭骨和2號頭骨。其中一種動物只吃植物，另外一種動物只吃其他動物。請判斷上面兩個頭骨哪一個是只吃其他動物的頭骨？哪一個是吃植物的頭骨？請將正確的號碼寫在下面的橫線上。

草食性動物骨頭是 _____ 號頭骨。

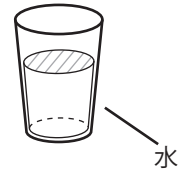
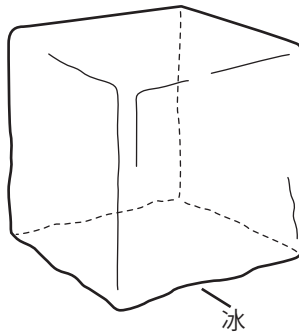
肉食性動物骨頭是 _____ 號頭骨。

請根據你對動物牙齒的認識，寫出你為什麼這樣認為的理由。

冬冬不小心割傷了他的手指，他的身體需要能量來幫助傷口復原，請問他可以從哪裡獲得能量來治好他的傷口？

- ① 從他包紮在傷口上的繃帶獲得能量。
- ② 從他塗在傷口上的藥膏獲得能量。
- ③ 從他所吃的食物獲得能量。
- ④ 從他所喝的水獲得能量。

S031319

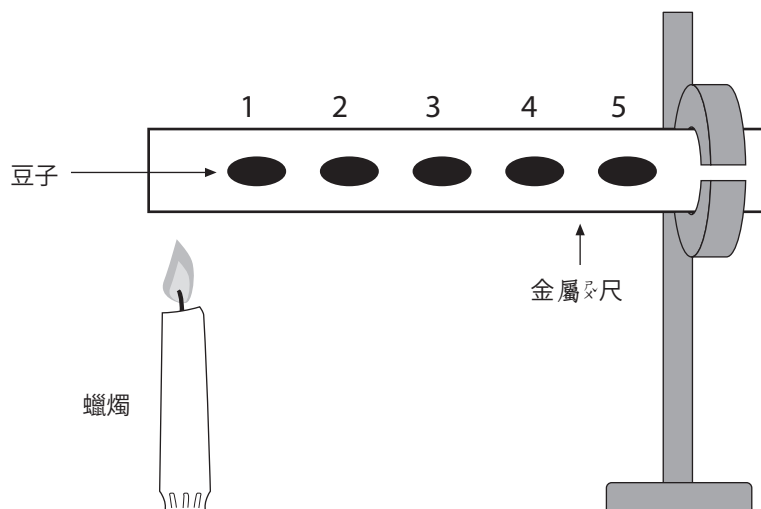


冰和水都是不同形態的水，每一種形態各有不同的用法，請針對每一種形態寫出一種人們的用法。

冰：

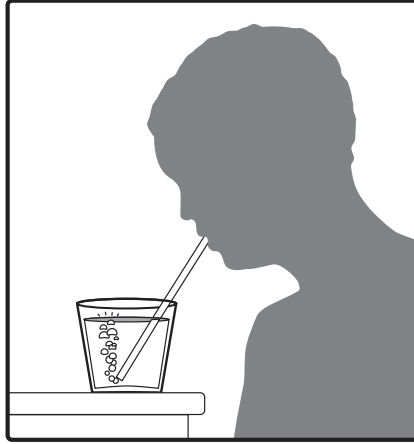
水：

S031414



如上圖所示，用奶油將豆子固定在金屬尺上。在尺的一端加熱，豆子會按哪一個順序掉下來呢？

- ① 1, 2, 3, 4, 5
- ② 5, 4, 3, 2, 1
- ③ 1, 3, 5, 4, 2
- ④ 全部同時掉下來



當你用吸管吹氣到水中，會有氣泡往上升，為什麼氣泡會往上升？

S031009

下面哪一個是我們看得到月亮的主要理由？

- ① 月亮會反射由地球發出來的光。
- ② 月亮會反射由太陽發出來的光。
- ③ 月亮會自己發光。
- ④ 月亮比其它星星大。

S031401

甲、請寫出一年中兩個不同季節的名稱。

答案 1: _____

答案 2: _____

乙、請寫出一個有關於上面兩個季節之間天氣的差別。

請問下列哪一種動物具有外骨骼？

①



貓

②



螞蟻

③



魚

④



蛇

S041165

一棵植物的種子可能遠離原來的植物。

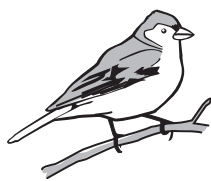
請描述一種發生此現象的方法。

S041023

下列哪一項是維他命及礦物質最好的來源？

- ① 水果及蔬菜
- ② 麵包、米飯、通心粉
- ③ 牛奶和乳製品
- ④ 肉、魚和家禽

S041047



鳥



雲

鳥是有生命的，雲是無生命的。

請舉出兩個理由，為什麼鳥被分在有生命的一類，而雲被分在無生命的一類？

1.

2.

志賢和怡君有來自同一棵向日葵的種子，他們用兩個相同的瓦盆，放入土壤然後擺進種子，志賢在他家照顧一個瓦盆，怡君也在她家照顧另一個瓦盆。

經過一段時間，他們比較瓦盆內的植株，發現它們在成長上有很大的不同，如下圖。



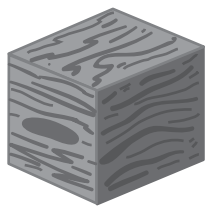
志賢的向日葵



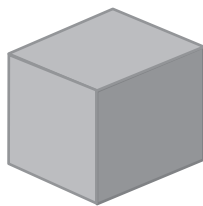
怡君的向日葵

請寫出一項志賢照顧盆栽的方法可能和怡君不同的地方。

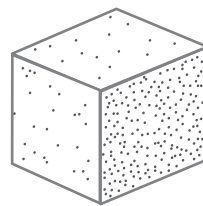
下面這三個物體是同樣的形狀和大小。



木頭



鐵塊



保麗龍塊

關於這三個物體重量的描述，下列哪一個最可能是正確的？

- ① 木頭做的物體最重
- ② 鐵做的物體最重
- ③ 保麗龍做的物體最重
- ④ 三個物體一樣重

S041054

下面哪一個例子中的物體是因為地心引力而移動的？

- ① 女孩用球棒擊了球
- ② 男孩在地板上推箱子
- ③ 女孩將鐵釘敲到牆壁裡
- ④ 男孩從樹上掉落到地面

S041308

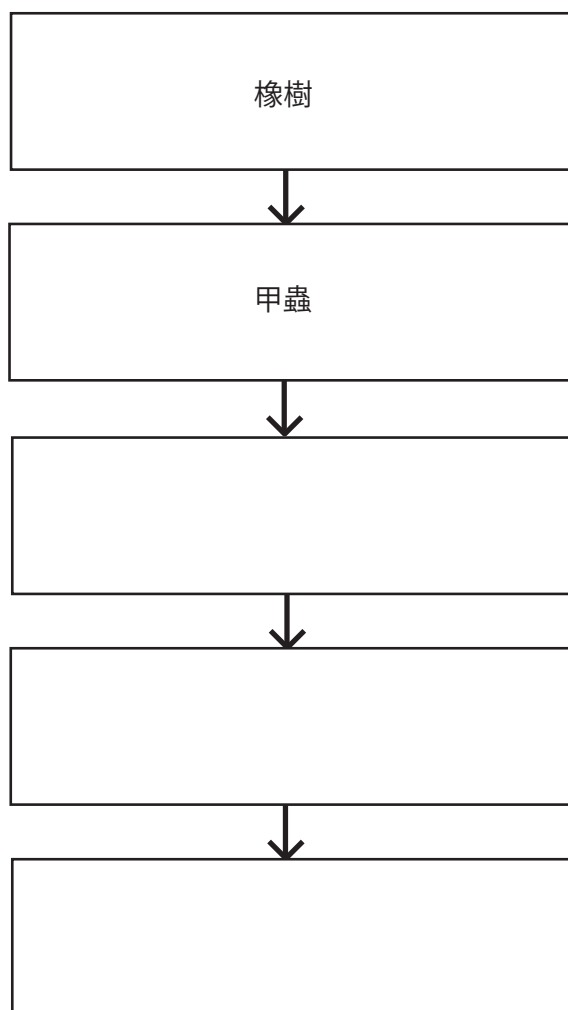
下表顯示一些生物在群落中的能量來源。

生物	能量來源
蛇	小型哺乳動物，鳥、青蛙
甲蟲	葉子
老鷹	小型哺乳動物，小爬蟲類、魚
橡樹	太陽
青蛙	昆蟲

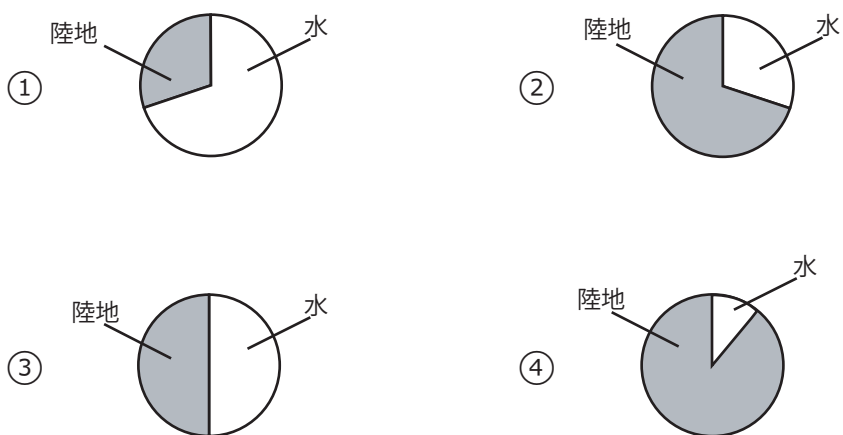
請利用上面的資料完成下列食物鏈。

在各空格內填寫一種生物。

(箭頭顯示能量流動的過程)

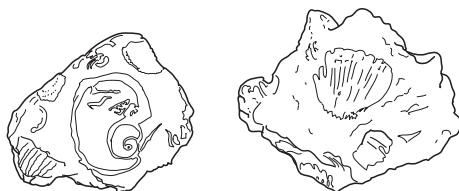


下面哪一個圓形統計圖，代表地球上陸地和水的比例？



S041087

下面的圖片表示兩塊在山邊的岩石層中發現的貝類化石。



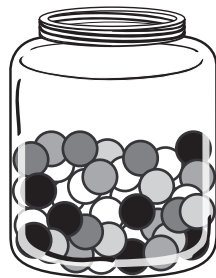
為什麼它們會在岩石層中被發現，最有可能的原因是下列哪一個？

- ① 貝類以前在陸地生活
- ② 在山中形成的化石保存得比較久
- ③ 人們把貝類的殼丟在山邊
- ④ 這些岩石層以前是海底的一部份

S041205

雅琪的玻璃罐裡有許多金屬球。

S04_11



每一個金屬球體積相同，但都是由不同的金屬所構成。

請舉出一項金屬的特性，讓她可以用來將玻璃罐中的金屬球做分類。

常溫中，物體以固體、液體或氣體三種狀態中的一種存在。

下表顯示幾種物體，已經以他們所屬的狀態而分類。

請在下列空格中，填入三類物體的狀態名稱。

S04_12



編號	物體	狀態
1	水和果汁	
2	空氣和氧	
3	石頭和黃金	

水是一種在地球發現的天然資源，也是每天生活要用的物質。

請再列出**另**一種每天都被使用的自然資源。

S041202

請描述該自然資源有何用處？

沙漠有很多不同的型態，請問它們有什麼共同點？

- ① 溫暖的冬天
- ② 漫長的夏天
- ③ 很低的降雨量
- ④ 日夜溫度都很低

S041215

小蚱蜢長大時會脫殼，下列哪一種動物長大時也會脫殼？

- ① 青蛙
- ② 蛇
- ③ 魚
- ④ 人類

S031255



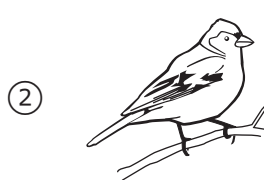
參考上圖有關人類骨骼的構造圖，寫下人類為何需要骨骼的兩種理由。

理由 1:

理由 2:

S031240

下列哪一種鳥類最有可能吃小型的哺乳動物？



S031239

海洋中發現一種新的動物，牠可能是魚類也可能是哺乳類。請寫出可以幫助你辨認它是魚類或是哺乳類的一種特徵來。

甲、魚類的特徵：

乙、哺乳類的特徵：

S031235

小柔分別測量一杯冷水、一杯溫水和一杯熱水各可以溶解多少糖，她很可能觀察到什麼現象？

- ① 冷水可以溶解最多糖
- ② 溫水可以溶解最多糖
- ③ 熱水可以溶解最多糖
- ④ 冷水、溫水與熱水溶解一樣多的糖

液體能夠變成氣體也能夠變成固體。

甲、請寫下來你如何使液體變成氣體？

乙、請寫下來你如何使液體變成固體？

S031393

人們爲什麼不應該直接飲用海水，請寫出理由？

S05_07

S031278

寫出一個太陽和月亮不一樣的地方：

S05_08

下列哪一種生物是利用太陽光來製造他們自己的食物？

- ① 蜥蜴
- ② 樹
- ③ 鹿
- ④ 老鷹

在一個小島上有一隻巨大的烏龜，牠是所剩下唯一的一隻特殊巨龜。牠能否再繁殖以避免絕種？

☐

（在下面 ☐ 中勾選一個）

- ☐ 是
- ☐ 否

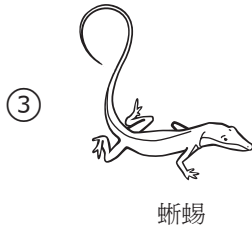
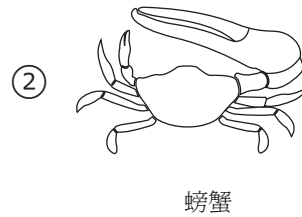
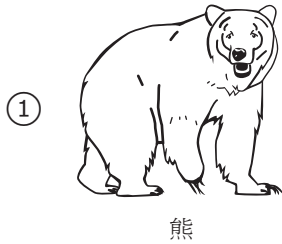
請寫出你選上面答案的一個理由。

生長在海洋中的海藻通常不容易在潮池中找到。（潮池是指在漲潮時會被海水灌滿的岩洞。）有四個學生想知道「海藻不能在潮池中生長」是否因為海水太鹹的緣故。於是每人設計了一組實驗來驗證這個想法。下面就是這四個學生所設計的實驗。

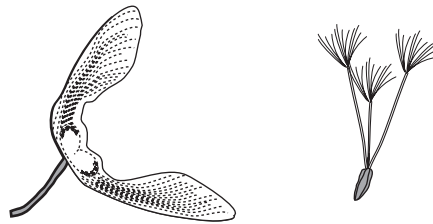
哪一組的實驗裝置最能調查出「海藻不能夠在非常鹹的海水中生存」？



下列哪一種動物最可能生活在沙漠裡？



S031283



上圖是二種種子的形狀，它們是如何散播的？

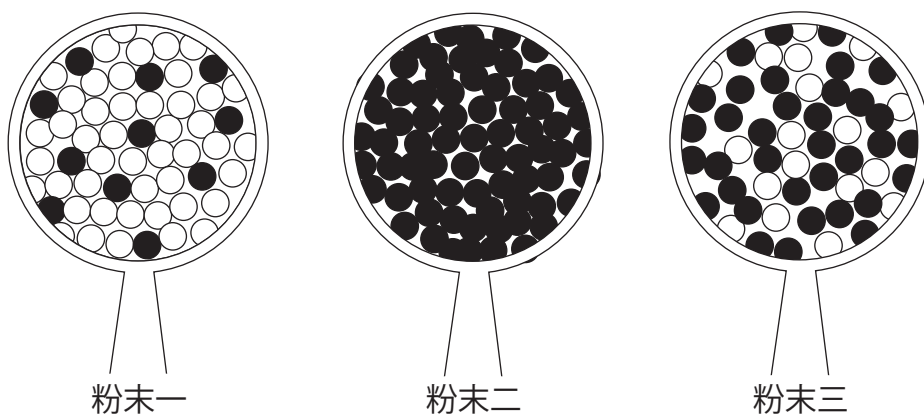
- ① 靠破裂力量彈出
- ② 靠黏在動物身上
- ③ 靠動物吃它
- ④ 靠風力

S031426

下列哪一種物體會生銹？

- ① 木屑
- ② 塑膠吸管
- ③ 鐵釘
- ④ 玻璃珠

S031422

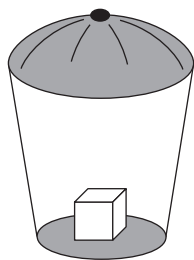


上圖表示在放大鏡下所看到的三種不同的粉末，哪些看起來最像混合物？

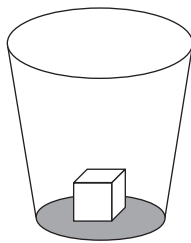
- ① 只有粉末一和粉末二
- ② 只有粉末一和粉末三
- ③ 只有粉末二和粉末三
- ④ 粉末一、二和三

S031427

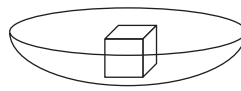
下面哪一個冰塊融化的最慢？



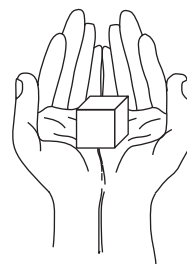
①



②



③



④

S031075

一些車子整晚停放在外面。晚上雖然沒有下雨，可是在清晨時看見車身是濕的。

水從哪兒來的呢？請將你的答案寫在下面。

S031047

科學家們相信很久以前海水曾淹沒過目前大部份的陸地。
科學家在陸地上發現了什麼，使他們相信這件事是真的？

- ① 地下水
- ② 沙土
- ③ 魚的化石
- ④ 鹹水湖

S031387

木頭是人類使用的一個資源。舉出二個不同的例子，說明木頭能做甚麼用？

1.

2.

S07_11

附錄五

TIMSS 2007 四年級數學可公布試題（中文版）

一個停車場中，共有 762 輛車子停成 6 排，每排停放的數量相同，請問每排停放了幾輛車？

M01_01

答：_____

$$\begin{array}{r} 942 \\ -5\clubsuit7 \\ \hline 415 \end{array}$$

M01_02

小玲回家做功課時，將桌上的飲料打翻了，不小心弄髒了減法題目的一個數字，算出的答案 415，請問被抹掉的數字是多少？

答：_____

去年陽光國小有 92 個男生及 83 個女生，今年全校有 210 個學生，其中 97 個是男生。請問今年的女生比去年的女生多幾個？請寫下你的作法。

M01_03

答：_____



阿信利用相同的規則 從 △ 內的數字得到 □ 內的數字。請問阿信的規則是什麼？

答：_____

M031227

某一天的早上 7 點時的氣溫為 12°C 。氣溫每小時會上升 2°C ，一直到早上 11 點時達到了 20°C 。請問在早上 9 點時的氣溫是多少？

- ① 14°C
- ② 15°C
- ③ 16°C
- ④ 17°C

M031335

小冬，小露，小琳下課一起從學校走路回家。到小琳的家需要花 25 分鐘，接著小冬、小露再花 10 分鐘到小露家。最後小冬又花 5 分鐘到自己家。如果小冬要在下午 3：50 回到家，他們必須在幾點離開學校？

M01_06

答：下午：_____

一只瓶子裝 1 公升的水。小華將瓶中的 250 毫公升的水倒入一個玻璃杯中，請問瓶子中還剩下多少水？

M01_07

答：_____ 毫公升

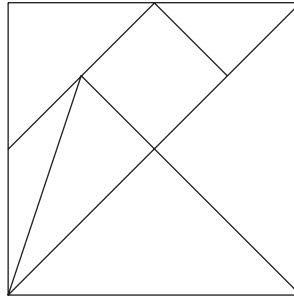
安安想知道他的貓咪有多重。首先他量出自己的體重是 57 公斤。然後，他抱著貓一起稱重，結果磅秤顯示 62 公斤。

M01_08

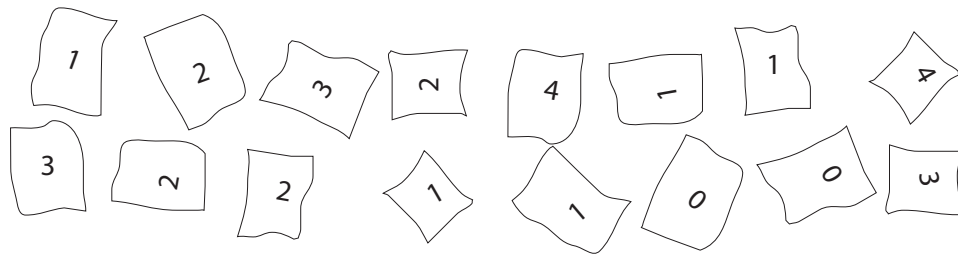
請問貓咪的重量是幾公斤？

答：_____ 公斤

一個正方形切割成 7 小塊。請在大小及形狀都相同的兩個三角形分別畫上 "X" 的記號。



M01_09



M01_10



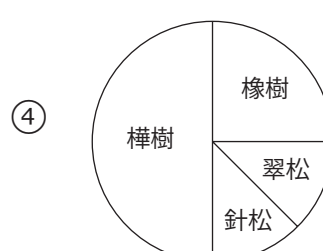
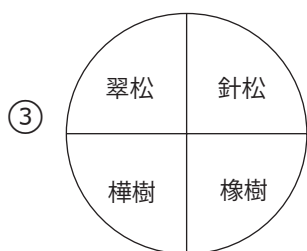
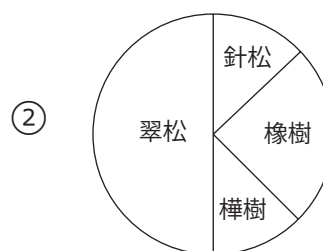
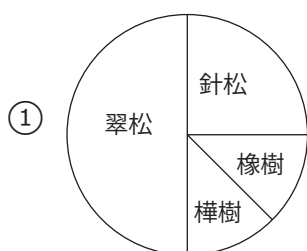
小敏請同學寫下他們各自家中的兄弟姐妹人數。上面是她收集的回覆紙條，她想要用畫記的方式來記錄所收集的結果。在表中 0 的地方她已經劃上了兩劃。

請你完成下面的表。

兄弟姐妹的人數	總數
0	//
1	
2	
3	
4	

樹木的種類	樹木的數量
翠松	200
針松	100
橡樹	50
樺樹	50

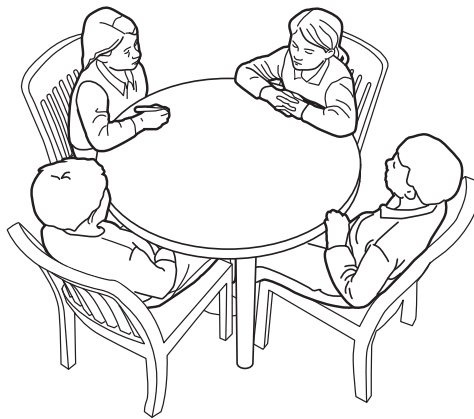
上表是表示一個公園內四種樹木的數量。請問下列哪個圓形圖能正確的表示上表所提供的數據資料？



下列哪一個是由**最大**排到**最小**？

- ① 36, 43, 66, 87
- ② 66, 43, 36, 87
- ③ 87, 66, 36, 43
- ④ 87, 66, 43, 36

M041014



一張桌子可以坐 4 個人。

如果有 28 個人，你要用什麼方法找出需要幾張桌子呢？

- ① 28 乘以 4
- ② 28 除以 4
- ③ 28 減 4
- ④ 28 加 4

M041039

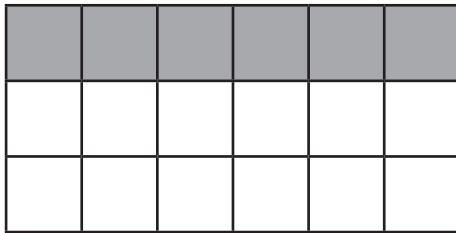
乘法：

$$53 \times 26$$

答：_____

M02_03

陰影的部份是這個長方形的幾分之幾？



- ① $\frac{1}{4}$
- ② $\frac{1}{3}$
- ③ $\frac{6}{12}$
- ④ $\frac{2}{3}$

M02_04

減法：

$$5.3 - 3.8$$

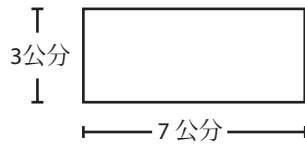
答：_____

M02_05

小傑有 10 元，午餐時間他買了 2.5 元的果汁和 3.85 元的三明治，請問小傑還剩下多少元？

- ① 3.65 元
- ② 4.75 元
- ③ 6.35 元
- ④ 16.35 元

M041094



請問這個長方形的周長是多少？

- ① 7 公分
- ② 10 公分
- ③ 20 公分
- ④ 21 公分

M041330

幾何板

說明：

在這個題目中，你會拿到一片厚紙板，它有六個拼圖，如下圖所示。將厚紙板上的六個拼圖撥開來。

若你沒有拿到厚紙板，請舉手。

4 個三角形板

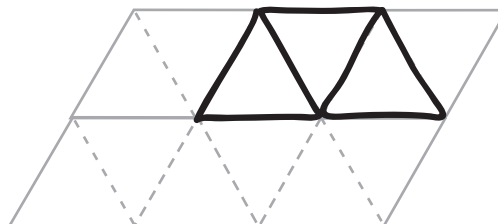


2 個梯形板



這些拼圖可以用來拼出新的圖形。這一題要拼的圖形已經幫你做好了。

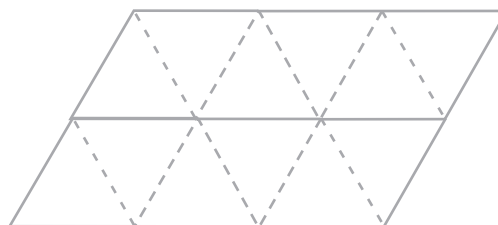
使用： 3 個三角板
做出： 1 個梯形
展示： 把拼出來的圖形畫在格子上



現在，試著做下面的題目。

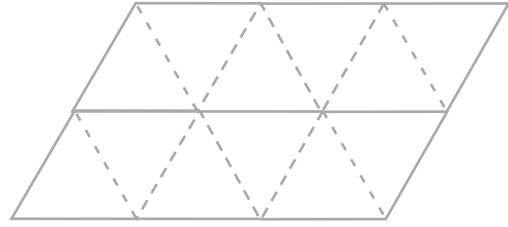
甲.

使用： 1 個三角板和
1 個梯形板
做出： 1 個四邊形
展示： 把拼出來的圖形畫在格子上



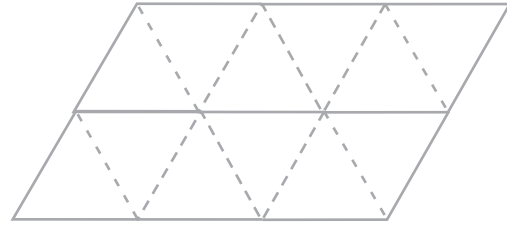
乙.

使用： 2 個梯形板
 做出： 1 個六邊形
 展示： 把拼出來的圖形畫在格子
 子上



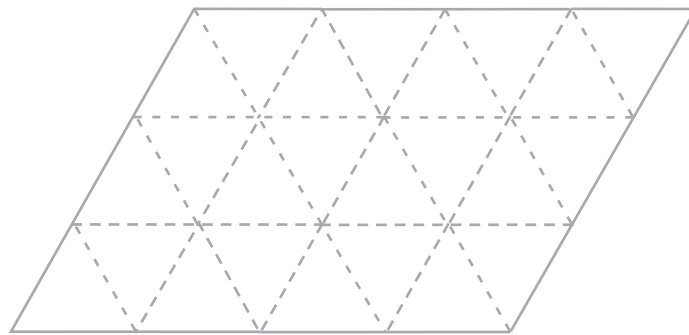
丙.

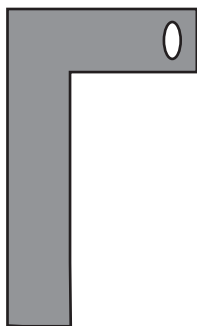
使用： 2 個梯形板
 做出： 1 個和第乙題形狀不一樣
 的六邊形
 展示： 把拼出來的圖形畫在格
 子上



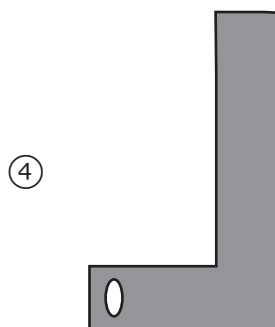
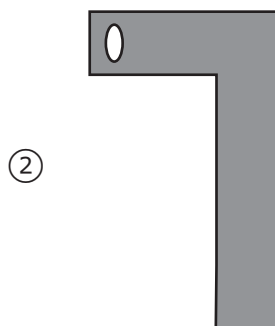
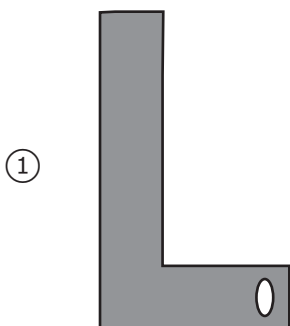
丁.

使用： 2 個三角板和
 1 個梯形板
 做出： 1 個七邊形
 展示： 把拼出來的圖形畫在格
 子上

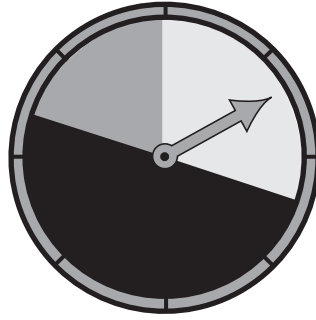




請將上方的圖形依順時針轉 90 度。下面哪一個是旋轉後的圖形？



指針可能會停在三個不同顏色的地方，下表是指針旋轉了一百次之後所得的結果。

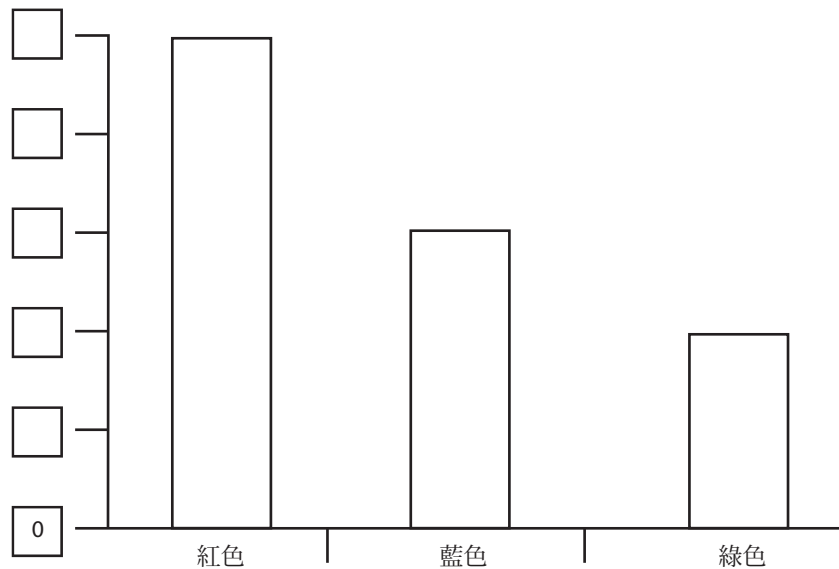


旋轉的結果

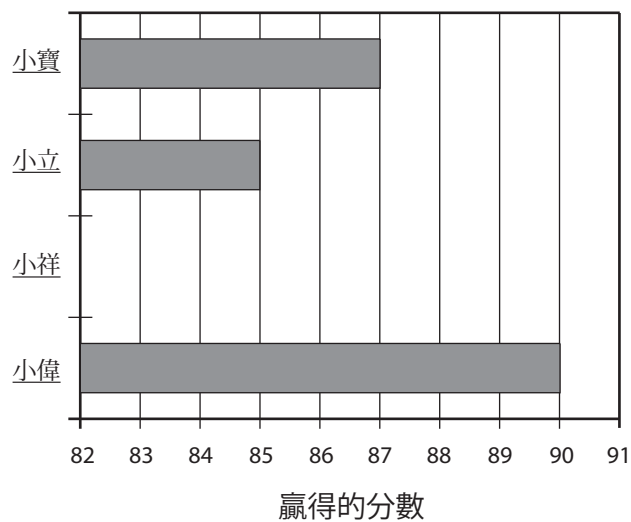
紅色	50
藍色	30
綠色	20

宛珍準備畫出下列的長條圖來表示旋轉的結果，請幫她在圖中的刻度□內填上正確的數字。

長條圖



下面的長條圖表示 4 位賽車手所贏得的分數。小偉得到第一名，小祥得到第三名，請您畫出正確的長條圖表示小祥贏得的分數。



M031235

8 個小朋友總共有 74 顆糖果。如果想讓每一個人都拿到一樣多的糖果並且沒有剩下糖果，請問最少還需要幾顆糖？

M03_01

答：_____

M031285

有兩個男生正在跑步，阿福每跑 2 公里時阿安就跑了 3 公里。結果阿福跑了 6 公里，請問阿安跑了幾公里？

M03_02

答：_____ 公里

M031050

小明用一支長 30 公分的尺量一塊黑板的長度，結果還差 6 公分就剛好量 9 次。請問黑板的長度是多少？

M03_03

- ① 264 公分
- ② 270 公分
- ③ 276 公分
- ④ 279 公分

安怡利用一個規則從美莉的數字去得到她的數字，如下表。

美莉的數字	安怡的數字
1	3
2	6
4	12
6	18

請問安怡的規則是什麼？

M031258

2, 5, 11, 23, ...

從 2 開始，請問要用什麼規則才能產生上面的數列？

- ① 將前一個數加 1 後再乘 2 即為下一個數字。
- ② 將前一個數乘 3 後再減 1 即為下一個數字。
- ③ 將前一個數乘 2 後再加 1 即為下一個數字。
- ④ 將前一個數減 1 後再乘 3 即為下一個數字。

M031334

$$64 \div \blacksquare = \blacksquare$$

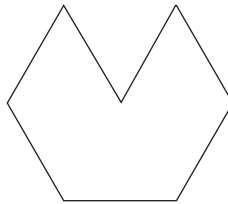
在這個算式中， $\blacksquare$ 代表同一個數。那麼 $\blacksquare$ 代表的數是多少？

- ① 4
- ② 8
- ③ 16
- ④ 32

M031255



請問需要幾個像上面這樣的三角形才能完全蓋住下面的圖形？



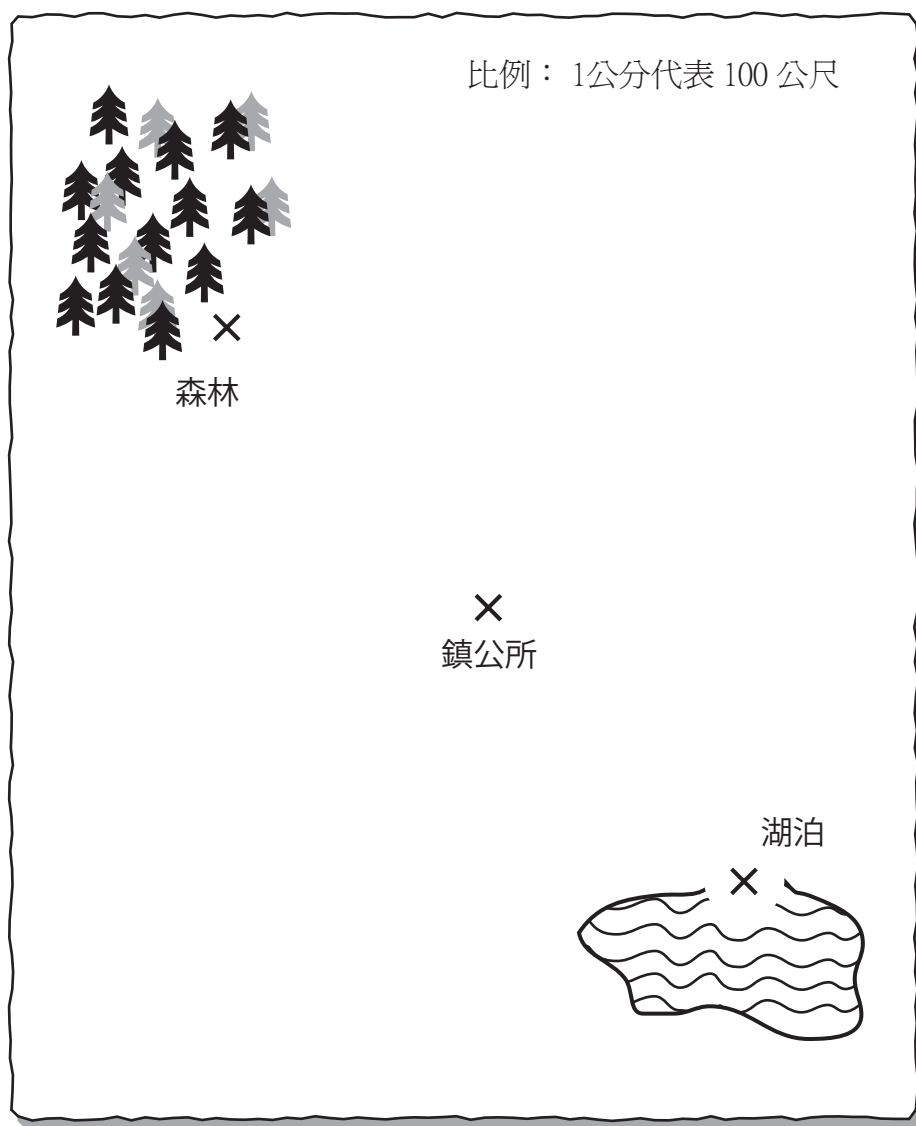
答：_____

M031041

這一大題，你會拿到一把紙板做的尺。如果你沒有拿到，請舉手。請用你的紙尺和這個地圖回答下列題組。

薔薇鎮是一個新的鄉鎮。薔薇鎮的居民正在規劃他們的新鎮。他們決定將鎮公所蓋在森林與湖泊之間的中間位置，如下圖所示。他們以 x 為測量點。

薔薇鎮



翻到下一頁的薔薇鎮題目，繼續作答。



請根據以下的條件在薔薇鎮的地圖上加上公園、圖書館、和學校。

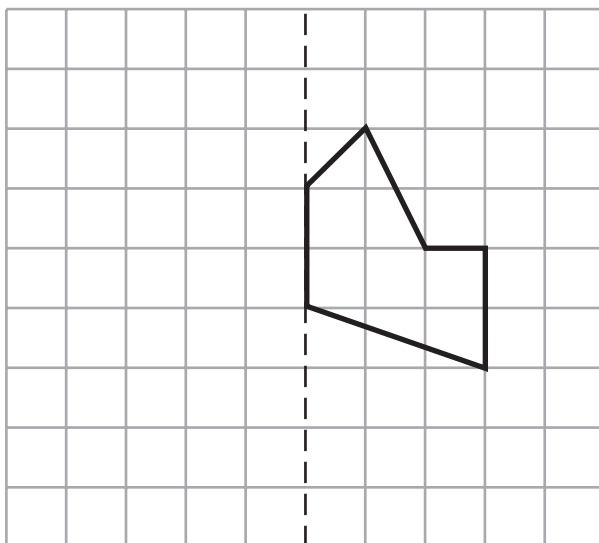
甲. **公園**是離湖泊 200 公尺的位置，這樣居民才可以方便去釣魚和游泳。
在地圖上用 x 標記出**公園**的位置，並在 x 下方寫上"**公園**"。

乙. **圖書館**至少要離鎮公所 300 公尺，但又不能超過 400 公尺的位置。在
地圖上用 x 標記出**圖書館**的位置，並在 x 下方寫上"**圖書館**"。

丙. **學校**是剛好位在公園與圖書館之間的中間距離。
地圖上用 x 標記出**學校**的位置，並在 x 下方寫上"**學校**"。



請在下面的方格上，以虛線為對稱軸，劃出此圖形的對稱圖形。



M031274

在公園裡有 5 個小孩。有一些小孩有戴帽子，有一些小孩則沒有戴帽子。



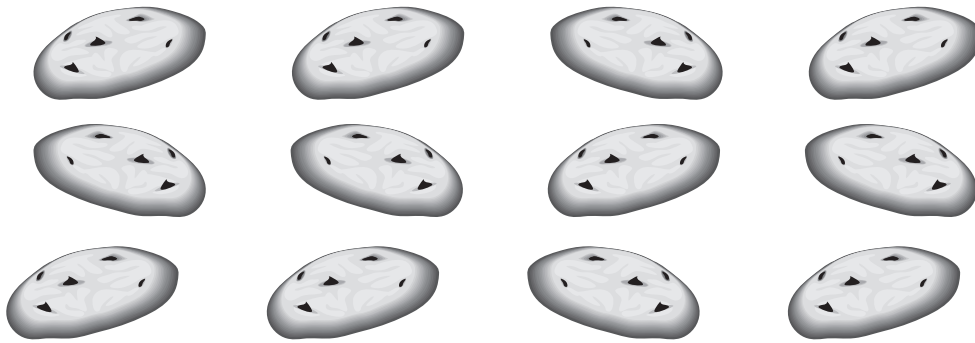
女生	男生
佳惠有戴帽子	小明有戴帽子
小佳沒有戴帽子	小華沒有戴帽子
小文沒有戴帽子	

請依據女生及男孩有戴帽子和沒有戴帽子的人數，完成下表：

	有戴帽子	沒有戴帽子
男生		
女生		

哪一個數與 3 個「一」加 2 個「十」加 4 個「百」等於以下哪一個數字？

- ① 432
- ② 423
- ③ 324
- ④ 234



上圖有 12 塊餅乾，請將這些餅乾的 $\frac{1}{3}$ 圈起來。

下面哪一個分數等於 $\frac{2}{3}$ ？

- ① $\frac{3}{4}$
- ② $\frac{4}{9}$
- ③ $\frac{4}{6}$
- ④ $\frac{3}{2}$

小傑花了零用錢的 $\frac{3}{10}$ 買了一支筆，花了零用錢的 $\frac{5}{10}$ 買了一本書。

請問他總共花掉了零用錢的幾分之幾？

答：_____

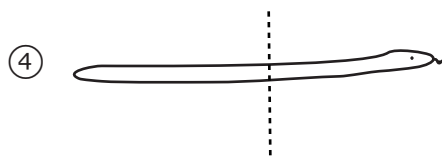
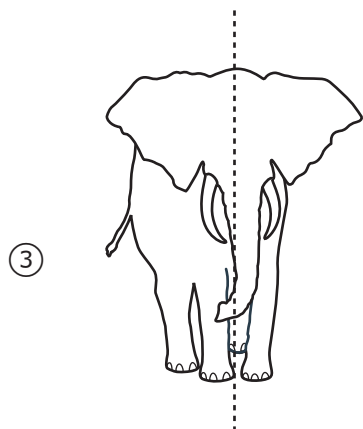
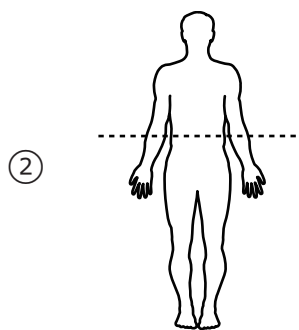
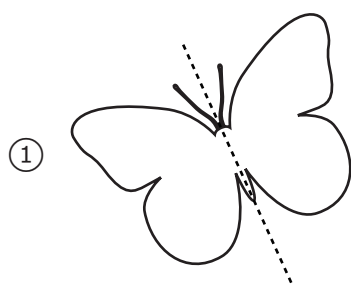
M04_04

國恩有 32 枝鉛筆，想分裝成 4 個盒子，每一個盒子裝的鉛筆一樣多。
下列哪一個算式可以告訴我們每個盒子要裝入多少枝鉛筆？

- ① $32 + 4 = \square$
- ② $32 - 4 = \square$
- ③ $32 \times 4 = \square$
- ④ $32 \div 4 = \square$

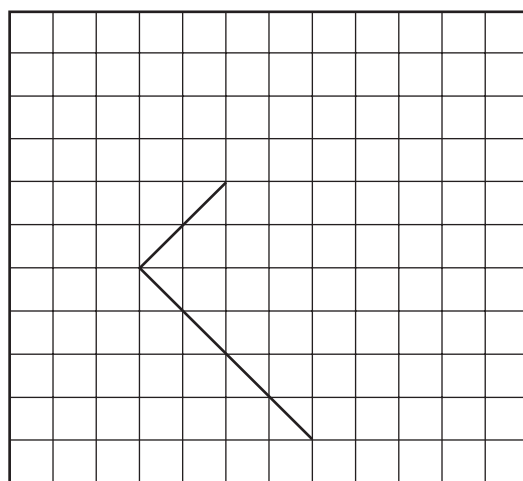
M04_05

下面哪一個圖形的虛線是這個圖形的對稱軸？

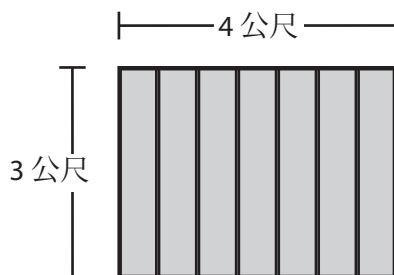


M041164

這是一個長方形的兩邊，請畫出另外兩邊。



M041146



小皮想要油漆籬笆的一個面，籬笆的長為 4 公尺，高為 3 公尺。

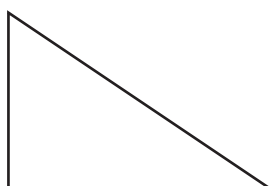
請問小皮需要漆的面積是多少？

- ① 4 平方公尺
- ② 7 平方公尺
- ③ 12 平方公尺
- ④ 14 平方公尺

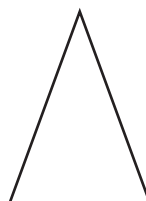
M041152

下面有兩個圖形，寫下一個它們相同的地方和一個不同的地方。

圖形 甲



圖形 乙



1. 相同的地方：

2. 不同的地方：

M041258



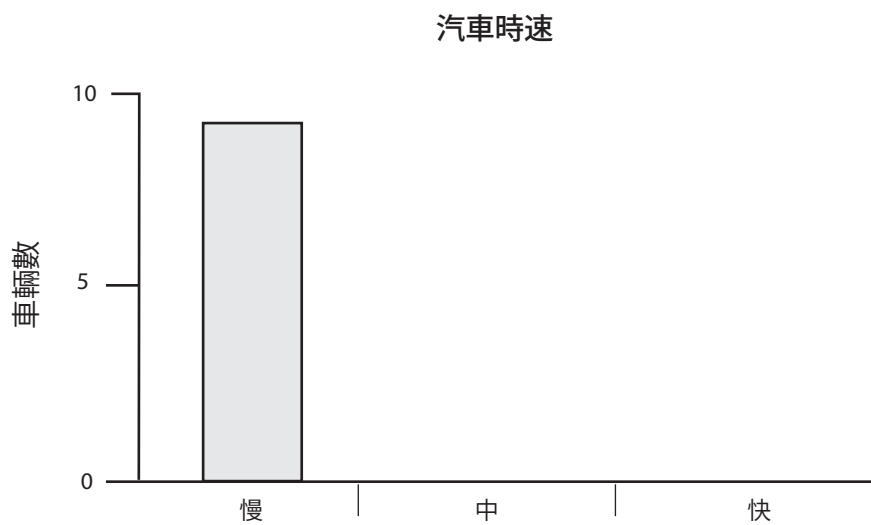
圖中的這個人的身高是 2 公尺，請問這棵樹大約有多高？

- ① 4 公尺
- ② 6 公尺
- ③ 8 公尺
- ④ 10 公尺

有一群學生正在調查關於學校車輛的開車速度資料，下表的數據是調查 20 輛車子得到的統計結果。

車輛	慢	中	快
1		X	
2	X		
3	X		
4			X
5			X
6	X		
7		X	
8		X	
9	X		
10	X		
11	X		
12		X	
13	X		
14			X
15			X
16	X		
17		X	
18	X		
19		X	
20			X

爲了讓結果更容易看懂，學生們開始將資料畫成長條圖。
請完成下面的長條圖。



下面的圖表是表示阿達每天摘下的蘋果數量。

每一個  代表 10 顆蘋果

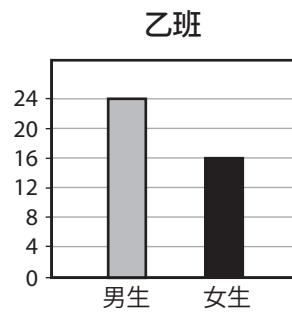
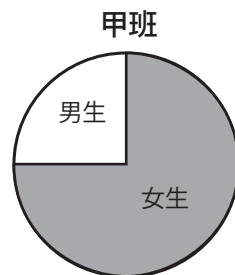
星期一	       
星期二	    
星期三	
星期四	

阿達是在星期幾摘下 5 顆蘋果？

- ① 星期一
- ② 星期二
- ③ 星期三
- ④ 星期四

M041186

甲班與乙班分別有 40 位學生。



甲班的女生比乙班多，請問多了幾位？

- ① 14
- ② 16
- ③ 24
- ④ 30

M041336

一排有 15 張椅子，9 排共有多少張椅子？下列哪一個算式可以代表椅子的總數？

- Ⓐ $15 \div 9$
- Ⓑ $15 - 9$
- Ⓒ 15×9
- Ⓓ $15 + 9$

將一條長 204 公分的繩子剪成等長的四段。每段的繩子有多長？

答：_____ 公分

$$12 \div 3 = \blacksquare \div 2$$

在這個算式中， $\blacksquare$ 代表的數字是什麼？

- Ⓐ 2
- Ⓑ 4
- Ⓒ 6
- Ⓓ 8

下面為租腳踏車的廣告文宣。

毛恩腳踏車出租行

最初的一小時收 8 zeds
每超過一小時加收 3 zeds



羅四腳踏車出租行

最初的一小時收 10 zeds
每超過一小時加收 2 zeds



甲、請依據廣告文宣中的資料，完成下表。

毛恩腳踏車出租行	
小時	租金 (zeds)
1	8
2	11
3	
4	
5	
6	

羅四腳踏車出租行	
小時	租金 (zeds)
1	10
2	12
3	
4	
5	
6	

乙、請問要租幾個小時才能讓兩家的租金相同？

答：_____

丙、如果我要租 12 個小時的腳踏車，哪一家的價錢會比較便宜？

- ① 毛恩出租行
- ② 羅四出租行
- ③ 兩家出租行都一樣
- ④ 無法比較

假日的時候，爸爸帶著他的 3 個孩子參觀市區博覽會。爸爸一個大人的門票票價是一個小孩子票價的兩倍。爸爸花 50 zeds 買了他們的 4 張票。

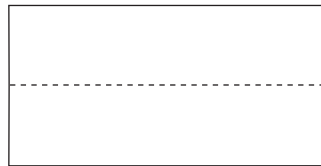
請問每位孩子的門票票價是多少？請寫出您的做法。

答：_____

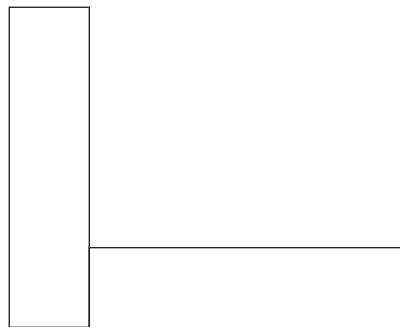
M05_05



小吉有一張長方形的紙。



她沿著虛線剪開，再拼成“L”的形狀。



下列哪一個敘述是對的？

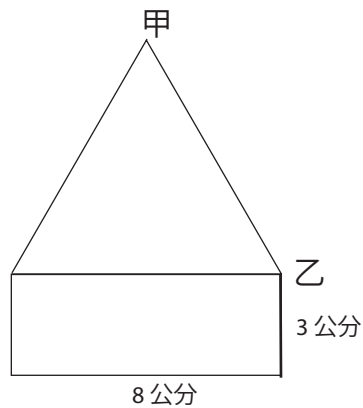
- ① “L” 形的面積比原來的長方形面積大。
- ② “L” 形的面積與原來的長方形面積相等。
- ③ “L” 形的面積比原來的長方形面積小。
- ④ 因為沒有經過測量，所以無法比較大小。

M05_06

小莉有 6 個紅色的盒子，每一個紅盒子中有 4 枝鉛筆。她還有 3 個藍色的盒子，每一個藍盒子中有 2 枝鉛筆。請問小莉共有幾枝鉛筆？

- ① 6
- ② 15
- ③ 24
- ④ 30

M031173



上面的圖形是由一個長方形和三邊等長的三角形所組成的。請問三角形的甲乙邊的長度為幾公分？

- ① 8
- ② 9
- ③ 10
- ④ 11

M031085

街名	房子間數
中華街	    
中正街	 
中山街	  
南陽街	

小真正在畫記表示某一街道上房子數的統計圖。每一個“”代表 5 間房子。在南陽街上有 20 間房子，請問小真應該在南陽街所對應的欄內要畫出幾個“”？

- ① 4
- ② 5
- ③ 15
- ④ 20

$$\frac{4}{5} - \frac{1}{5} =$$

① $\frac{3}{5}$

② $\frac{3}{10}$

③ $\frac{3}{25}$

④ 3

M031029

$$12.36 - 9.7 =$$

答：_____

M031030

下列哪個數字最接近 10 ？

- ① 0.10
- ② 9.99
- ③ 10.10
- ④ 10.90

M031332

在一個有規律的數列中，前面的四個數為

2, 4, 8, 16, ...

請問下一個數字是多少？

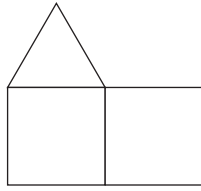
- ① 24
- ② 30
- ③ 32
- ④ 64

M031098

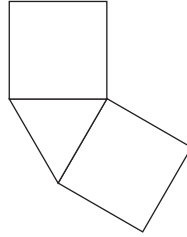
一個書架的長度為 240 公分。小可要在書架上放一些小盒子，每一個小盒子佔用了書架的 20 公分。下列哪一個算式可以表示書架可以放置的盒子數量？▲為盒子的數量。

- ① $240 - 20 = \blacktriangle$
- ② $240 \div 20 = \blacktriangle$
- ③ $240 + 20 = \blacktriangle$
- ④ $240 \times 20 = \blacktriangle$

M031254



小莉



小美

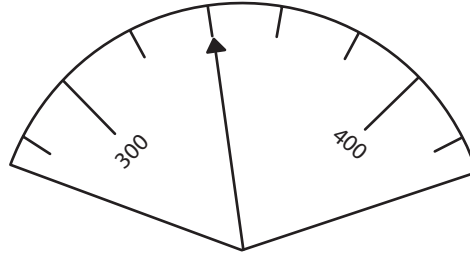


小娜

小莉、小美和小娜輪流排 3 個積木，每個人排出上面不同的形狀。下列的敘述是有關於她們所排出形狀的面積，哪一個敘述是對的？

- ① 小莉排出的形狀，其面積最大。
- ② 小美所排出的形狀，其面積最大。
- ③ 小娜所排出的形狀，其面積最大。
- ④ 三個人所排的形狀，其面積一樣大。

M031038



根據上面的刻度，指針所指的數是多少？

- ① 302
- ② 310
- ③ 320
- ④ 340

M031276

小文要烤餅乾。首先，他要先打開烤箱熱 10 分鐘，接著再花 12 分鐘烤餅乾。如果小文希望在 11:00 烤完餅乾，他最晚應該在什麼時候打開烤箱？

- ① 10:38
- ② 10:48
- ③ 10:50
- ④ 11:22

小可有許多像右圖的瓷磚：



小歐有許多像右圖的瓷磚：



小丹有許多像右圖的瓷磚：

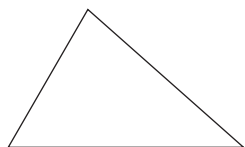


阿尼有許多像右圖的瓷磚：

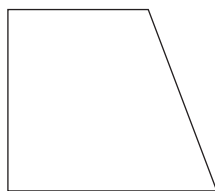


若要用瓷磚鋪教室地板，請問誰用的瓷磚數量最少？

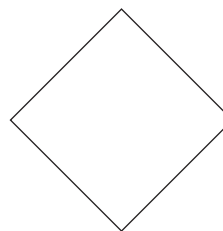
- ① 小可
- ② 小歐
- ③ 小丹
- ④ 阿尼



甲



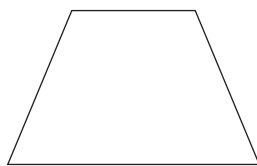
乙



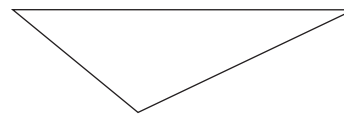
丙



丁



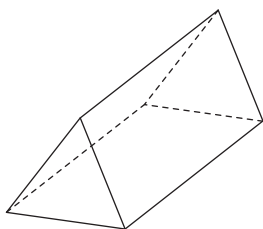
戊



己

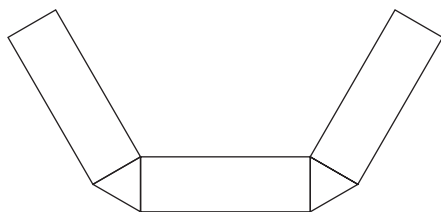
請列出所有三角形的代號。

答：_____

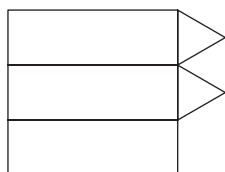


下列哪一個圖可以組合成上面的立體圖？

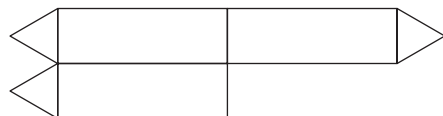
①



②

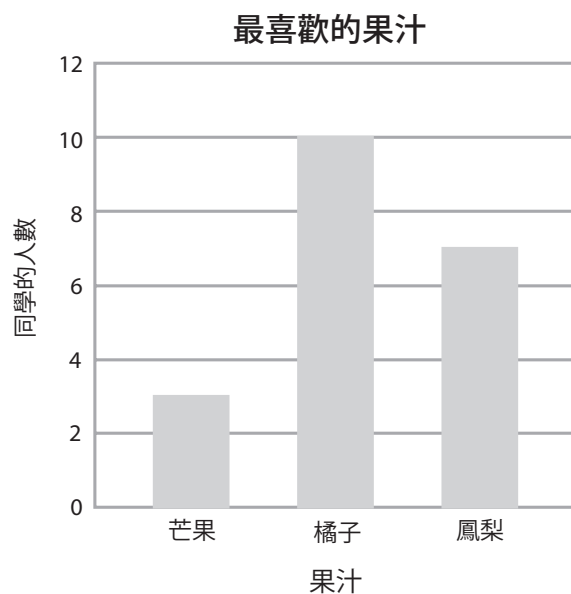


③



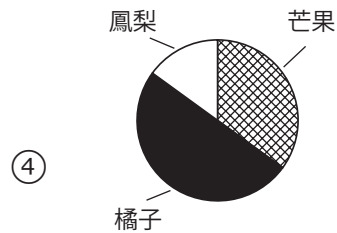
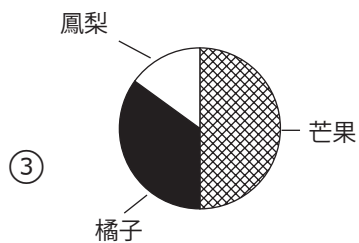
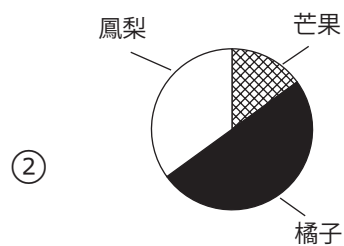
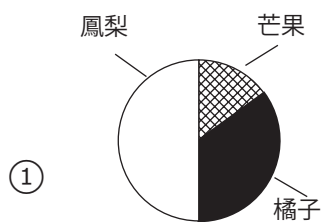
④





小琳想要調查 20 個同學最喜歡喝的果汁是橘子、芒果、還是鳳梨的口味？她將調查所得的資料製成如上面的長條圖。

她也用上圖的資料繪製成一個圓形圖。下面哪個圓形圖能正確的表示上面的資料呢？



附錄六

TIMSS 2007 八年級問卷（中文版）

識 別 資 料

學校代碼：

學校名稱：

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

學校問卷

八年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

感謝 貴校同意參與2 0 0 7 國際數學與科學教育研究趨勢之調查(Trends in International Mathematics and Science Study, TIMSS-2 0 0 7)，這是一個有關數學與自然科學學習狀況之大型國際性調查研究，全球有超過六十個國家參加。本調查研究由國際教育成就調查委員會(International Association for the Evaluation Achievement, IEA)所主辦，目的在測量學生成就的趨勢與瞭解世界各國教育制度的差異，以求提昇全球數學及自然科學之教育水準。

這份問卷主要是詢問各校校長及各處室主任一些學校的問題。由於 貴校被選為全國受測樣本之一，您回答的資料對協助了解我國的學校系統是非常重要的。

請仔細回答每一個問題，以使您提供的訊息能充分反映出 貴校的情況。在您填答的時候，有些問題也許必須查閱 貴校資料方能作答，因此您可能需要相關人員的協助來取得資料。

請在不受干擾的時間及地點完成問卷，回答此問卷大約需要三十分鐘。為了作答方便，大部份問題僅須填選即可。

當您做完這份問卷之後，請將問卷交給貴校之學校聯絡人，連同試題一併寄回國立台灣師範大學科學教育中心。謝謝您的作答。填答過程中若有疑問可以與下列人員聯絡詢問。

劉佩娟小姐	02-29316273分機314
王華鈺小姐	02-29316273分機341
任宗浩助理研究員	02-29316273分機324
洪志明教授	02-29316273分機312

1

A. 請問 貴校註冊的學生總人數有多少？

學生總人數:_____人

B. 請問 貴校註冊的國中八年級學生人數有多少？

學生總人數:_____人

2

貴校所在的鄉、鎮或市有多少居住人口？

請只勾選一項

- 超過 500,000 人 -----○
100,001 ~ 500,000 人 -----○
50,001 ~ 100,000 人 -----○
15,001 ~ 50,000 人 -----○
3,001 ~ 15,000 人 -----○
3,000 人以下 -----○

3

貴校學生中，下列背景的學生之百分比約為何？

每小題請勾選一項

- | | 0 ~ 10% | 11 ~ 25% | 26 ~ 50% | 多於 50% |
|--------------|---------|----------|----------|--------|
| a) 家境困難----- | ○ | ○ | ○ | ○ |
| b) 家境富裕----- | ○ | ○ | ○ | ○ |

4

貴校學生中，以國語作為母語之百分比約為何？

請只勾選一項

- 高於 90% -----○
76 ~ 90% -----○
50 ~ 75% -----○
低於 50%-----○

5

針對 貴校八年級學生情形回答：

A. 貴校八年級學生一年上課幾天？

_____天
(請填寫數字)

B. 貴校八年級全天課(扣除午餐休息，自修課及課後活動時間等)每天實際授課時間為多久？

_____時 _____分
(請填寫數字)

C. 貴校八年級學生每週上課幾天？

請只勾選一項

- 6 天 -----○
5 1/2 天 -----○
5 天 -----○
4 1/2 天 -----○
4 天 -----○
其他 -----○
請寫出天數 _____

6

在這個學年結束時，您以校長身份從事下列活動的時間百分比為何？

每小題請輸入百分比
總和為 100%

- a) 行政工作(如：約聘教職員、編預算、擬定行事曆、開會)----- %
- b) 教育領導
(如：發展課程及教學)----- %
- c) 教職員視察與考核----- %
- d) 教學 ----- %
- e) 公關及募款----- %
- f) 其他 ----- %
- 總共 ----- 100%

7

貴校是否要求家長參與下列活動？

每小題請勾選一項

否
是

- a) 出席特別的活動
(如：科學展覽、音樂會、運動會) ---- ○---○
- b) 為學校募款----- ○---○
- c) 擔任學校計畫、正式/非正式課程或校外旅遊活動的志工----- ○---○
- d) 督促孩子完成家庭作業----- ○---○
- e) 擔任學校委員
(如：遴聘教職員及審核學校財務) ---- ○---○

8

您對 貴校以下各項的評比如何？

每小題請勾選一項

- 很高 高 普通 低 很低
- a) 教師的工作滿意度--- ○ -- ○ -- ○ -- ○ -- ○
- b) 教師對學校課程目標的了解----- ○ -- ○ -- ○ -- ○ -- ○
- c) 教師達成學校課程進度的程度----- ○ -- ○ -- ○ -- ○ -- ○
- d) 教師對學生學習成就的期望----- ○ -- ○ -- ○ -- ○ -- ○
- e) 家長對學生學習成就的支持----- ○ -- ○ -- ○ -- ○ -- ○
- f) 家長對學校活動的參與度----- ○ -- ○ -- ○ -- ○ -- ○
- g) 學生愛惜學校資源的程度----- ○ -- ○ -- ○ -- ○ -- ○
- h) 學生力求在校有好表現的意願----- ○ -- ○ -- ○ -- ○ -- ○

9

貴校八年級的數學教學，是否採能力分班上課？

請勾選是或否 ----- 是 否

10

請問 貴校對八年級學生是否有採取下列措施？

每小題請勾選是或否

- 是 否
- a) 提供數學課外補充教學----- ○ -- ○
- b) 提供數學補救教學----- ○ -- ○

11

貴校八年級的自然科教學，是否採能力分班上課？

請勾選是或否 ----- 是 否

12

請問 貴校對八年級學生是否有採取下列措施？

每小題請勾選是或否

- 是 否
- a) 提供自然科課外補充教學----- ○ -- ○
- b) 提供自然科補救教學----- ○ -- ○

13

在過去兩年中，貴校八年級教師曾經參與下列各項以數學及科學為主的教師專業發展研習活動之百分比為何？

每小題請勾選一項

	從未	25% 或更低	26 - 50%	51 - 75%	76 - 100%
a) 支援全國性或區域性課程的實施-----	☐	☐	☐	☐	☐
b) 設計或支援校內的改進方案-----	☐	☐	☐	☐	☐
c) 加強學科知識的研習活動-----	☐	☐	☐	☐	☐
d) 加強教學技能的研習活動-----	☐	☐	☐	☐	☐
e) 資訊科技融入教學的研習-----	☐	☐	☐	☐	☐

14

貴校是否使用下列方式來評鑑八年級數學教師的教學實務？

每小題請勾選是或否

	否	是
a) 校長或資深教師之觀察-----	☐	☐
b) 由督學或校外人士之觀察-----	☐	☐
c) 學生之學習成就-----	☐	☐
d) 教師互評-----	☐	☐

15

貴校是否使用下列方式來評鑑八年級自然科教師的教學實務？

每小題請勾選是或否

	否	是
a) 校長或資深教師之觀察-----	☐	☐
b) 由督學或校外人士之觀察-----	☐	☐
c) 學生之學習成就-----	☐	☐
d) 教師互評-----	☐	☐

16

這學年，貴校對下列科目之八年級教師缺額遞補的困難程度為何？

每小題請勾選是或否

	該科並無缺額	容易遞補	有些困難	非常困難
a) 數學-----	☐	☐	☐	☐
b) 自然科-----	☐	☐	☐	☐
c) 電腦/資訊科技-----	☐	☐	☐	☐

17

貴校目前是否有任何優惠方案（如薪水、住宿、簽約獎金、小班級等）來遴聘或留住下列科目的八年級教師？

每小題請勾選是或否

	否	是
a) 數學-----	☐	☐
b) 自然科-----	☐	☐
c) 其他-----	☐	☐

貴校八年級學生出現下列問題行為的次數及其嚴重程度為何？

A. 發生頻率

每小題請各勾選一項

- | | 從未 | 很少 | 每月 | 每週 | 每天 |
|--------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 上課遲到----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| b) 無故缺席----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| c) 翹課----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| d) 服裝儀容不合規定----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| e) 擾亂教室秩序----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| f) 作弊----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| g) 說髒話----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| h) 破壞公物----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| i) 偷竊----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| j) 言語恐嚇嘲弄其他同學----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| k) 肢體傷害其他同學----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| l) 言語恐嚇污辱教職員----- | ☐ | ☐ | ☐ | ☐ | ☐ |
| m) 肢體傷害教職員----- | ☐ | ☐ | ☐ | ☐ | ☐ |

B. 問題嚴重性

每小題請各勾選一項

- | | 不成問題 | 輕微問題 | 嚴重問題 |
|-------|-----------------------|-----------------------|-----------------------|
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |
| ----- | ☐ | ☐ | ☐ |

貴校的教學，受到下列各項資源的缺乏或不足之影響為何？

每小題請勾選一項

沒有 很少 一些 很大

- a) 教材(如：教科書) ----- ○ -- ○ -- ○ -- ○
- b) 庶務預算(如：紙、筆) ---- ○ -- ○ -- ○ -- ○
- c) 學校建築及場地----- ○ -- ○ -- ○ -- ○
- d) 空調及照明系統 ----- ○ -- ○ -- ○ -- ○
- e) 教學空間(如：教室) ----- ○ -- ○ -- ○ -- ○
- f) 殘障學生所需的特殊設備-- ○ -- ○ -- ○ -- ○
- g) 數學教學用的電腦----- ○ -- ○ -- ○ -- ○
- h) 數學教學用的電腦軟體---- ○ -- ○ -- ○ -- ○
- i) 數學教學用的計算機----- ○ -- ○ -- ○ -- ○
- j) 與數學教學相關的圖書---- ○ -- ○ -- ○ -- ○
- k) 數學教學用的視聽設備---- ○ -- ○ -- ○ -- ○

每小題請勾選一項

沒有 很少 一些 很大

- l) 自然科學實驗室的設備與器材----- ○ -- ○ -- ○ -- ○
- m) 自然科學教學用的電腦---- ○ -- ○ -- ○ -- ○
- n) 自然科學教學用的電腦軟體 ○ -- ○ -- ○ -- ○
- o) 自然科學教學用的計算機 - ○ -- ○ -- ○ -- ○
- p) 與自然科學教學相關的圖書 ○ -- ○ -- ○ -- ○
- q) 自然科學教學用的視聽設備 ○ -- ○ -- ○ -- ○
- r) 師資----- ○ -- ○ -- ○ -- ○
- s) 電腦技術支援人員 ----- ○ -- ○ -- ○ -- ○

20 _____

A. 貴校有自然科學實驗室嗎？

請勾選是或否 ----- ☐ 是 ☐ 否

B. 當學生進行科學實驗時，教師是否能獲得支援？

請勾選是或否 ----- ☐ 是 ☐ 否

21 _____

A. 貴校在教學時提供八年級學生使用的電腦總數有幾台？

_____ 台

如果填 0，請直接回答第 22 題



B. 這些供教學用的電腦中，有多少台可以上網？
（包括e-mail或WWW）

請只勾選一項

全部都可以 ----- ☐
大部份可以 ----- ☐
有一些可以 ----- ☐
都不可以 ----- ☐

22 _____

貴校是否有人可支援教師在教學與學習時使用資訊科技？

請勾選是或否 ----- ☐ 是 ☐ 否

感謝您撥冗填答



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

學校問卷

八年級

識 別 資 料

學校代碼：

學生代碼：

學校名稱：

學生姓名：

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

學生問卷

八年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

這份問卷要問你一些問題，有些問題是問關於你的一些事情，有些問題則是問你的看法。

請仔細讀每一個問題，並確實作答。如果你有不明白的地方或是不知道怎麼回答，可以舉手問老師。

有些問題的後面跟著幾個答案，這些答案是以加圈圈的數字來表示，回答問題時，請把你所選擇的答案圈起來，就像下面的例題一、二、三。

例題 1

你上學嗎？

請只圈選一項

是----- ①
否----- ②

例題 2

你常常做下列的事情嗎？

每項圈選一個

	每天都有 ↓	每週至少 一次 ↓	每月 1~2次 ↓	每年 數次 ↓	從來 沒有 ↓
a) 聽音樂 -----	①	②	③	④	⑤
b) 和朋友聊天 -----	①	②	③	④	⑤
c) 運動 -----	①	②	③	④	⑤

例題 3

你對下列的說法同不同意？

每項圈選一個

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| a) 看電影很有趣 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| b) 我喜歡吃冰淇淋 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

請仔細讀每一個問題後，再選出你認為最合適的答案。如果你要更改答案，請先在原來的答案上打「X」，再重新選擇。如果你有任何問題或不知道如何作答，請舉手問老師。

謝謝你的作答

關於你自己

1

你是民國幾年幾月出生的？

A. 請把你出生時的年份圈起來

B. 請把你出生時的月份圈起來

年

78 - ①

79 - ②

80 - ③

81 - ④

82 - ⑤

83 - ⑥

84 - ⑦

85 - ⑧

其他 ⑨

月

一月 - ○

二月 - ○

三月 - ○

四月 - ○

五月 - ○

六月 - ○

七月 - ○

八月 - ○

九月 - ○

十月 - ○

十一月 - ○

十二月 - ○

2

你是男生還是女生？

只圈選一個

女生----- ①

男生----- ②

3

你在家常說國語嗎？

只圈選一個

- 總是----- ①
常常----- ②
有時----- ③
從不----- ④

4

你家大約有多少本書？（雜誌、報紙和學校的課本不算）

只圈選一個

- 沒有或很少
(0-10 本書) ----- ①
可以放滿一排
(11-25 本書) ----- ②
可以放滿一個書架
(26-100 本書) ----- ③
可以放滿兩個書架
(101-200 本書) ----- ④
可以放滿三個或三個以上的書架
(200本書以上)----- ⑤

關於你自己（繼續）

5

你家有下面這些東西嗎？

每小題只圈選一個

是



否



- a) 計算機 ----- ① ----- ②
- b) 電腦
（不包括任天堂和電視/電腦遊樂器） ① ----- ②
- c) 你個人專用的書桌或桌子 ----- ① ----- ②
- d) 字典 ----- ① ----- ②
- e) 連接網路 ----- ① ----- ②
- f) 學習數學的光碟、軟體、錄影帶 -- ① ----- ②
- g) 學習數學的參考書 ----- ① ----- ②
- h) 學習自然科的光碟、軟體、錄影帶 ① ----- ②
- i) 學習自然科的參考書 ----- ① ----- ②

A. 你的母親(或繼母、養母、女性監護人等)的最高學歷為何？*只圈選一項*

- 小學肄業或沒有上過學----- ①
國小畢業----- ②
國中畢業----- ③
高中/職畢業----- ④
五專或二專畢業----- ⑤
二技或四技畢業----- ⑥
大學畢業----- ⑦
碩士以上學位----- ⑧
我不知道----- ⑨

B. 你的父親(或繼父、養父、男性監護人等)的最高學歷為何？*只圈選一項*

- 小學肄業或沒有上過學----- ①
國小畢業----- ②
國中畢業----- ③
高中/職畢業----- ④
五專或二專畢業----- ⑤
二技、四技畢業----- ⑥
大學畢業----- ⑦
碩士以上學位----- ⑧
我不知道----- ⑨

關於你自己（繼續）

7

你期望自己的最高教育程度為何？

只圈選一項

高中/職畢業 ----- ①

五專或二專畢業 ----- ②

二技、四技畢業 ----- ③

大學畢業 ----- ④

碩士或以上學位 ----- ⑤

我不知道 ----- ⑥

8

你對下列學習數學的說法同不同意？

每小題圈選一項

- | | 很同意
↓ | 有點同意
↓ | 不太同意
↓ | 很不同意
↓ |
|--------------------------------|----------|-----------|-----------|-----------|
| a) 我在數學科的表現通常不錯 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| b) 我希望在學校多上一些數學課 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| c) 相對於許多班上其他的同學，我覺得數學比較難 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| d) 我喜歡學數學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| e) 數學不是我擅長的科目之一 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| f) 與數學有關的事我學得很快 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| g) 數學很無趣 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| h) 我喜歡數學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

9

你對下面各項有關數學的敘述同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|------------------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| a) 我認為學數學對我的日常生活有幫助 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| b) 我需要用數學去學習其他學科 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| c) 我需要學好數學以進入我心目中理想的學校 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| d) 我需要把數學學好才能得到我想要的職業 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

你上數學課時，常做下面的這些事情嗎？

每小題圈選一項

	幾乎 每節課	約有一 半的課	有些課	從來 沒有過
	↓	↓	↓	↓
a) 不用計算機來做加減乘除的運算 -----	①-----	②-----	③-----	④-----
b) 作小數及分數的計算 -----	①-----	②-----	③-----	④-----
c) 解有關幾何圖形、直線和角的問題 -----	①-----	②-----	③-----	④-----
d) 解讀圖表中的資料 -----	①-----	②-----	③-----	④-----
e) 寫出方程式和函數來表示關係 -----	①-----	②-----	③-----	④-----
f) 背誦公式和過程 -----	①-----	②-----	③-----	④-----
g) 解釋自己的答案 -----	①-----	②-----	③-----	④-----
h) 將所學的數學聯結到日常生活上 -----	①-----	②-----	③-----	④-----
i) 為了解一些複雜的問題時，我們自己決定解題 步驟 -----	①-----	②-----	③-----	④-----
j) 檢討家庭作業 -----	①-----	②-----	③-----	④-----
k) 聽老師講課 -----	①-----	②-----	③-----	④-----
l) 自己做數學題目 -----	①-----	②-----	③-----	④-----
m) 小組共同合作 -----	①-----	②-----	③-----	④-----
n) 在課堂上就開始做家庭作業 -----	①-----	②-----	③-----	④-----
o) 平時考或測驗 -----	①-----	②-----	③-----	④-----
p) 使用計算機 -----	①-----	②-----	③-----	④-----
q) 使用電腦 -----	①-----	②-----	③-----	④-----

你對下列學習自然科學的說法同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|----------------------------------|---------|---------|---------|------|
| | ↓ | ↓ | ↓ | ↓ |
| a) 我在自然科學方面的表現通常不錯 | ① ----- | ② ----- | ③ ----- | ④ |
| b) 我希望在學校多上一些自然科學課 | ① ----- | ② ----- | ③ ----- | ④ |
| c) 相對於許多班上其他的同學，我覺得自然科學比較難 ----- | ① ----- | ② ----- | ③ ----- | ④ |
| d) 我喜歡學自然科學 ----- | ① ----- | ② ----- | ③ ----- | ④ |
| e) 自然科學不是我擅長的科目之一 -- | ① ----- | ② ----- | ③ ----- | ④ |
| f) 與自然科學有關的事我學得很快 -- | ① ----- | ② ----- | ③ ----- | ④ |
| g) 自然科學很無趣 ----- | ① ----- | ② ----- | ③ ----- | ④ |
| h) 我喜歡自然科學 ----- | ① ----- | ② ----- | ③ ----- | ④ |

你對下面各項有關自然科學的敘述同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|--------------------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| a) 我認為學自然科學對我的日常生活有幫助 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| b) 我需要用自然科學去學習其他學科 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| c) 我需要學好自然科學以進入我心目中理想的學校 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| d) 我需要把自然科學學好才能得到我想要的職業 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

13

你上自然課時，常做下面的事情嗎？

每小題圈選一項

	幾乎 每節課 ↓	約有一 半的課 ↓	有些課 ↓	從來 沒有過 ↓
a) 觀察並做紀錄 -----	①-----	②-----	③-----	④-----
b) 看老師示範實驗或演示探究活動 -----	①-----	②-----	③-----	④-----
c) 設計及規劃實驗或探討活動 -----	①-----	②-----	③-----	④-----
d) 做實驗或進行探究 -----	①-----	②-----	③-----	④-----
e) 以小組方式做實驗或進行探究 -----	①-----	②-----	③-----	④-----
f) 閱讀我們的學科教科書和其他教材相關資料 -----	①-----	②-----	③-----	④-----
g) 背誦科學知識和原理 -----	①-----	②-----	③-----	④-----
h) 用科學公式和定律解題 -----	①-----	②-----	③-----	④-----
i) 解釋學習和研究的結果 -----	①-----	②-----	③-----	④-----
j) 將所學的自然科學連結到日常生活上 -----	①-----	②-----	③-----	④-----
k) 檢討家庭作業 -----	①-----	②-----	③-----	④-----
l) 聽老師講課 -----	①-----	②-----	③-----	④-----
m) 自己做自然科學題目 -----	①-----	②-----	③-----	④-----
n) 在課堂上就開始做家庭作業 -----	①-----	②-----	③-----	④-----
o) 平時考或測驗 -----	①-----	②-----	③-----	④-----
p) 使用電腦 -----	①-----	②-----	③-----	④-----

14

A. 你使用過電腦嗎？（不包括電視/電腦遊樂器）

是 否
↓ ↓

請圈選是或否 ----- ① ----- ②

如果選答「否」，請直接回答第15題



B. 你在下列哪些地方使用過電腦？

每小題圈選一項

是 否
↓ ↓

a) 家裡 ----- ① ----- ②

b) 學校 ----- ① ----- ②

c) 其他地方 (例如：圖書館、朋友家
裡、網咖等) ----- ① ----- ②

C. 你常用電腦做學校功課嗎(在校內或校外)？

每小題圈選一項

每天 每週
 至少
 一次
↓ ↓ 每月 每年 從來
 1~2次 數次 沒有過
↓ ↓ ↓ ↓ ↓

a) 數學科 ----- ① ----- ② ----- ③ ----- ④ ----- ⑤

b) 自然科 ----- ① ----- ② ----- ③ ----- ④ ----- ⑤

你的學校

15

你對下面對學校的說法同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|----------------------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| a) 我喜歡待在學校 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| b) 我認為學校裡的學生都試著盡力
做好 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| c) 我認為學校裡的老師都希望學生
能盡力做好 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

16

上個月你在學校有發生過下面的事情嗎？

每小題圈選一項

- | | 是 | 否 |
|------------------------|---------|---------|
| | ↓ | ↓ |
| a) 我有東西被偷了 ----- | ① ----- | ② ----- |
| b) 有同學傷害我(如：推撞、踢打) --- | ① ----- | ② ----- |
| c) 有同學叫我做我不願意做的事 ---- | ① ----- | ② ----- |
| d) 有同學嘲笑我或戲弄我 ----- | ① ----- | ② ----- |
| e) 同學不讓我參加他們的活動 ----- | ① ----- | ② ----- |

校外活動

17

在上學的日子，你每天在上學前或放學後通常會花多少時間做下面的事情？

每小題圈選一項

	沒有 ↓	1小時 以下 ↓	1-2 小時 ↓	2~4 小時 ↓	4小時 以上 ↓
a) 看電視、錄影帶 -----	① -----	② -----	③ -----	④ -----	⑤ -----
b) 玩電腦、電動遊戲 -----	① -----	② -----	③ -----	④ -----	⑤ -----
c) 和朋友玩耍或聊天 -----	① -----	② -----	③ -----	④ -----	⑤ -----
d) 做家事 -----	① -----	② -----	③ -----	④ -----	⑤ -----
e) 打工賺錢 -----	① -----	② -----	③ -----	④ -----	⑤ -----
f) 運動 -----	① -----	② -----	③ -----	④ -----	⑤ -----
g) 看自己喜歡看的書 -----	① -----	② -----	③ -----	④ -----	⑤ -----
h) 上網 -----	① -----	② -----	③ -----	④ -----	⑤ -----
i) 做家庭作業 -----	① -----	② -----	③ -----	④ -----	⑤ -----

家庭作業

18

A. 你的數學老師多常指定家庭作業？

只圈選一項

- 每天 ----- ①
- 一週 3~4 次 ----- ②
- 一週 1~2 次 ----- ③
- 一週不到 1 次 ----- ④
- 從來沒有做數學作業 ----- ⑤

如果勾選「從來沒有做數學作業」，請直接寫第19題



B. 當數學老師指定家庭作業時，通常你每次花大約多少時間做數學作業？

只圈選一項

- 0 分鐘 ----- ①
- 1~15 分鐘 ----- ②
- 16~30 分鐘 ----- ③
- 31~60 分鐘 ----- ④
- 61~90 分鐘 ----- ⑤
- 超過 90 分鐘 ----- ⑥

A. 你的老師多常指定自然科作業？

只圈選一項

- 每天----- ①
一週 3~4 次----- ②
一週 1~2 次----- ③
一週不到 1 次----- ④
從來沒有做自然科作業----- ⑤

如果勾選「從來沒有做自然科作業」，請直接寫第20題



B. 當老師指定自然科作業時，通常你每次花多少時間做？

只圈選一項

- 0 分鐘----- ①
1~15 分鐘----- ②
16~30 分鐘----- ③
31~60 分鐘----- ④
61~90 分鐘----- ⑤
超過 90 分鐘----- ⑥

更多關於你自己

20

A. 你的母親(或繼母、養母、女性監護人等)是在中華民國出生的嗎？

是 否
↓ ↓

請圈選是或否 ----- ① ----- ②

B. 你的父親(或繼父、養父、男性監護人等)是在中華民國出生的嗎？

是 否
↓ ↓

請圈選是或否 ----- ① ----- ②

21

A. 你是在中華民國出生的嗎？

是 否
↓ ↓

請圈選是或否 ----- ① ----- ②

如果圈選「是」，則你已完成此問卷

B. 如果你不是在中華民國出生的，你是幾歲時回到國內的？

只圈選一項

滿 10 歲以後 ----- ①

5~10 歲之間 ----- ②

未到 5 歲 ----- ③

謝謝你填答這份問卷



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

學生問卷

八年級

識別資料

教師姓名: _____

班級名稱: _____

教師代碼: _____ 教師連結 # _____

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

數學教師問卷

八年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

感謝 貴校已同意參與2007國際數學與科學教育成就趨勢調查(Trends in International Mathematics and Science Study, TIMSS 2007)，這是一個有關數學與自然科學學習狀況之大型國際調查，全球有超過六十個國家參加。本教育研究專案由國際教育成就調查委員會(International Association for the Evaluation Achievement, IEA)所主辦，目的在測量學生成就的趨勢與瞭解世界各國教育制度的差異，以求提昇全球數學及自然科學之教育水準。

本研究從我國國中八年級學生樣本中抽出部分學生來填寫T I M S S 的測驗卷。這份問卷主要是為受測班級學生的數學科教師所設計，問卷內容包括教師的學經歷背景，教學實務和教授數學的態度。由於您任課的班級為全國受測樣本之一，您回答的資料對協助了解我國的數學教育是非常重要的。

此問卷中有些問題提及受測班級，指的是此問卷封面上所標示之班級，也就是接受TIMSS 2007測驗的班級。請仔細回答每一個問題，以使您提供的訊息能充分反映出實際狀況。

請在不受干擾的時間及地點完成問卷，回答此問卷大約需要四十五分鐘。為了作答方便，大部份問題僅須填選即可。

當您做完這份問卷之後，請將問卷交給貴校聯絡人，連同試題一併寄回國立台灣師範大學科學教育中心。謝謝您的作答。填答過程中有疑問可以與下列人員聯絡詢問。

劉佩娟小姐	02-29316273分機314
王華鈺小姐	02-29316273分機341
任宗浩助理研究員	02-29316273分機324
洪志明教授	02-29316273分機312

1

您今年幾歲？

請只勾選一項

- 25歲以下 -----○
 25 - 29歲 -----○
 30 - 39歲 -----○
 40 - 49歲 -----○
 50 - 59歲 -----○
 60 歲以上 -----○

2

您的性別為何？

請只勾選一項

- 女性 -----○
 男性 -----○

3

到本學年結束時，您將有幾年的教學年資？

（請填寫數字）年

4

您的最高學歷為何？

請只勾選一項

- 高中/職肄業 -----○
 師範學校/高中/職畢業 -----○
 師專或其他五專、二專畢業 -----○
 二技或四技畢業 -----○
 大學 -----○
 碩士或博士學位 -----○

5

您在中學以後所接受的教育，是主修哪些科或哪些領域？

每小題請勾選是或否

- | | | |
|----------------|---|---|
| | 是 | 否 |
| a) 數學 -----○ | 是 | 否 |
| b) 數學教育 -----○ | 是 | 否 |
| c) 自然科學 -----○ | 是 | 否 |
| d) 科學教育 -----○ | 是 | 否 |
| e) 教育 -----○ | 是 | 否 |
| f) 其他 -----○ | 是 | 否 |

6

您有教師證嗎？

- 請勾選是或否 -----○

你覺得能勝任下列各主題的教學嗎？

每小題請勾選一項

不能勝任
尚能勝任
非常勝任
無法回答

A. 數

- a) 用正整數計算，估計或近似值 ----- ○ -- ○ -- ○ -- ○
- b) 用文字、數字或模型（包括數線）來表示小數和分數 ----- ○ -- ○ -- ○ -- ○
- c) 分數和小數計算 ----- ○ -- ○ -- ○ -- ○
- d) 用整數表示、比較、排序和計算 ----- ○ -- ○ -- ○ -- ○
- e) 牽涉百分比和比例的解題 ----- ○ -- ○ -- ○ -- ○

B. 代數

- a) 數、代數、幾何規律或數列（延伸、缺項、將規律一般化） ----- ○ -- ○ -- ○ -- ○
- b) 簡化和計算代數式 ----- ○ -- ○ -- ○ -- ○
- c) 簡單的一元一次方程式、不等式及二元一次聯立方程式 ----- ○ -- ○ -- ○ -- ○
- d) 函數的各種等價表徵，如：數對、表格、圖形、文字或方程式 ----- ○ -- ○ -- ○ -- ○

C. 幾何

- a) 角和幾何圖形的性質（三角形、四邊形和常見的多邊形） ----- ○ -- ○ -- ○ -- ○
- b) 全等圖形和相似三角形 ----- ○ -- ○ -- ○ -- ○
- c) 立體圖與其平面圖表示法的關係 ----- ○ -- ○ -- ○ -- ○
- d) 使用適當的公式測量矩形周長、圓周長、圓面積、表面積和體積 ----- ○ -- ○ -- ○ -- ○
- e) 座標平面—數對、方程式、截距、交點與斜率 ----- ○ -- ○ -- ○ -- ○
- f) 平移、鏡射、旋轉 ----- ○ -- ○ -- ○ -- ○

D. 數據與機率

- a) 用表、統計圖表、長條圖、圓形比例圖、曲線圖讀取和展示數據 ----- ○ -- ○ -- ○ -- ○
- b) 說明數據集合（如下結論、作預測、內插和外推） ----- ○ -- ○ -- ○ -- ○
- c) 判斷、預測和決定可能結果的機率 ----- ○ -- ○ -- ○ -- ○

8

您多常與其他教師有如下的互動？

- 每小題請勾選一項
- 每天或幾乎每天
每週1~3次
每月2~3次
從不或幾乎沒有
- a) 討論某個概念的教學方式-- ○ -- ○ -- ○ -- ○
b) 一起準備教材----- ○ -- ○ -- ○ -- ○
c) 觀摩其他教師上課----- ○ -- ○ -- ○ -- ○
d) 其他教師到**我**班上
作非正式的觀摩----- ○ -- ○ -- ○ -- ○

9

過去兩年中，您是否曾參與和下列各項相關的研習活動？

- 每小題請勾選一項
- 否
是
- a) 數學學科內容----- ○ -- ○
b) 數學教材教法----- ○ -- ○
c) 數學課程 ----- ○ -- ○
d) 將資訊科技融入數學課程中 ----- ○ -- ○
e) 增進學生批判性思考或解題技能 ---- ○ -- ○
f) 數學評量 ----- ○ -- ○

10

有關貴校，您對下列各項同意的程度為何？

- 每小題請勾選一項
- 非常不同意
不同意
同意
非常同意
- a) 本校附近治安良好----- ○ -- ○ -- ○ -- ○
b) 我覺得在學校很安全----- ○ -- ○ -- ○ -- ○
c) 本校有完備的保全措施---- ○ -- ○ -- ○ -- ○

11

有關貴校，下列問題的嚴重性為何？

- 每小題請勾選一項
- 很嚴重的問題
小問題
不是問題
- a) 本校建築物該大幅整修----- ○ -- ○ -- ○
b) 教室太擁擠 ----- ○ -- ○ -- ○
c) 教師在教室以外無適當
的工作空間 ----- ○ -- ○ -- ○

12

您對貴校以下各項的評比為何？

每小題請勾選一項

- 很高 高 普通 低 很低
- a) 教師的工作滿意度--- ○ -- ○ -- ○ -- ○ -- ○
- b) 教師對學校課程目標的了解----- ○ -- ○ -- ○ -- ○ -- ○
- c) 教師達成學校課程進度的程度----- ○ -- ○ -- ○ -- ○ -- ○
- d) 教師對學生學習成就的期望----- ○ -- ○ -- ○ -- ○ -- ○
- e) 家長對學生學習成就的支持----- ○ -- ○ -- ○ -- ○ -- ○
- f) 家長對學校活動的參與度----- ○ -- ○ -- ○ -- ○ -- ○
- g) 學生愛惜學校資源的程度----- ○ -- ○ -- ○ -- ○ -- ○
- h) 學生力求在校有好表現的意願----- ○ -- ○ -- ○ -- ○ -- ○

受測班級

以下是有關受測班級（即標示在本問卷封面右上角之班級）的問題，這個班級也是 貴校參與TIMSS-2007研究計畫的受試樣本。

13

該班共有多少位學生？

_____位（請填寫人數）

14

您每週上該班數學課的時間為幾分鐘？

_____分鐘（請填寫數字）

15

A. 您上該班數學課時，是否有使用教科書？

請勾選是或否 ----- ☐ 是 ☐ 否

如果選答「否」，請跳至第16題

B. 在該班數學教學過程中，您如何使用教科書？

請只勾選一項

作為主要教學依據 ----- ☐

作為補充教材 ----- ☐

16

受測班級每週上數學課時，學生進行下列各項活動所佔時間的百分比為何？

每小題請填寫百分比
各項百分比的總和為100%

- a) 檢討家庭作業 ----- %
- b) 聽老師講課 ----- %
- c) 在您的指導下進行解題 ----- %
- d) 在沒有您的指導下自行解題 ----- %
- e) 聽您重新講解及澄清教學內容或步驟 ----- %
- f) 進行測驗或小考 ----- %
- g) 與教學內容/目的無關的教室管理活動
（如處理干擾或維持秩序） ----- %
- h) 其他學生活動 ----- %
- 總和 ----- 100%

17

上該班數學課時，您要求學生進行下列活動的情況為何？

每小題請勾選一項

- 從來沒有
有些課
約一半的課
每節或幾乎每節課
- a) 不用計算機來做四則運算-- ○ -- ○ -- ○ -- ○
 - b) 做分數及小數的題目----- ○ -- ○ -- ○ -- ○
 - c) 用圖形、直線、角性質的知識解題 ----- ○ -- ○ -- ○ -- ○
 - d) 解釋圖表的資料----- ○ -- ○ -- ○ -- ○
 - e) 寫出方程式和函數來表示關係----- ○ -- ○ -- ○ -- ○
 - f) 背誦公式與過程----- ○ -- ○ -- ○ -- ○
 - g) 應用事實、概念和過程解一般問題----- ○ -- ○ -- ○ -- ○
 - h) 解釋答案----- ○ -- ○ -- ○ -- ○
 - i) 將所學的數學連結到日常生活上----- ○ -- ○ -- ○ -- ○
 - j) 在解一些複雜的問題時，讓學生自己決定解題步驟----- ○ -- ○ -- ○ -- ○
 - k) 做那些沒有顯而易見解法的問題----- ○ -- ○ -- ○ -- ○
 - l) 小組討論----- ○ -- ○ -- ○ -- ○

18

依您的看法，下列各項因素影響您教導該班的程度為何？

每小題請勾選一項

- 無法回答
沒有影響
少許影響
有些影響
很有影響
- 學生**
- a) 學生的學業能力不同- ○ -- ○ -- ○ -- ○
 - b) 學生的背景差異很大（如：家庭經濟、語言等）----- ○ -- ○ -- ○ -- ○
 - c) 有特殊需求的學生（如：聽障、視障、語障、肢體障礙、心智或情緒/心理障礙）----- ○ -- ○ -- ○ -- ○
 - d) 學生缺乏學習興趣--- ○ -- ○ -- ○ -- ○
 - e) 搗亂的學生----- ○ -- ○ -- ○ -- ○
- 資源**
- f) 電腦硬體不足----- ○ -- ○ -- ○ -- ○
 - g) 電腦軟體不足----- ○ -- ○ -- ○ -- ○
 - h) 使用電腦的支援不足- ○ -- ○ -- ○ -- ○
 - i) 學生的教科書不足--- ○ -- ○ -- ○ -- ○
 - j) 學生用的其他教學設備不足----- ○ -- ○ -- ○ -- ○
 - k) 教師演示教學或其他活動中使用的設備不足- ○ -- ○ -- ○ -- ○
 - l) 學校硬體設施不完備- ○ -- ○ -- ○ -- ○
 - m) 學生/老師的比值高 -- ○ -- ○ -- ○ -- ○

至本學年結束，您在該班教導下列各項數學領域
所佔的時間百分比大約為何？

(每小題請填寫百分比，總和為100%)

- a) 數(如：正整數、分數、小數、比率、比例、百分比) ----- %
- b) 代數(如：規律、方程式、公式、關係式) ----- %
- c) 幾何(如：直線及角、形狀、全等與相似、空間關係、對稱與轉換) ----- %
- d) 數據與機率(如數據的讀取、組織和表示，解釋數據和機率) ----- %
- e) 其他，請說明：
----- %

總和 ----- 100%



下列為TIMSS數學測驗的主題。請選擇一項最能描述該班在何時上這些數學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”。

每小題請勾選一項

尚未教過或僅簡單介紹
大部分在今年教過
大部分以前教過

A. 數

- | | | | | |
|---------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 正整數，包括位值、因數分解及四則運算----- | ☐ | ☐ | ☐ | ☐ |
| b) 正整數的計算、估算或近似值----- | ☐ | ☐ | ☐ | ☐ |
| c) 分數，包括等值分數和分數排序----- | ☐ | ☐ | ☐ | ☐ |
| d) 小數，包括位值、排序與分數的換算（等等）----- | ☐ | ☐ | ☐ | ☐ |
| e) 使用文字、數字或模型（含數線）來表示小數及分數----- | ☐ | ☐ | ☐ | ☐ |
| f) 分數的計算----- | ☐ | ☐ | ☐ | ☐ |
| g) 小數的計算----- | ☐ | ☐ | ☐ | ☐ |
| h) 整數的表示、比較、排序和計算----- | ☐ | ☐ | ☐ | ☐ |
| i) 比例（等值，將一個數量按比例分配）----- | ☐ | ☐ | ☐ | ☐ |
| j) 百分比與分數或小數的相互換算----- | ☐ | ☐ | ☐ | ☐ |

B. 代數

- | | | | | |
|-------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 數字、代數和幾何規律或數列（延伸、缺項、將規律一般化）----- | ☐ | ☐ | ☐ | ☐ |
| b) 代數式的和、乘積及次方----- | ☐ | ☐ | ☐ | ☐ |
| c) 將給定的數值代入代數式作計算----- | ☐ | ☐ | ☐ | ☐ |
| d) 代數式的簡化或比較----- | ☐ | ☐ | ☐ | ☐ |
| e) 根據條件列出算式----- | ☐ | ☐ | ☐ | ☐ |
| f) 用變數的已知值估算函數/算式----- | ☐ | ☐ | ☐ | ☐ |
| g) 簡單的一元一次方程式、不等式及二元一次聯立方程式----- | ☐ | ☐ | ☐ | ☐ |
| h) 函數的各種表徵，如：數對、表格、圖形、文字或方程式----- | ☐ | ☐ | ☐ | ☐ |

20 接續前一頁

下列為TIMSS數學測驗的主題。請選擇一項最能描述該班在何時上這些數學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”。

	每小項請勾選一項		
	尚未教過或僅簡單介紹	大部分在今年教過	大部分以前教過
C. 幾何			
a) 角—銳角、直角、平角、鈍角、反射角-----	☐	☐	☐
b) 角與角之間的關係—共點的角、一條線上的角、對頂角、平行線與其截線所產生的角及垂直所產生的角-----	☐	☐	☐
c) 幾何形狀的性質：三角形、四邊形和常見多邊形-----	☐	☐	☐
d) 用作圖或畫出符合已知量的三角形與長方形-----	☐	☐	☐
e) 全等圖形（三角形、四邊形）與其對應值-----	☐	☐	☐
f) 相似三角形及其性質-----	☐	☐	☐
g) 平面圖形與立體圖形的關係-----	☐	☐	☐
h) 用畢氏定理求邊長（不需證明畢氏定理）-----	☐	☐	☐
i) 測量、作圖和估計角的大小、直線長度、面積和體積-----	☐	☐	☐
j) 測量公式：矩形周長、圓周長、圓面積、表面積和體積-----	☐	☐	☐
k) 測量不規則或複合圖形的面積（如用方格、切割和重組方式）-----	☐	☐	☐
l) 座標平面—數對、方程式、截距、交點與斜率-----	☐	☐	☐
m) 平面圖形的線對稱和旋轉對稱-----	☐	☐	☐
n) 平移、鏡射與旋轉-----	☐	☐	☐
D. 資料與機率			
a) 從表、統計圖表、長條圖、圓形圖和曲線圖讀取資料-----	☐	☐	☐
b) 用表、統計圖表、長條圖、圓形圖和曲線圖組織和展示資料-----	☐	☐	☐
c) 數據資料的特徵，包括平均數、中位數、全距及用一般語言形容分佈的形狀-----	☐	☐	☐
d) 詮釋資料（如：下結論、作預測、應用內插法和外插法）-----	☐	☐	☐
e) 展示的資料可能導致錯誤的詮釋（如不適當的彙整、量尺的誤導或曲解）-----	☐	☐	☐
f) 用實驗所得數據預測未來成果的機率-----	☐	☐	☐
g) 用特定結果的機率來解題-----	☐	☐	☐

受測班級的計算機和電腦

21

該班上數學課時，是否允許使用計算機？

請只勾選一項

是，且沒有使用限制-----○

是，但有使用限制-----○

否，不可以使用-----○

如果選答「否」，請跳至第23題 →

22

該班學生上數學課時，在下列活動中使用計算機的次數為何？

每小題勾選一項

從未使用

有些課

約有一半的課

每節或幾乎每節課

a) 驗算答案-----○--○--○--○

b) 一般的計算-----○--○--○--○

c) 解複雜的問題-----○--○--○--○

d) 探索數的概念-----○--○--○--○

23

A. 該班學生上數學課時，是否有電腦可用？

否

是

請勾選是或否-----○--○

如果選答「否」，請跳至第25題 →

B. 是否有可以上網的電腦？

否

是

請勾選是或否-----○--○

24

該班學生上數學課時，您多常讓他們用電腦進行下列的活動？

每小題請勾選一項

從未使用

有些課

約有一半的課

每節或幾乎每節課

a) 找出數學原理和概念-----○--○--○--○

b) 練習技巧和過程-----○--○--○--○

c) 查詢各種觀點和資訊-----○--○--○--○

d) 處理和分析資料-----○--○--○--○

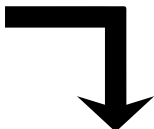
家庭作業

25

您是否會給該班學生做數學家庭作業？

請勾選是或否-----
是 否
○ ○

如果選答「否」，請跳至第30題



26

您通常多久給該班學生一次數學家庭作業？

請只勾選一項
每節或幾乎每節課-----○
約有一半的課-----○
有些課-----○

27

您給該班學生的數學家庭作業，每次需要花多少時間完成？（請以普通程度學生所需要時間來作估計）

請只勾選一項
少於 15 分鐘-----○
15-30 分鐘-----○
31-60 分鐘-----○
61-90 分鐘-----○
超過 90 分鐘-----○

28

您多久給該班做一次下列各種數學科的家庭作業？

每小題請勾選一項
從未或幾乎沒有
有時
總是或經常
a) 做習題-----○--○--○
b) 蒐集資料和做報告-----○--○--○
c) 找出一或多個上課內容
的應用例子-----○--○--○

29

您多常使用下列方式處理學生的數學科家庭作業？

每小題請勾選一項
從未或幾乎沒有
有時
總是或經常
a) 記錄作業是否完成-----○--○--○
b) 批改作業並給予回饋-----○--○--○
c) 要學生在教室訂正自己的作業---○--○--○
d) 用作業作為課堂討論之素材-----○--○--○
e) 做為學期成績的一部分-----○--○--○

30

您有多強調用下列來源考評學生的數學進步情形？

請只勾選一項

不強調
很少強調
有些強調
很強調

- a) 教室內測驗（教師自編或教科書上的測驗）----- ○ -- ○ -- ○ -- ○
- b) 國家或地區的成就測驗----- ○ -- ○ -- ○ -- ○
- c) 您的專業判斷----- ○ -- ○ -- ○ -- ○

31

您多常在該班實施數學測驗或考試？

請只勾選一項

- 約每週一次----- ○
- 約兩週一次----- ○
- 約每月一次----- ○
- 一年數次----- ○
- 從未實施----- ○

如果回答「從未實施」，您已完成問卷。
回答其他選項，請繼續作答。

32

您通常採用的數學科測驗題型為何？

請勾選一項

- 只用非選擇題----- ○
- 大多為非選擇題----- ○
- 約半數為非選擇題，半數為是非題或選擇題型----- ○
- 大多為是非題或選擇題型----- ○
- 只用是非題或選擇題型----- ○

33

您多常將以下各類題目納入數學測驗中？

每小題請勾選一項

從未或幾乎沒有
有時
總是或經常

- a) 回憶事實和過程的題目----- ○ -- ○ -- ○ -- ○
- b) 要應用到數學過程的題目----- ○ -- ○ -- ○ -- ○
- c) 找出規律與關係的題目----- ○ -- ○ -- ○ -- ○
- d) 需要提出解釋或證明的題目----- ○ -- ○ -- ○ -- ○

感謝您撥冗填答



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

數學教師問卷

八年級

識 別 資 料

教師姓名：_____

班級名稱：_____

教師代碼：_____ 教師連結 # _____

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

科學教師問卷

八年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

感謝 貴校已同意參與2007國際數學與科學教育成就趨勢調查(Trends in International Mathematics and Science Study, TIMSS 2007)，這是一個有關數學與自然科學學習狀況之大型國際調查，全球有超過六十個國家參加。本教育研究專案由國際教育成就調查委員會(International Association for the Evaluation Achievement, IEA)所主辦，目的在測量學生成就的趨勢與瞭解世界各國教育制度的差異，以求提昇全球數學及自然科學之教育水準。

本研究從我國國中八年級學生樣本中抽出部分學生來填寫T I M S S 的測驗卷。這份問卷主要是為受測班級學生的自然科教師所設計，問卷內容包括教師的學經歷背景，教學實務和教授自然科的態度。由於您任課的班級為全國受測樣本之一，您回答的資料對協助了解我國的自然科教育是非常重要的。

此問卷中有些問題提及受測班級，指的是此問卷封面上所標示之班級，也就是接受TIMSS 2007測驗的班級。請仔細回答每一個問題，以使您提供的訊息能充分反映出實際狀況。

請在不受干擾的時間及地點完成問卷，回答問卷大約需要四十五分鐘。為了作答方便，大部份問題僅須打勾即可。

當您做完這份問卷之後，請將問卷交給貴校聯絡人，連同試題一併寄回國立台灣師範大學科學教育中心。填答過程中若有疑問可以與下列人員聯絡詢問。

劉佩娟小姐	02-29316273分機314
王華鈺小姐	02-29316273分機341
任宗浩助理研究員	02-29316273分機324
洪志明教授	02-29316273分機312

1

您今年幾歲？

請只勾選一項

- 25歲以下 -----○
 25 - 29歲 -----○
 30 - 39歲 -----○
 40 - 49歲 -----○
 50 - 59歲 -----○
 60 歲以上 -----○

2

您的性別為何？

請只勾選一項

- 女性 -----○
 男性 -----○

3

到本學年結束時，您將有幾年的教學年資？

 （請填寫數字）年

4

您的最高學歷為何？

請只勾選一項

- 高中/職肄業 -----○
 師範學校/高中/職畢業-----○
 師專或其他五專、二專畢業-----○
 二技或四技畢業-----○
 大學 -----○
 碩士或博士學位-----○

5

您在中學以後所接受的教育，是主修哪些科或哪些領域？

每小題請勾選是或否

- | | 是 | 否 |
|---------------|---|---|
| a) 生物-----○ | ○ | ○ |
| b) 物理-----○ | ○ | ○ |
| c) 化學-----○ | ○ | ○ |
| d) 地球科學-----○ | ○ | ○ |
| e) 科學教育-----○ | ○ | ○ |
| f) 數學-----○ | ○ | ○ |
| g) 數學教育-----○ | ○ | ○ |
| h) 教育-----○ | ○ | ○ |
| i) 其他-----○ | ○ | ○ |

6

您有教師證嗎？

是 否

請勾選是或否-----○

7

您覺得能勝任下列八年級各主題的教學嗎？

每小題請勾選一項

	無法回答	非常勝任	尚能勝任	不能勝任
A. 生物				
a) 人體與其他生物的主要器官和系統（結構/功能，維持穩定生理狀況之生命過程）-----	☐	☐	☐	☐
b) 細胞及其功能，包括細胞進行之呼吸和光合作用-----	☐	☐	☐	☐
c) 生殖（有性與無性）和遺傳（性狀的遺傳，先天與後天/習得性狀的比較）-----	☐	☐	☐	☐
d) 在變動的環境下，變異與適應對物種生存/滅絕所扮演的角色-----	☐	☐	☐	☐
e) 在生態系統中生物與外在環境的交互作用 （能量流、食物網、環境變化之影響、物質循環）-----	☐	☐	☐	☐
f) 人口趨勢及其對環境的影響-----	☐	☐	☐	☐
g) 自然災害對人數、野生動物和環境的衝擊-----	☐	☐	☐	☐
B. 化學				
a) 物質分類與組成（元素、化合物、混合物之特性）-----	☐	☐	☐	☐
b) 物質的粒子結構（分子、原子、質子、中子和電子）-----	☐	☐	☐	☐
c) 溶液（溶劑、溶質、濃度/稀釋、溫度對溶解度之影響）-----	☐	☐	☐	☐
d) 常見酸鹼的特性與用途-----	☐	☐	☐	☐
e) 化學變化（反應物的變化、化學變化的證據、質量守恆、常見氧化反應—燃燒和生鏽）-----	☐	☐	☐	☐
C. 物理				
a) 物質的物理狀態與變化（以粒子間的運動/距離來解釋物質的性質； 熱/能量的增/減造成的物相變化、熱膨脹、體積與/或壓力之改變）-----	☐	☐	☐	☐
b) 能的形式、轉換和溫度-----	☐	☐	☐	☐
c) 光的基本性質/行為（反射、折射、光與顏色、簡單成像圖） 及聲音的性質/行為（振動產生聲音、透過介質之傳播、響度、 音調、音色、振幅、頻率、光和聲音的相對速度）-----	☐	☐	☐	☐
d) 電路（電流、電路類型、通路/斷路和並聯/串聯、電流/電壓的關係）-----	☐	☐	☐	☐
e) 永久磁鐵和電磁鐵的性質-----	☐	☐	☐	☐
f) 力與運動（力的種類、運動的基本描述、距離/時間關係圖之使用、密度與壓力之效應）-----	☐	☐	☐	☐

7 接續前一頁

您覺得能勝任下列八年級各主題的教學嗎？

每小題請勾選一項

	無法回答	非常勝任	尚能勝任	不能勝任
D. 地球科學				
a) 地球的結構與地貌（地殼、地函與地核；地形圖之使用）-----	○	○	○	○
b) 地球上的各種作用、循環與歷史（岩石循環、水循環、天氣規律、重大地質事件、化石與化石燃料之形成）-----	○	○	○	○
c) 環境影響（如污染、全球暖化、酸雨）-----	○	○	○	○
d) 地球自然資源之使用與保育（再生與非再生資源，土地/土壤及水資源的人為利用）---	○	○	○	○
e) 太陽系與宇宙中的地球（地球上的各種現象：晝夜、潮汐、月亮的盈虧、日月蝕、季節；地球與其他天體的物理特徵之比較；太陽是恆星）-----	○	○	○	○

8

您多常與其他教師有如下的互動？

每小題請勾選一項

- 每天或幾乎每天
每週1~3次
每月2~3次
從不或幾乎沒有
- a) 討論某個概念的教學方式-- ○ -- ○ -- ○ -- ○
b) 一起準備教材----- ○ -- ○ -- ○ -- ○
c) 觀摩其他教師上課----- ○ -- ○ -- ○ -- ○
d) 其他教師到**我**班上
作非正式的觀摩----- ○ -- ○ -- ○ -- ○

10

有關貴校，您對下列各項同意的程度為何？

每小題請勾選一項

- 非常不同意
不同意
同意
非常同意
- a) 本校附近治安良好----- ○ -- ○ -- ○ -- ○
b) 我覺得在學校很安全----- ○ -- ○ -- ○ -- ○
c) 本校有完備的保全措施---- ○ -- ○ -- ○ -- ○

9

過去兩年中，您是否曾參與和下列各項相關的研習活動？

每小題請勾選一項

- 否
是
- a) 自然科學學科內容----- ○ -- ○
b) 自然科學教材教法----- ○ -- ○
c) 自然科學課程----- ○ -- ○
d) 將資訊科技融入自然科學課程中----- ○ -- ○
e) 增進學生批判性思考或探究技能----- ○ -- ○
f) 自然科學評量----- ○ -- ○

11

有關貴校，下列問題的嚴重性為何？

每小題請勾選一項

- 很嚴重的問題
小問題
不是問題
- a) 本校建築物該大幅整修----- ○ -- ○ -- ○
b) 教室太擁擠----- ○ -- ○ -- ○
c) 教師在教室以外無
適當的工作空間----- ○ -- ○ -- ○
d) 進行科學實驗或探究時
沒有材料----- ○ -- ○ -- ○

您對貴校以下各項的評比為何？

每小題請勾選一項

- | | 很高 | 高 | 普通 | 低 | 很低 |
|------------------|----|---|----|---|----|
| a) 教師的工作滿意度 | ○ | ○ | ○ | ○ | ○ |
| b) 教師對學校課程目標的了解 | ○ | ○ | ○ | ○ | ○ |
| c) 教師達成學校課程進度的程度 | ○ | ○ | ○ | ○ | ○ |
| d) 教師對學生學習成就的期望 | ○ | ○ | ○ | ○ | ○ |
| e) 家長對學生學習成就的支持 | ○ | ○ | ○ | ○ | ○ |
| f) 家長對學校活動的參與度 | ○ | ○ | ○ | ○ | ○ |
| g) 學生愛惜學校資源的程度 | ○ | ○ | ○ | ○ | ○ |
| h) 學生力求在校有好表現的意願 | ○ | ○ | ○ | ○ | ○ |



受測班級

以下是有關受測班級（即標示在本問卷封面右上角之班級）的問題，這個班級也是 貴校參與TIMSS-2007研究計畫的受試樣本。

13

該班共有多少位學生？

_____ 位（請填寫人數）

14

您每週上該班自然科學課的時間為幾分鐘？

_____ 分鐘（請填寫數字）

15

A. 您上該班自然科學課時，是否有使用教科書？

_____ 否
_____ 是
請勾選是或否 ----- ○ ----- ○

如果選答「否」，請跳至第16題

B. 在該班自然科學教學過程中，您如何使用教科書？

請只勾選一項

作為主要教學依據 ----- ○

作為補充教材 ----- ○

16

受測班級每週上自然科學課時，學生進行下列各項活動所佔時間的百分比為何？

每小題請填寫百分比
各項百分比的總和為100%

- a) 檢討家庭作業 ----- %
- b) 聽老師講課 ----- %
- c) 在您的指導下進行解題 ----- %
- d) 在沒有您的指導下自行解題 ----- %
- e) 聽您重新講解及澄清教學內容或步驟 ----- %
- f) 進行測驗或小考 ----- %
- g) 與教學內容/目的無關的教室管理活動
（如處理干擾或維持秩序） ----- %
- h) 其他學生活動 ----- %
- 總和 ----- 100%

17

上該班自然科學課時，您要求學生進行下列活動的情況為何？

每小題請勾選一項

從來沒有
有些課
約一半的課
每節或幾乎每節課

- a) 觀察自然現象並描述他們所看到的 ☐ ☐ ☐ ☐ ☐
- b) 看教師示範實驗或探究活動 ☐ ☐ ☐ ☐ ☐
- c) 設計與規劃實驗或探究活動 ☐ ☐ ☐ ☐ ☐
- d) 進行實驗或探究活動 ☐ ☐ ☐ ☐ ☐
- e) 以小組方式進行實驗或探究活動 ☐ ☐ ☐ ☐ ☐
- f) 閱讀他們的教科書或其他的教材資料 ☐ ☐ ☐ ☐ ☐
- g) 讓學生背誦事實或原理 ☐ ☐ ☐ ☐ ☐
- h) 用科學的公式或定律解一般問題 ☐ ☐ ☐ ☐ ☐
- i) 對他們所學的事物提出解釋 ☐ ☐ ☐ ☐ ☐
- j) 將所學的科學連結到日常生活上 ☐ ☐ ☐ ☐ ☐

18

依您的看法，下列各項因素影響您教導該班的程度為何？

每小題請勾選一項

很有影響
有些影響
少許影響
沒有影響
無法回答

學生

- a) 學生的學業能力不同 ☐ ☐ ☐ ☐ ☐
 - b) 學生的背景差異很大（如：家庭經濟、語言等） ☐ ☐ ☐ ☐ ☐
 - c) 有特殊需求的學生（如：聽障、視障、語障、肢體障礙、心智或情緒/心理障礙） ☐ ☐ ☐ ☐ ☐
 - d) 學生缺乏學習興趣 ☐ ☐ ☐ ☐ ☐
 - e) 搗亂的學生 ☐ ☐ ☐ ☐ ☐
- 資源
- f) 電腦硬體不足 ☐ ☐ ☐ ☐ ☐
 - g) 電腦軟體不足 ☐ ☐ ☐ ☐ ☐
 - h) 使用電腦的支援不足 ☐ ☐ ☐ ☐ ☐
 - i) 學生的教科書不足 ☐ ☐ ☐ ☐ ☐
 - j) 學生用的其他教學設備不足 ☐ ☐ ☐ ☐ ☐
 - k) 教師演示教學或其他活動中使用的設備不足 ☐ ☐ ☐ ☐ ☐
 - l) 學校硬體設施不完備 ☐ ☐ ☐ ☐ ☐
 - m) 學生/老師的比值高 ☐ ☐ ☐ ☐ ☐

至本學年結束，您在該班教導下列各領域科學內容所佔的時間百分比大約為何？

（每小題請填寫百分比，總和為100%）

- a) 生命科學（如：構造及功能、生命過程、生殖和遺傳；天擇；生態系；身體保健）----- %
- b) 化學（如：物質的分類、組成及性質；化學變化）----- %
- c) 物理（如：物質的物理狀態與變化；能；光；聲音；電與磁；力與運動）----- %
- d) 地球科學（如：地球結構；變化過程和資源；太陽系與宇宙）----- %
- e) 其他，請說明：

----- %

總和 ----- 100%

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

A. 生物

- | | | | | | |
|--|-----------------------|----|-----------------------|----|-----------------------|
| a) 根據各種生理與行為特徵來分類生物----- | ☐ | -- | ☐ | -- | ☐ |
| b) 人類與其他生物的主要器官系統----- | ☐ | -- | ☐ | -- | ☐ |
| c) 系統如何運作以維持穩定的身體狀況----- | ☐ | -- | ☐ | -- | ☐ |
| d) 細胞構造與功能----- | ☐ | -- | ☐ | -- | ☐ |
| e) 在細胞與生物體內所進行的光合作用與呼吸作用，包含物質的使用與產生----- | ☐ | -- | ☐ | -- | ☐ |
| f) 生物的生命週期，包括人類、植物、鳥類、昆蟲----- | ☐ | -- | ☐ | -- | ☐ |
| g) 生殖（有性與無性）和遺傳（性狀之遺傳、先天與後天/習得性狀的比較）----- | ☐ | -- | ☐ | -- | ☐ |
| h) 在變動的環境下，變異與適應對物種生存/滅絕所扮演的角色----- | ☐ | -- | ☐ | -- | ☐ |
| i) 在生態系統中生物間的交互作用（能量流、食物鏈與食物網、食物塔以及變化對生態系的影響）----- | ☐ | -- | ☐ | -- | ☐ |
| j) 自然界的物質循環（水、碳/氧循環、生物體的分解）----- | ☐ | -- | ☐ | -- | ☐ |
| k) 人口趨勢及其對環境的影響----- | ☐ | -- | ☐ | -- | ☐ |
| l) 自然災害對人類、野生動物和環境的衝擊----- | ☐ | -- | ☐ | -- | ☐ |
| m) 常見傳染病的成因、感染/散播的方式、預防、身體的自然抵抗力和恢復能力----- | ☐ | -- | ☐ | -- | ☐ |
| n) 保健預防的方法（飲食、衛生、運動和生活方式）----- | ☐ | -- | ☐ | -- | ☐ |



20 接續前一頁

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分在以前教過

B. 化學

- | | | | | | |
|---|-----------------------|----|-----------------------|----|-----------------------|
| a) 物質分類與組成（物理性質及化學性質、純物質與混合物、分離方法）----- | ☐ | -- | ☐ | -- | ☐ |
| b) 物質的粒子結構（分子、原子、質子、中子和電子）----- | ☐ | -- | ☐ | -- | ☐ |
| c) 溶液（溶劑、溶質、溫度對溶解度之影響）----- | ☐ | -- | ☐ | -- | ☐ |
| d) 水的特性與用途（組成、熔/沸點、密度/體積的改變）----- | ☐ | -- | ☐ | -- | ☐ |
| e) 常見酸鹼的特性與用途----- | ☐ | -- | ☐ | -- | ☐ |
| f) 化學變化（反應物的變化、化學變化的證據、質量守恒）----- | ☐ | -- | ☐ | -- | ☐ |
| g) 需要氧參與的常見氧化反應（燃燒、生鏽）及常見物質發生氧化反應難易程度的傾向----- | ☐ | -- | ☐ | -- | ☐ |
| h) 將常見的化學反應分成放熱或吸熱反應----- | ☐ | -- | ☐ | -- | ☐ |

20 接續前一頁

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

C. 物理

- | | | | | | |
|--|-----------------------|----|-----------------------|----|-----------------------|
| a) 物質的物理狀態與變化（以粒子間的運動/距離來解釋體積、形狀、密度與可壓縮性等性質，質量在物理變化過程中是守恆的）----- | ☐ | -- | ☐ | -- | ☐ |
| b) 熔化、凝固、蒸發及凝結的過程（物相變化、熔/沸點、壓力與物質純度的效應）----- | ☐ | -- | ☐ | -- | ☐ |
| c) 能的形式、轉換，熱與溫度，包含熱傳遞----- | ☐ | -- | ☐ | -- | ☐ |
| d) 溫度的改變與體積及（或）壓力變化，以及與
粒子運動或速率變化之關係----- | ☐ | -- | ☐ | -- | ☐ |
| e) 光的基本性質/行為（反射、折射、光與顏色、簡單成像圖）----- | ☐ | -- | ☐ | -- | ☐ |
| f) 聲音的性質
（透過介質之傳播、描述聲音的方式（響度、音調、振幅、頻率）、在不同介質中的速率）- | ☐ | -- | ☐ | -- | ☐ |
| g) 電路（電流、電路類型、並聯/串聯）及電流與電壓的關係----- | ☐ | -- | ☐ | -- | ☐ |
| h) 永久磁鐵與電磁鐵的性質----- | ☐ | -- | ☐ | -- | ☐ |
| i) 力與運動（力的種類、運動的基本描述）、距離/時間關係圖之使用----- | ☐ | -- | ☐ | -- | ☐ |
| j) 密度與壓力之效應----- | ☐ | -- | ☐ | -- | ☐ |



20 接續前一頁

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

每一小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分在以前教過

D. 地球科學

- | | | | | | |
|--|-----------------------|----|-----------------------|----|-----------------------|
| a) 地球結構與地貌（地殼、地函與地核；地形圖）----- | ☐ | -- | ☐ | -- | ☐ |
| b) 地球上水的物理狀態、運動、組成及相對分布----- | ☐ | -- | ☐ | -- | ☐ |
| c) 地球大氣及其主要成分的相對含量----- | ☐ | -- | ☐ | -- | ☐ |
| d) 地球上的水循環（步驟、太陽能的角色、淡水的循環與淨化）----- | ☐ | -- | ☐ | -- | ☐ |
| e) 岩石循環及火成岩、變質岩、沈積岩之形成過程----- | ☐ | -- | ☐ | -- | ☐ |
| f) 氣象資料／圖表、天氣型態變化（如：季節性變化、緯度、高度及地理上的影響）----- | ☐ | -- | ☐ | -- | ☐ |
| g) 數十億年來發生的地質作用（如：侵蝕、造山運動、板塊運動）----- | ☐ | -- | ☐ | -- | ☐ |
| h) 化石及化石燃料之形成----- | ☐ | -- | ☐ | -- | ☐ |
| i) 環境議題（如污染、全球暖化、酸雨）----- | ☐ | -- | ☐ | -- | ☐ |
| j) 地球的資源（再生/非再生、資源保存、廢物處理）----- | ☐ | -- | ☐ | -- | ☐ |
| k) 土地經營（如蟲害控制）和人類使用（如耕作）的關係----- | ☐ | -- | ☐ | -- | ☐ |
| l) 淡水資源的供應與需求----- | ☐ | -- | ☐ | -- | ☐ |
| m) 根據天體在太陽系及宇宙中的位置/運動來解釋地球上的現象（如：晝夜、潮汐、年、
月球盈虧、日月蝕、季節；所看到太陽、月球、行星、星座的樣子）----- | ☐ | -- | ☐ | -- | ☐ |
| n) 地球與月球及其他行星地表特徵的比較
（如：大氣、溫度、水、與太陽的距離、公轉/自轉的週期，維持生命的能力）----- | ☐ | -- | ☐ | -- | ☐ |

■ 受測班級的計算機和電腦

21

A. 該班學生上自然科學課時，是否有電腦可用？

請勾選是或否-----
是 否
○ ○

如果選答「否」，請跳至第23題 ➡

B. 是否有可以上網的電腦？

請勾選是或否-----
是 否
○ ○

22

該班學生上自然科學課時，您多常讓他們用電腦進行下列的活動？

每小題請勾選一項

- | | 從未使用 | 有些課 | 約有一半的課 | 每節或幾乎每節課 |
|----------------------|------|-----|--------|----------|
| a) 進行科學活動或實驗----- | ○ | ○ | ○ | ○ |
| b) 以模擬的方式探討自然現象----- | ○ | ○ | ○ | ○ |
| c) 練習技巧和過程----- | ○ | ○ | ○ | ○ |
| d) 查詢各種觀點與資訊----- | ○ | ○ | ○ | ○ |
| e) 處理和分析資料----- | ○ | ○ | ○ | ○ |

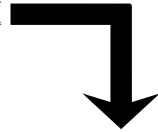
家庭作業

23

您是否會給該班學生做自然科學家庭作業？

請勾選是或否-----○-----○
否
是

如果選答「否」，請跳至第28題



24

您通常多久給該班學生一次自然科學家庭作業？

請只勾選一項
每節或幾乎每節課-----○
約有一半的課-----○
有些課-----○

25

您給該班學生的自然科學家庭作業，每次需要花多少時間完成？（請以普通程度學生所需要時間來作估計）

請只勾選一項
少於15分鐘-----○
15-30分鐘-----○
31-60分鐘-----○
61-90分鐘-----○
超過90分鐘-----○

26

您多久給該班做一次下列各種自然科學的家庭作業？

每小題請勾選一項
從未或幾乎沒有
有時
總是或經常
a) 做習題-----○--○--○
b) 找出一或多個上課內容的應用例子○--○--○
c) 閱讀課本或補充教材-----○--○--○
d) 寫定義或其他簡短的書面報告---○--○--○
e) 做專題-----○--○--○
f) 做小型研究或蒐集資料-----○--○--○
g) 準備報告-----○--○--○

27

您多常使用下列方式處理學生的自然科學家庭作業？

每小題請勾選一項
從未或幾乎沒有
有時
總是或經常
a) 記錄作業是否完成-----○--○--○
b) 批改作業並給予回饋-----○--○--○
c) 要學生在教室訂正自己的作業---○--○--○
d) 用作業作為課堂討論之素材-----○--○--○
e) 做為學期成績的一部分-----○--○--○

28

您有多強調用下列來源評定學生自然科學的進步情形？

請只勾選一項

不強調
|
很少強調
|
有些強調
|
很強調

- a) 教室內測驗（教師自編或教科書上的測驗）----- ○ -- ○ -- ○ -- ○
b) 國家或地區的成就測驗----- ○ -- ○ -- ○ -- ○
c) 您的專業判斷----- ○ -- ○ -- ○ -- ○

29

您多常在該班實施自然科學測驗或考試？

請只勾選一項

- 約每週一次----- ○
約兩週一次----- ○
約每月一次----- ○
一年數次----- ○
從未實施----- ○

如果回答「從未實施」，您已完成問卷。
回答其他選項，請繼續作答。

30

您通常採用的自然科學測驗題型為何？

請勾選一項

- 只用非選擇題----- ○
大多為非選擇題----- ○
約半數為非選擇題，半數為是非題或選擇題型--- ○
大多為是非題或選擇題型----- ○
只用是非題或選擇題型----- ○

31

您多常將以下各類題目納入自然科學測驗中？

每小題請勾選一項

從未或幾乎沒有
|
有時
|
總是或經常

- a) 需要知道事實與概念的題目----- ○ -- ○ -- ○
b) 知識應用與理解的題目----- ○ -- ○ -- ○
c) 涉及發展假設與設計科學探究的題目----- ○ -- ○ -- ○
d) 需要解釋或證明的題目----- ○ -- ○ -- ○

感謝您撥冗填答



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

科學教師問卷

八年級

附錄七

TIMSS 2007 四年級問卷（中文版）

識 別 資 料

學校代碼：

學校名稱：

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

學校問卷

四年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

感謝 貴校同意參與2 0 0 7 國際數學與科學教育研究趨勢之調查(Trends in International Mathematics and Science Study, TIMSS-2 0 0 7)，這是一個有關數學與自然科學學習狀況之大型國際性調查研究，全球有超過六十個國家參加。本調查研究由國際教育成就調查委員會(International Association for the Evaluation Achievement, IEA)所主辦，目的在測量學生成就的趨勢與瞭解世界各國教育制度的差異，以求提昇全球數學及自然科學之教育水準。

這份問卷主要是詢問各校校長及各處室主任一些學校的問題。由於 貴校被選為全國受測樣本之一，您回答的資料對協助了解我國的學校系統是非常重要的。

請仔細回答每一個問題，以使您提供的訊息能充分反映出 貴校的情況。在您填答的時候，有些問題也許必須查閱 貴校資料方能作答，因此您可能需要相關人員的協助來取得資料。

請在不受干擾的時間及地點完成問卷，回答此問卷大約需要三十分鐘。為了作答方便，大部份問題僅須填選即可。

當您做完這份問卷之後，請將問卷交給貴校之學校聯絡人，連同試題一併寄回國立台灣師範大學科學教育中心。謝謝您的作答。填答過程中若有疑問可以與下列人員聯絡詢問。

王華鈺小姐	02-29316273分機341
劉佩娟小姐	02-29316273分機314
任宗浩助理研究員	02-29316273分機324
洪志明教授	02-29316273分機312

學校特色

1

A. 請問 貴校註冊的學生總人數有多少？

學生總人數:_____人

B. 請問 貴校註冊的國小四年級學生人數有多少？

學生總人數:_____人

2

貴校所在的鄉、鎮或市有多少居住人口？

請只勾選一項

超過 500,000 人 -----○

100,001 ~ 500,000 人 -----○

50,001 ~ 100,000 人 -----○

15,001 ~ 50,000 人 -----○

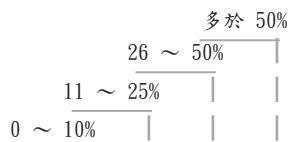
3,001 ~ 15,000 人 -----○

3,000 人以下 -----○

3

貴校學生中，下列背景的學生之百分比約為何？

每小題請勾選一項



a) 家境困難-----○--○--○--○

b) 家境富裕-----○--○--○--○

4

貴校學生中，以國語作為母語之百分比約為何？

請只勾選一項

高於 90% -----○

76 ~ 90% -----○

50 ~ 75% -----○

低於 50%-----○

5

針對 貴校四年級學生情形回答：

A. 貴校四年級學生一年上課幾天？

____天
(請填寫數字)

B. 貴校四年級全天課(扣除午餐休息，自修課及課後活動時間等)每天實際授課時間為多久？

____時____分
(請填寫數字)

C. 貴校四年級學生每週上課幾天？

請只勾選一項

6 天 -----○

5 1/2 天 -----○

5 天 -----○

4 1/2 天 -----○

4 天 -----○

其他 -----○

請寫出天數 _____

6

在這個學年結束時，您以校長身份從事下列活動的時間百分比為何？

每小題請輸入百分比
總共不可超過100%

- a) 行政工作
(如：約聘教職員、編預算、擬定行事曆、開會)----- %
- b) 教育領導
(如：發展課程及教學) ----- %
- c) 教職員視察與考核----- %
- d) 教學 ----- %
- e) 公關及募款----- %
- f) 其他 ----- %
- 總共 ----- 100%

7

貴校是否要求家長參與下列活動？

每小題請勾選一項

- | | 是 | 否 |
|-------------------------------------|-----------------------|-----------------------|
| a) 出席特別的活動
(如：科學展覽、音樂會、運動會) ---- | ☐ | ☐ |
| b) 為學校募款----- | ☐ | ☐ |
| c) 擔任學校計畫、正式/非正式課程或校外旅遊活動的志工----- | ☐ | ☐ |
| d) 督促孩子完成家庭作業----- | ☐ | ☐ |
| e) 擔任學校委員
(如：遴聘教職員及審核學校財務) ---- | ☐ | ☐ |

8

您對 貴校以下各項的評比如何？

每小題請勾選一項

- 很高 高 普通 低 很低
- a) 教師的工作滿意度 --- ○ --- ○ --- ○ --- ○ --- ○
- b) 教師對學校課程目標的了解
----- ○ --- ○ --- ○ --- ○ --- ○
- c) 教師達成學校課程進度的程度
----- ○ --- ○ --- ○ --- ○ --- ○
- d) 教師對學生學習成就的期望
----- ○ --- ○ --- ○ --- ○ --- ○
- e) 家長對學生學習成就的支持
----- ○ --- ○ --- ○ --- ○ --- ○
- f) 家長對學校活動的參與度
----- ○ --- ○ --- ○ --- ○ --- ○
- g) 學生愛惜學校資源的程度
----- ○ --- ○ --- ○ --- ○ --- ○
- h) 學生力求在校有好表現的意願
----- ○ --- ○ --- ○ --- ○ --- ○

9

貴校四年級的數學教學，是否採能力分班上課？

請勾選是或否 ----- ○ --- ○

是 否

10

請問 貴校對四年級學生是否有採取下列措施？

每小題請勾選是或否

- a) 提供數學課外補充教學----- ○ --- ○
- b) 提供數學補救教學----- ○ --- ○
- 是 否

11

貴校四年級的自然科教學，是否採能力分班上課？

請勾選是或否 ----- ○ --- ○

是 否

12

請問 貴校對四年級學生是否有採取下列措施？

每小題請勾選是或否

- a) 提供自然科課外補充教學----- ○ --- ○
- b) 提供自然科補救教學----- ○ --- ○
- 是 否

13

在過去兩年中，貴校四年級教師曾經參與下列各項以數學及科學為主的教師專業發展研習活動之百分比為何？

每小題請勾選一項

	從未	25% 或更低	26 - 50%	51 - 75%	76 - 100%
a) 支援全國性或區域性課程的實施	☐	☐	☐	☐	☐
b) 設計或支援校內的改進方案	☐	☐	☐	☐	☐
c) 加強學科知識的研習活動	☐	☐	☐	☐	☐
d) 加強教學技能的研習活動	☐	☐	☐	☐	☐
e) 資訊科技融入教學的研習	☐	☐	☐	☐	☐

14

貴校是否使用下列方式來評鑑四年級教師的教學實務？

每小題請勾選是或否

	是	否
a) 校長或資深教師之觀察	☐	☐
b) 由督學或校外人士之觀察	☐	☐
c) 學生之學習成就	☐	☐
d) 教師互評	☐	☐

15

這學年貴校對四年級教師缺額遞補的困難程度為何？

請只勾選一項

沒有缺額	☐
容易遞補	☐
有些困難	☐
很困難	☐

16

貴校目前是否有任何優惠方案（如薪水、住宿、簽約獎金、小班級等）來遴聘或留住四年級教師？

請勾選是或否

	是	否
	☐	☐

貴校四年級學生出現下列問題行為的次數及其嚴重程度為何？

A. 發生頻率

每小題請各勾選一項

從未 很少 每月 每週 每天

- a) 上課遲到-----○--○--○--○--○
- b) 無故缺席-----○--○--○--○--○
- c) 翹課-----○--○--○--○--○
- d) 服裝儀容不合規定-----○--○--○--○--○
- e) 擾亂教室秩序-----○--○--○--○--○
- f) 作弊-----○--○--○--○--○
- g) 說髒話-----○--○--○--○--○
- h) 破壞公物-----○--○--○--○--○
- i) 偷竊-----○--○--○--○--○
- j) 言語恐嚇嘲弄其他同學-----○--○--○--○--○
- k) 肢體傷害其他同學-----○--○--○--○--○
- l) 言語恐嚇污辱教職員-----○--○--○--○--○
- m) 肢體傷害教職員-----○--○--○--○--○

B. 問題嚴重性

每小題請勾選一項

不成問題 輕微問題 嚴重問題

-
-
-
-
-
-
-
-
-
-
-
-
-

貴校的教學，受到下列各項資源的缺乏或不足之影響為何？

每小題請勾選一項

沒有 很少 一些 很大

- a) 教材(如：教科書) ----- ○ -- ○ -- ○ -- ○
- b) 庶務預算(如：紙、筆) ---- ○ -- ○ -- ○ -- ○
- c) 學校建築及場地----- ○ -- ○ -- ○ -- ○
- d) 空調及照明系統 ----- ○ -- ○ -- ○ -- ○
- e) 教學空間(如：教室) ----- ○ -- ○ -- ○ -- ○
- f) 殘障學生所需的特殊設備-- ○ -- ○ -- ○ -- ○
- g) 數學教學用的電腦----- ○ -- ○ -- ○ -- ○
- h) 數學教學用的電腦軟體---- ○ -- ○ -- ○ -- ○
- i) 數學教學用的計算機----- ○ -- ○ -- ○ -- ○
- j) 與數學教學相關的圖書---- ○ -- ○ -- ○ -- ○
- k) 數學教學用的視聽設備---- ○ -- ○ -- ○ -- ○

每小題請勾選一項

沒有 很少 一些 很大

- l) 自然科學實驗室的設備與器材 ----- ○ -- ○ -- ○ -- ○
- m) 自然科學教學用的電腦---- ○ -- ○ -- ○ -- ○
- n) 自然科學教學用的電腦軟體 ○ -- ○ -- ○ -- ○
- o) 自然科學教學用的計算機 - ○ -- ○ -- ○ -- ○
- p) 與自然科學教學相關的圖書 ○ -- ○ -- ○ -- ○
- q) 自然科學教學用的視聽設備 ○ -- ○ -- ○ -- ○
- r) 師資----- ○ -- ○ -- ○ -- ○
- s) 電腦技術支援人員 ----- ○ -- ○ -- ○ -- ○

19

A. 貴校有自然科學實驗室嗎？

請勾選是或否 ----- ☐ 是 ☐ 否

B. 當學生進行科學實驗時，教師是否能獲得支援？

請勾選是或否 ----- ☐ 是 ☐ 否

20

A. 貴校在教學時提供四年級學生使用的電腦總數有幾台？

_____ 台

如果填 0，請直接回答第 21 題



B. 這些供教學用的電腦中，有多少台可以上網？
（包括e-mail或WWW）

請只勾選一項

全部都可以 ----- ☐

大部份可以 ----- ☐

有一些可以 ----- ☐

都不可以 ----- ☐

21

貴校是否有人可支援教師在教學與學習時使用資訊科技？

請勾選是或否 ----- ☐ 是 ☐ 否

感謝您撥冗填答



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

學校問卷

四年級

識 別 資 料

學校代碼：

學生代碼：

學校名稱：

學生姓名：

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

學生問卷

四年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

這份問卷要問你一些問題，有些問題是問關於你的一些事情，有些問題則是問你的看法。

請仔細讀每一個問題，並確實作答。如果你有不明白的地方或是不知道怎麼回答，可以舉手問老師。

有些問題的後面跟著幾個答案，這些答案是以加圈圈的數字來表示，回答問題時，請把你所選擇的答案圈起來，就像下面的例題一、二、三。

例題 1

你上學嗎？

請只圈選一項

是 ----- ①
否 ----- ②

例題 2

你常常做下列的事情嗎？

每項圈選一個

	每天 都有 ↓	每週至少 一次 ↓	每月 1~2次 ↓	每年 數次 ↓	從來 沒有 ↓
1) 聽音樂 -----	①	②	③	④	⑤
2) 和朋友聊天 -----	①	②	③	④	⑤
3) 運動 -----	①	②	③	④	⑤

例題 3

你對下列的說法同不同意？

每項圈選一個

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| 1) 看電影很有趣 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 我喜歡吃冰淇淋 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

請仔細讀每一個問題後，再選出你認為最合適的答案。如果你要更改答案，請先在原來的答案上打「X」，再重新選擇。如果你有任何問題或不知道如何作答，請舉手問老師。

謝謝你的作答

關於你自己

1

你是民國幾年幾月出生的？

1) 請把你出生時的年份圈起來

年

83 - ①

84 - ②

85 - ③

86 - ④

87 - ⑤

88 - ⑥

89 - ⑦

其他 - ⑧

2) 請把你出生時的月份圈起來

月

一月 - ○

二月 - ○

三月 - ○

四月 - ○

五月 - ○

六月 - ○

七月 - ○

八月 - ○

九月 - ○

十月 - ○

十一月 - ○

十二月 - ○

2

你是男生還是女生？

只圈選一個

女生----- ①

男生----- ②

3

你在家常說國語嗎？

只圈選一個

總是----- ①

常常----- ②

偶爾----- ③

從不----- ④

關於你自己（繼續）

4

你家大約有多少本書？（雜誌、報紙和學校的課本不算）

只圈選一個

沒有或很少

(0-10 本書)----- ① 這代表 10 本書



可以放滿一排

(11-25 本書)----- ② 這代表 25 本書



可以放滿一個書架

(26-100 本書)----- ③ 這代表 100 本書



可以放滿兩個書架

(101-200 本書)----- ④ 這代表 200 本書



可以放滿三個或三個以上的書架

(200 本書以上)----- ⑤ 這代表 200 本書以上



你家有下面這些東西嗎？

每小題只圈選一個

是



否



- | | | |
|----------------------|---------|---|
| 1) 計算機 ----- | ① ----- | ② |
| 2) 電腦 | | |
| (不包括任天堂和電視/電腦遊樂器) | ① ----- | ② |
| 3) 個人專用的書桌或桌子 ----- | ① ----- | ② |
| 4) 字典 ----- | ① ----- | ② |
| 5) 連接網路 ----- | ① ----- | ② |
| 6) 學習數學的光碟、軟體、錄影帶 -- | ① ----- | ② |
| 7) 學習數學的參考書 ----- | ① ----- | ② |
| 8) 學習自然科的光碟、軟體、錄影帶 | ① ----- | ② |
| 9) 學習自然科的參考書 ----- | ① ----- | ② |

6

你對下列學習數學的說法同不同意？

每小題圈選一項

- | | 很同意
↓ | 有點同意
↓ | 不太同意
↓ | 很不同意
↓ |
|--------------------------------|----------|-----------|-----------|-----------|
| 1) 我的數學不錯 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 我希望在學校多上一些數學課 ---- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 3) 相對於許多班上其他的同學，我覺得數學比較難 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 4) 我喜歡學習數學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 5) 我的數學不怎麼好 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 6) 與數學有關的事我學得很快 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 7) 數學很無趣 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 8) 我喜歡數學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

你上數學課時，多常做下面的這些事情？

每小題圈選一項

- | | 幾乎
每節課
↓ | 約有一
半的課
↓ | 有些課
↓ | 從來
沒有過
↓ |
|----------------------------------|----------------|-----------------|----------|----------------|
| 1) 不用計算機來做加減乘除的運算 -- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 作小數及分數的計算 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 3) 測量教室及學校裡的東西 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 4) 畫圖、表 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 5) 學習圓形、三角形、長方形和立方
體等形狀 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 6) 記住解題的方法 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 7) 和其他同學進行小組討論或活動 -- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 8) 說明自己的答案 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 9) 自己做數學題目 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 10) 使用計算機 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 11) 使用電腦 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

學校自然科學課

8

你對下列學習自然科學的說法同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|--------------------------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| 1) 我的自然科學不錯 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 我希望在學校多上一些自然科學課 | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 3) 相對於許多班上其他的同學，我覺
得自然科學比較難 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 4) 我喜歡學習自然科學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 5) 我的自然科學不怎麼好 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 6) 與自然科學有關的事我學得很快 -- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 7) 自然科學很無趣 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 8) 我喜歡自然科學 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

你上自然科學課時，多常做下面的事情？

每小題圈選一項

- | | 每週至少
一次
↓ | 每月做
1~2次
↓ | 每年做
數次
↓ | 從來
沒有
↓ |
|-----------------------------------|-----------------|------------------|----------------|---------------|
| 1) 我觀察如天氣或植物生長等現象，
並做紀錄 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 我看老師做科學實驗 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 3) 我自己設計一個科學實驗或研究 -- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 4) 我做科學實驗或研究 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 5) 我和同學分組一起做科學實驗或研究① ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 6) 我閱讀有關自然科學的書籍 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 7) 我記住科學知識 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 8) 我用寫的或用說的說明在自然課所
學到的東西 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 9) 我自己做自然科題目 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 10) 我在自然課使用電腦 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |

10

甲. 你使用過電腦嗎？（不包括電視/電腦遊樂器）

是 否
↓ ↓

請圈選是或否 ----- ① ----- ②

如果選答「否」，請直接回答第11題



乙. 你在下列哪些地方使用過電腦？

每小題圈選一項

是 否
↓ ↓

1) 家裡 ----- ① ----- ②

2) 學校 ----- ① ----- ②

3) 其他地方（如：公立圖書館、朋友家裡
、網咖） ----- ① ----- ②

丙. 你多常使用電腦做學校功課（包含在校內和校外）？

每小題圈選一項

每天 每週至少一次 每月1~2次 每年數次 從來沒有過
↓ ↓ ↓ ↓ ↓

1) 數學科 ----- ① ----- ② ----- ③ ----- ④ ----- ⑤

2) 自然科 ----- ① ----- ② ----- ③ ----- ④ ----- ⑤

你的學校

11

你對下面對學校的說法同不同意？

每小題圈選一項

- | | 很同意 | 有點同意 | 不太同意 | 很不同意 |
|---------------------------|---------|---------|---------|---------|
| | ↓ | ↓ | ↓ | ↓ |
| 1) 我喜歡待在學校 ----- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 2) 我們學校的學生都想盡力做好 ---- | ① ----- | ② ----- | ③ ----- | ④ ----- |
| 3) 我們學校的老師都希望學生盡力做好① ---- | ② ----- | ③ ----- | ④ ----- | |

12

上個月你在學校有發生過下面的事情嗎？

每小題圈選一項

- | | 是 | 否 |
|------------------------|---------|---------|
| | ↓ | ↓ |
| 1) 我有東西被偷了 ----- | ① ----- | ② ----- |
| 2) 有同學傷害我(如：推撞、踢打) --- | ① ----- | ② ----- |
| 3) 有同學叫我做我不願意做的事 ---- | ① ----- | ② ----- |
| 4) 有同學嘲笑我或戲弄我 ----- | ① ----- | ② ----- |
| 5) 同學不讓我參加他們的活動 ----- | ① ----- | ② ----- |

校外活動

13

在上學的日子，你每天在上學前或放學後通常會花多少時間做下面的事情？

每小題圈選一項

	沒有 ↓	1小時 以下 ↓	1-2 小時 ↓	2~4 小時 ↓	4小時 以上 ↓
1) 看電視、錄影帶 -----	① -----	② -----	③ -----	④ -----	⑤ -----
2) 玩電腦遊戲 -----	① -----	② -----	③ -----	④ -----	⑤ -----
3) 和朋友玩耍或聊天 -----	① -----	② -----	③ -----	④ -----	⑤ -----
4) 做家事 -----	① -----	② -----	③ -----	④ -----	⑤ -----
5) 運動 -----	① -----	② -----	③ -----	④ -----	⑤ -----
6) 看自己喜歡看的書 -----	① -----	② -----	③ -----	④ -----	⑤ -----
7) 上網 -----	① -----	② -----	③ -----	④ -----	⑤ -----
8) 做家庭作業 -----	① -----	② -----	③ -----	④ -----	⑤ -----

家庭作業

14

甲. 你的數學老師多常指定家庭作業？

只圈選一項

- 每天 ----- ①
- 一週 3~4 次 ----- ②
- 一週 1~2 次 ----- ③
- 一週不到 1 次 ----- ④
- 從來沒有指定數學作業 ----- ⑤

如果圈選「從來沒有指定數學作業」，請直接寫第15題



乙. 你的數學老師每次指定的家庭作業，你通常需要花多少分鐘才能做完？

只圈選一項

- 0 分鐘 ----- ①
- 1~15 分鐘 ----- ②
- 16~30 分鐘 ----- ③
- 31~60 分鐘 ----- ④
- 61~90 分鐘 ----- ⑤
- 超過 90 分鐘 ----- ⑥

家庭作業（繼續）

15

甲. 你的自然科學老師多常指定家庭作業？

只圈選一項

每天 ----- ①

一週 3~4 次 ----- ②

一週 1~2 次 ----- ③

一週不到 1 次 ----- ④

從來沒有指定自然科學作業 ----- ⑤

如果圈選「從來沒有指定自然科學作業」，請直接寫第 16 題



乙. 你的自然科學老師每次指定的家庭作業，你通常需要花多少分鐘才能做完？

只圈選一項

0 分鐘 ----- ①

1~15 分鐘 ----- ②

16~30 分鐘 ----- ③

31~60 分鐘 ----- ④

61~90 分鐘 ----- ⑤

超過 90 分鐘 ----- ⑥

更多關於你的事

16

甲. 你的母親(或繼母、養母、女性監護人等)是在中華民國出生的嗎？

是



否



請圈選是或否 ----- ① ----- ②

乙. 你的父親(或繼父、養父、男性監護人等)是在中華民國出生的嗎？

是



否



請圈選是或否 ----- ① ----- ②

17

甲. 你是在中華民國出生的嗎？

是



否



請圈選是或否 ----- ① ----- ②

如果圈選「否」，請作答下一題

乙. 如果你不是在中華民國出生的，你是幾歲時回到國內的？

只圈選一項

滿 5 歲以後 ----- ①

1~5 歲之間 ----- ②

不到 1 歲 ----- ③

謝謝你填答這份問卷



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

學生問卷

四年級

識 別 資 料

教師姓名：_____

班級名稱：_____

教師代碼：_____ 教師連結 # _____

國際數學與科學教育成就趨勢調查 2007

Trends in International Mathematics and Science Study

TIMSS 2007



實 測

教師問卷

四年級



International Association for the Evaluation
of Educational Achievement
© Copyright IEA, 2007

主辦單位：行政院國家科學委員會
教育部

承辦單位：國立台灣師範大學科學教育中心

填答說明

感謝 貴校已同意參與2007國際數學與科學教成就趨勢調查(Trends in International Mathematics and Science Study, TIMSS 2007)，這是一個有關數學與自然科學學習狀況之大型國際調查，全球有超過六十個國家參加。本教育研究專案由國際教育成就調查委員會(International Association for the Evaluation Achievement, IEA)所主辦，目的在測量學生成就的趨勢與瞭解世界各國教育制度的差異，以求提昇全球數學及自然科學之教育水準。

本研究從我國國小四年級學生樣本中抽出部分學生來填寫TIMSS的測驗卷。這份問卷主要是為受測班級學生的數學與自然學科教師所設計，問卷內容包括教師的學經歷背景，教學實務和教授數學與自然學科的態度。由於您任課的班級為全國受測樣本之一，您回答的資料對協助了解我國的數學教育是非常重要的。

此問卷中有些問題提及受測班級，指的是此問卷封面上所標示之班級，也就是接受TIMSS 2007測驗的班級。請仔細回答每一個問題，以使您提供的訊息能充分反映出實際狀況。

請在不受干擾的時間及地點完成問卷，回答問卷大約需要四十五分鐘。為了作答方便，大部份問題僅須打勾即可。

當您做完這份問卷之後，請將問卷交給貴校聯絡人，連同試題一併寄回國立台灣師範大學科學教育中心。填答過程中若有疑問可以與下列人員聯絡詢問。

王華鈺小姐	02-29316273分機341
劉佩娟小姐	02-29316273分機314
任宗浩助理研究員	02-29316273分機324
洪志明教授	02-29316273分機312

教師基本資料

1 您今年幾歲？

- 請只勾選一項
- 25歲以下 -----○
- 25 - 29歲 -----○
- 30 - 39歲 -----○
- 40 - 49歲 -----○
- 50 - 59歲 -----○
- 60 歲以上 -----○

2 您的性別為何？

- 請只勾選一項
- 女性 -----○
- 男性 -----○

3 到本學年結束時，您將有幾年的教學年資？

_____年
(請填寫數字)

4 您有教師證嗎？

- 否
是 |
請勾選是或否 -----○--○

5 您的最高學歷為何？

- 請只勾選一項
- 高中/職肄業或 -----○
- 師範學校/高中/職畢業 -----○
- 師專或其他五專、二專畢業 -----○
- 二技或四技畢業 -----○
- 大學 -----○
- 碩士或博士學位 -----○

6 您在中學以後所接受的教育，是主修哪些科或哪些領域？

A. 您在中學以後所接受的教育，是主修哪些科或哪些領域？

每小題請勾選是或否
否
是

- a) 初等教育 -----○--○
- b) 中等教育 -----○--○
- c) 數學 -----○--○
- d) 科學 -----○--○
- e) 其他 -----○--○

B. 如果您的主修是教育，您有下列領域的專長嗎？

每小題請勾選是或否
否
是 |

- a) 數學 -----○--○
- b) 科學 -----○--○
- c) 語言/語文 -----○--○
- d) 其他科目 -----○--○

7

您多常與其他教師有如下的互動？

- 每小題請勾選一項
- 每天或幾乎每天
每週1~3次
每月2~3次
從不或幾乎沒有
- a) 討論某個概念的教學方式-- ○ -- ○ -- ○ -- ○
b) 一起準備教材----- ○ -- ○ -- ○ -- ○
c) 觀摩其他教師上課----- ○ -- ○ -- ○ -- ○
d) 其他教師到**我**班上
作非正式的觀摩----- ○ -- ○ -- ○ -- ○

8

有關貴校，您對下列各項敘述的同意程度為何？

- 每小題請勾選一項
- 非常不同意
不同意
同意
非常同意
- a) 本校附近治安良好----- ○ -- ○ -- ○ -- ○
b) 我覺得在學校很安全----- ○ -- ○ -- ○ -- ○
c) 本校有完備的保全措施---- ○ -- ○ -- ○ -- ○

9

有關貴校，下列問題的嚴重性為何？

- 每小題請勾選一項
- 很嚴重的問題
小問題
不是問題
- a) 本校建築物該大幅整修----- ○ -- ○ -- ○
b) 教室太擁擠 ----- ○ -- ○ -- ○
c) 教師在教室以外無
適當的工作空間 ----- ○ -- ○ -- ○
d) 做實驗或探究的材料不足 ----- ○ -- ○ -- ○

10

有關貴校，您對以下的評比為何？

- 每小題請勾選一項
- 非常低
低
普通
高
非常高
- a) 教師的工作滿意度--- ○ -- ○ -- ○ -- ○
b) 教師對學校課程目標
的了解----- ○ -- ○ -- ○ -- ○
c) 教師達成學校課程進
度的程度----- ○ -- ○ -- ○ -- ○
d) 教師對學生學習成就
的期望----- ○ -- ○ -- ○ -- ○
e) 家長對學生學習成就
的支持----- ○ -- ○ -- ○ -- ○
f) 家長對學校活動的參
與度----- ○ -- ○ -- ○ -- ○
g) 學生愛惜學校資源的
程度----- ○ -- ○ -- ○ -- ○
h) 學生在校求好表現的
意願----- ○ -- ○ -- ○ -- ○

你覺得自己能勝任下列四年級各主題的教學嗎？

每小題請勾選一項

	無法回答	非常勝任	能勝任	不能勝任
A. 數				
a) 整數包括位值與排序-----	☐	☐	☐	☐
b) 整數的四則運算-----	☐	☐	☐	☐
c) 分數(整體或集合的一部分、在數線上的位置) -----	☐	☐	☐	☐
d) 使用文字、數字或模型來表示分數-----	☐	☐	☐	☐
e) 分數的比較與排序-----	☐	☐	☐	☐
f) 分數的加減-----	☐	☐	☐	☐
g) 小數的加減-----	☐	☐	☐	☐
h) 算式(找出缺少的數、用算式表示簡單的關係) -----	☐	☐	☐	☐
i) 數列的規律(數列的延伸與缺項的填補) -----	☐	☐	☐	☐
j) 表達已知整數數對間的關係-----	☐	☐	☐	☐
B. 幾何圖形與測量				
a) 角的比較與作圖-----	☐	☐	☐	☐
b) 常見幾何圖形的基本特性-----	☐	☐	☐	☐
c) 平面與立體圖形之關係-----	☐	☐	☐	☐
d) 求面積與周長-----	☐	☐	☐	☐
e) 估計面積與體積-----	☐	☐	☐	☐
f) 利用非正式座標系在平面上指出固定點的位置-----	☐	☐	☐	☐
g) 鏡射與旋轉-----	☐	☐	☐	☐
C. 資料處理				
a) 從表格、統計圖、長條圖或圓形比例圖中讀取數據-----	☐	☐	☐	☐
b) 根據所呈現的資料下結論-----	☐	☐	☐	☐
c) 從表格、統計圖、長條圖或圓形比例圖來呈現資料-----	☐	☐	☐	☐

受測班級的數學教學

以下第12題到第26題是有關受測班級（即標示在本問卷封面右上角之班級）的問題，這個班級也是貴校參與TIMSS-2007研究計畫的受試樣本。

12

A. 上數學課時，該班共有幾位學生？

_____位(請填寫人數)

B. 上題的學生中有幾位是四年級的學生？

_____位(請填寫四年級「數學」學生的人數)

13

您每週上該班四年級數學課的時間為幾分鐘？

_____分鐘（請填寫數字）

14

A. 您上該班四年級數學課時，是否有使用教科書？

_____否
_____是
請勾選是或否-----○-----○

如果選答「否」，請跳至第15題

B. 在教學過程中，您如何使用教科書？

請只勾選一項

作為主要教學依據-----○

作為補充教材-----○

15

一般而言，受測班級每週上數學課時，學生進行下列各項活動所佔時間的百分比為何？

每小題請填寫百分比
各項百分比的總和為100%

- a) 檢討家庭作業-----○%
- b) 聽老師講課-----○%
- c) 在您的指導下進行解題-----○%
- d) 在沒有您的指導下自行解題-----○%
- e) 聽您重新講解及澄清教學內容或步驟-----○%
- f) 進行測驗或小考-----○%
- g) 與教學內容/目的無關的教室管理活動
（如處理干擾或維持秩序）-----○%
- h) 其他學生活動-----○%
- 總和----- 100%

16

該班上數學課時，是否允許使用計算機？

請只勾選一項

- 是，且沒有使用限制-----○
 是，但有使用限制-----○
 否，不可以使用-----○

如果選答「否」，請跳至第18題 →

17

該班學生上數學課時，在下列活動中使用計算機的次數為何？

每小題勾選一項

- | | 從未使用 | 有些課 | 約有一半的課 | 每節或幾乎每節課 |
|----------------|------|-----|--------|----------|
| a) 驗算答案----- | ○ | ○ | ○ | ○ |
| b) 做一般的計算----- | ○ | ○ | ○ | ○ |
| c) 解複雜的問題----- | ○ | ○ | ○ | ○ |
| d) 探索數的概念----- | ○ | ○ | ○ | ○ |

18

A. 該班學生上數學課時，是否有電腦可用？

- | | 否 | 是 |
|-------------|---|---|
| 請勾選是或否----- | ○ | ○ |

如果選答「否」，請跳至第20題 →

B. 是否有可以上網的電腦？

- | | 否 | 是 |
|-------------|---|---|
| 請勾選是或否----- | ○ | ○ |

19

該班學生上數學課時，您多常讓他們用電腦進行下列的活動？

每小題請勾選一項

- | | 從未使用 | 有些課 | 約有一半的課 | 每節或幾乎每節課 |
|-------------------|------|-----|--------|----------|
| a) 找出數學原理和概念----- | ○ | ○ | ○ | ○ |
| b) 練習技巧和過程----- | ○ | ○ | ○ | ○ |
| c) 查詢資料----- | ○ | ○ | ○ | ○ |

20

上該班數學課時，您多常要求學生進行下列的活動？

每小題請勾選一項

從來沒有
有些課
約有一半的課
每節或幾乎每節課

- a) 不用計算機來做四則運算-- ○ -- ○ -- ○ -- ○
 b) 做分數及小數的題目----- ○ -- ○ -- ○ -- ○
 c) 測量教室及學校中的物件-- ○ -- ○ -- ○ -- ○
 d) 作圖表----- ○ -- ○ -- ○ -- ○
 e) 學習圓形、三角形、矩形、立方體等形狀 ----- ○ -- ○ -- ○ -- ○
 f) 寫出算式來解應用問題---- ○ -- ○ -- ○ -- ○
 g) 解釋他們的答案----- ○ -- ○ -- ○ -- ○
 h) 將所學的數學連結到日常生活----- ○ -- ○ -- ○ -- ○
 i) 記住公式與過程----- ○ -- ○ -- ○ -- ○

21

至本學年結束，您本學年在該班教導下列各項數學主題所佔的時間百分比為何？

(每小題請填寫百分比，總和為100%)

- a) 數(含：整數、分數、小數的計算以及數的規律性) ----- %
 b) 幾何圖形與測量(含平面及立體圖形、長度、面積和體積) ----- %
 c) 資料呈現(含讀、作和解釋圖表) ----- %
 d) 其他，請說明：
 ----- %

總和 ----- 100%

下列為TIMSS數學測驗的主題。請選擇一項最能描述該班在何時上這些數學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

		每小題請勾選一項		
		尚未教過或僅簡單介紹		
		大部分在今年教過		
		大部分以前教過		
A. 數				
a)	使用文字、圖表或符號來表示整數-----	☐	☐	☐
b)	整數，包含位值、排序-----	☐	☐	☐
c)	整數的計算-----	☐	☐	☐
d)	數的乘除-----	☐	☐	☐
e)	整數的估算-----	☐	☐	☐
f)	牽涉到比例的問題-----	☐	☐	☐
g)	分數(整體或集合的一部分、在數線上的位置)-----	☐	☐	☐
h)	等值分數-----	☐	☐	☐
i)	簡單分數大小的比較與排序-----	☐	☐	☐
j)	使用文字、數字或模型來表示分數或小數-----	☐	☐	☐
k)	相同分母之分數的加減-----	☐	☐	☐
l)	小數數值包括用文字或數字寫小數-----	☐	☐	☐
m)	小數的加減-----	☐	☐	☐
n)	求方程式中的未知數(例如： $17 + \underline{\quad} = 29$ ，空格中該填入什麼數值能使等式成立?)-----	☐	☐	☐
o)	用包含未知數的方程式表示簡單的情況-----	☐	☐	☐
p)	序列的延伸及找出規律中的缺項-----	☐	☐	☐
q)	描述序列中相鄰兩項的關係-----	☐	☐	☐
r)	符合已知規則的數對(例如：將第一個數乘以3再加上2，得到第二個數)-----	☐	☐	☐
s)	找出適當的規律來表達已知數對間的關係-----	☐	☐	☐

22 接續前一頁

下列為TIMSS數學測驗的主題。請選擇一項最能描述該班在何時上這些數學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選“大部分在今年教過”的選項；如果這個主題沒有在課程中，請勾選“尚未教過或僅簡單介紹”的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

B. 幾何圖形與測量

- | | | | |
|------------------------------------|-----------------------|-----------------------|-----------------------|
| a) 測量和估計長度----- | ☐ | ☐ | ☐ |
| b) 平行線與垂直線----- | ☐ | ☐ | ☐ |
| c) 比較角的大小和畫角(例如：直角、比直角大或小的角) ----- | ☐ | ☐ | ☐ |
| d) 常見幾何圖形的基本性質 ----- | ☐ | ☐ | ☐ |
| e) 認識平面表示法和立體表示法的關係----- | ☐ | ☐ | ☐ |
| f) 計算正方形和長方形的面積和周長----- | ☐ | ☐ | ☐ |
| g) 用已知的圖形或方格來計算面積----- | ☐ | ☐ | ☐ |
| h) 估計面積與體積----- | ☐ | ☐ | ☐ |
| i) 用非正式的座標系來標示平面上點的位置----- | ☐ | ☐ | ☐ |
| j) 線對稱圖形----- | ☐ | ☐ | ☐ |
| k) 反射與鏡射----- | ☐ | ☐ | ☐ |

C. 資料處理

- | | | | |
|--|-----------------------|-----------------------|-----------------------|
| a) 從表格、長條圖或圓形比例圖中讀取數據----- | ☐ | ☐ | ☐ |
| b) 從相關的數據組中比較資訊(例如：用圖表示每個班級中，學生所喜愛的冰淇淋口味，指出巧克力口味在哪個班級最受歡迎) ----- | ☐ | ☐ | ☐ |
| c) 用顯示的資料來回答無法直接從資料讀取的問題(例如：經過計算下結論和作預測) ----- | ☐ | ☐ | ☐ |
| d) 比對同一資料之不同呈現方式----- | ☐ | ☐ | ☐ |
| e) 使用表格、統計圖、長條圖或圓形比例圖來組織和呈現資料----- | ☐ | ☐ | ☐ |

23

您是否會給該學生做數學家庭作業？

請勾選是或否-----☐是☐否

如果選答「否」，請跳至第26題➡

24

您通常多久給該班學生一次數學家庭作業？

請只勾選一項
 每節或幾乎每節課-----☐
 約有一半的課-----☐
 有些課-----☐

25

您給該班學生的數學家庭作業，每次需要花多少時間完成？（請以普通程度學生所需要時間來作估計）

請只勾選一項
 少於15分鐘-----☐
 15-30分鐘-----☐
 31-60分鐘-----☐
 61-90分鐘-----☐
 超過90分鐘-----☐

26

依您的看法，下列各項因素影響您對於該班數學的教學之程度為何？

每小題請勾選一項
 很有影響
 有些影響
 有一點影響
 沒有影響
 無法回答

a) 學生的學業能力不同- ☐-☐-☐-☐-☐

b) 學生的背景差異很大
 （如經濟、語言）-- ☐-☐-☐-☐-☐

c) 有特殊需求的學生
 （如：聽障、視障、語障、肢體障礙、心智或情緒/心理障礙）---- ☐-☐-☐-☐-☐

d) 學生缺乏學習興趣--- ☐-☐-☐-☐-☐

e) 搗亂的學生----- ☐-☐-☐-☐-☐

27

過去兩年中，您是否曾參與和下列各項相關的研習活動？

每小題請勾選一項
 否
 是

a) 數學學科內容-----☐-☐-☐-☐-☐

b) 數學教材教法-----☐-☐-☐-☐-☐

c) 數學課程-----☐-☐-☐-☐-☐

d) 將資訊科技融入數學課程中-----☐-☐-☐-☐-☐

e) 增進學生批判性思考或解題技能----☐-☐-☐-☐-☐

f) 數學評量-----☐-☐-☐-☐-☐

您覺得能勝任下列各主題的教學嗎？

每小題請勾選一項

不能勝任

尚能勝任

非常勝任

無法回答

A. 生命科學

- | | | | | |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 人類和其他生物(動、植物)的主要身體結構及其功能 ----- | ☐ | ☐ | ☐ | ☐ |
| b) 動、植物的生殖及發育(特徵遺傳；常見的生物之生命週期) ----- | ☐ | ☐ | ☐ | ☐ |
| c) 不同環境中生物之外型特徵、行為模式及生存方式----- | ☐ | ☐ | ☐ | ☐ |
| d) 生物族群關係(如：簡單食物鏈、掠食者和被掠食者的關係) ----- | ☐ | ☐ | ☐ | ☐ |
| e) 環境變化(人類活動的影響、污染及其防制) ----- | ☐ | ☐ | ☐ | ☐ |
| f) 人類保健(如：傳染病之散播/預防、健康/疾病之徵兆、飲食、運動) ----- | ☐ | ☐ | ☐ | ☐ |

B. 物質科學

- | | | | | |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 依據物體或物質之物理性質分類(如：質量、形狀、體積、顏色、硬度、
材質、導熱/電、磁性)----- | ☐ | ☐ | ☐ | ☐ |
| b) 混合物之形成及分離----- | ☐ | ☐ | ☐ | ☐ |
| c) 物質三態(固態、液態、氣態)及其物理性質的差異(形狀、體積)，含物質
經過加熱或冷卻後的狀態改變(熔化、凝固、沸騰、蒸發、凝結)----- | ☐ | ☐ | ☐ | ☐ |
| d) 物質常見的變化(如：動/植物之腐化、燃燒、生鏽、烹調)----- | ☐ | ☐ | ☐ | ☐ |
| e) 常見的能源及其形式和他們實際上的用途
(如：風力、太陽能、電力、燃料、水力、食物) ----- | ☐ | ☐ | ☐ | ☐ |
| f) 光(如：光源和行為) ----- | ☐ | ☐ | ☐ | ☐ |
| g) 電路----- | ☐ | ☐ | ☐ | ☐ |
| h) 磁鐵的性質----- | ☐ | ☐ | ☐ | ☐ |
| i) 使物體運動的力(如：重力、推/拉力)----- | ☐ | ☐ | ☐ | ☐ |

C. 地球科學

- | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 地表特徵(如：山脈、平原、河流、沙漠) ----- | ☐ | ☐ | ☐ | ☐ |
| b) 地表上的水(位置、型態及流動) ----- | ☐ | ☐ | ☐ | ☐ |
| c) 空氣(組成、存在的證據、用途、維持生命的重要性) ----- | ☐ | ☐ | ☐ | ☐ |
| d) 常見的地表特徵(如：山脈、平原、河流、沙漠)及其與人
類土地利用之關係(如：農耕、灌溉、土地開發)----- | ☐ | ☐ | ☐ | ☐ |
| e) 每日或季節性氣候----- | ☐ | ☐ | ☐ | ☐ |
| f) 動、植物化石(如：年代、形成過程) ----- | ☐ | ☐ | ☐ | ☐ |
| g) 地球所處的太陽系(行星、太陽、月球) ----- | ☐ | ☐ | ☐ | ☐ |

受測班級的自然科學教學

第29題到第40題是有關受測班級（即標示在本問卷封面右上角之班級）的問題，這個班級也是 貴校參與TIMSS-2007研究計畫的受試樣本。

29

A. 上自然科學課時，該班共有幾位學生？

_____位(請填寫人數)

B. 上題的學生中有幾位是四年級的學生？

_____位(請填寫四年級「自然科學」學生的人數)

30

對於受測班級中的四年級學生，自然科學是否通常被視為一個獨立的科目(例如：並沒有和其他科目統整在一起)？

請勾選是或否-----
是 否
☐ ☐

A. 如果您填答「是」，您每週上該班

自然科學課的時間為幾分鐘？

_____分鐘（請填寫實際時間）

B. 如果您填答「否」，請估計您每週上該班有關

自然科學主題的時間大約為幾分鐘？

_____分鐘（請填寫實際時間）

31

A. 該班學生上自然科學課時，是否有電腦可用？

是 否
☐ ☐
請勾選是或否-----

如果選答「否」，請跳至第33題



B. 是否有可以上網的電腦？

是 否
☐ ☐
請勾選是或否-----

32

該班學生上自然科學課時，您多常讓他們用電腦進行下列的活動？

每小題請勾選一項

從未使用
有些課
約有一半的課
每節或幾乎每節課
☐ ☐ ☐ ☐

- a) 進行科學活動或實驗----- ☐ ☐ ☐ ☐
- b) 以模擬的方式探討自然現象 ☐ ☐ ☐ ☐
- c) 練習技巧和過程----- ☐ ☐ ☐ ☐
- d) 查詢各種觀點與資料----- ☐ ☐ ☐ ☐

33

上該班自然科學課時，您多常要求學生進行下列活動？

每小題請勾選一項

從來沒有

有些課

約一半的課

每節或幾乎每節課

- a) 觀察例如天氣或植物的生長等自然現象並做紀錄----- ☐ -- ☐ -- ☐ -- ☐
- b) 看教師的示範實驗----- ☐ -- ☐ -- ☐ -- ☐
- c) 實驗或探究活動的設計或規劃----- ☐ -- ☐ -- ☐ -- ☐
- d) 進行實驗或探究活動----- ☐ -- ☐ -- ☐ -- ☐
- e) 以小組方式進行實驗或探究活動----- ☐ -- ☐ -- ☐ -- ☐
- f) 閱讀他們的教科書或資源材料----- ☐ -- ☐ -- ☐ -- ☐
- g) 背誦事實或原理----- ☐ -- ☐ -- ☐ -- ☐
- h) 解釋一些他們探究的活動-- ☐ -- ☐ -- ☐ -- ☐
- i) 將所學的科學連結到日常生活上----- ☐ -- ☐ -- ☐ -- ☐
- j) 在自己的位置上獨自做功課 ☐ -- ☐ -- ☐ -- ☐

34

至本學年結束，您本學年在該班教導下列自然科學主題時所佔的時間百分比為何？

每小題請填寫百分比
各項百分比的總和為100%

- a) 生命科學（包括環境議題）----- _____ %
- b) 物質科學（包括物理及化學單元）----- _____ %
- c) 地球科學(包括地球系統及太陽系等)---- _____ %
- d) 其他，請說明：

_____ %

總和 ----- 100%

35

A. 您上該班自然課時，是否有使用教科書？

否

是

請勾選是或否----- ☐ -- ☐

如果選答「否」，請跳至第36題

B. 在自然科教學過程中，您如何使用教科書？

請只勾選一項

作為主要教學依據----- ☐

作為補充教材----- ☐

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選"大部分在今年教過"的選項；如果這個主題沒有在課程中，請勾選"尚未教過或僅簡單介紹"的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

A. 生命科學

- | | | | | | |
|---|-----------------------|----|-----------------------|----|-----------------------|
| a) 生物型態、特徵、分類----- | ☐ | -- | ☐ | -- | ☐ |
| b) 人類和其他生物(動、植物)的主要身體結構及其功能----- | ☐ | -- | ☐ | -- | ☐ |
| c) 常見生物的生命週期之共同階段(如：人類、蝴蝶、青蛙、植物)----- | ☐ | -- | ☐ | -- | ☐ |
| d) 動植物之生殖(特徵之遺傳)----- | ☐ | -- | ☐ | -- | ☐ |
| e) 不同環境中生物之外型特徵、行為模式及生存方式----- | ☐ | -- | ☐ | -- | ☐ |
| f) 身體對外界狀況(如：熱、冷、危險)及活動(如運動)之反應----- | ☐ | -- | ☐ | -- | ☐ |
| g) 植物和動物的能量需求(用太陽能製造食物和提供生長和治療的能量)----- | ☐ | -- | ☐ | -- | ☐ |
| h) 生物族群關係(如：簡單食物鏈、掠食者和被掠食者的關係)----- | ☐ | -- | ☐ | -- | ☐ |
| i) 環境變化(人類活動的影響、污染及其防制)----- | ☐ | -- | ☐ | -- | ☐ |
| j) 常見的傳染疾病(如：流行性感冒)的感染途徑、徵兆、預防和治療----- | ☐ | -- | ☐ | -- | ☐ |
| k) 保健方法，包含飲食、運動----- | ☐ | -- | ☐ | -- | ☐ |



36 接續上一頁

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選"大部分在今年教過"的選項；如果這個主題沒有在課程中，請勾選"尚未教過或僅簡單介紹"的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

B. 物質科學

- | | | | |
|--|-----------------------|-----------------------|-----------------------|
| a) 依據物體與物質的物理性分類----- | ☐ | ☐ | ☐ |
| b) 金屬特性與用途----- | ☐ | ☐ | ☐ |
| c) 混合物之形成與分離----- | ☐ | ☐ | ☐ |
| d) 水的特性與用途----- | ☐ | ☐ | ☐ |
| e) 物質三態(固態、液態、氣態)及以形狀體積來分辨其物理性質的差異----- | ☐ | ☐ | ☐ |
| f) 物質經加熱或冷卻後的狀態改變(溶化、凝固、沸騰、蒸發、凝結)----- | ☐ | ☐ | ☐ |
| g) 物質的變化(如：動/植物之腐化、燃燒、生鏽、烹調)----- | ☐ | ☐ | ☐ |
| h) 常見的能源及其形式和它們實際上的用途
(如：風力、太陽能、電力、燃料、水力、食物)----- | ☐ | ☐ | ☐ |
| i) 熱與溫度----- | ☐ | ☐ | ☐ |
| j) 常見的光源及與光有關的現象(如影子和彩虹
之形成、物體的可見度、鏡子、顏色)----- | ☐ | ☐ | ☐ |
| k) 由震動產生聲音----- | ☐ | ☐ | ☐ |
| l) 電路----- | ☐ | ☐ | ☐ |
| m) 磁鐵(南北極、相吸及相斥)----- | ☐ | ☐ | ☐ |
| n) 使物體運動的力(如：重力、推/拉力)----- | ☐ | ☐ | ☐ |

36 接續上一頁

下列為TIMSS自然科學測驗的主題。請選擇一項最能描述該班在何時上這些自然科學的主題。如果這個主題今年教了一半，但還沒全部教完的話，請勾選"大部分在今年教過"的選項；如果這個主題沒有在課程中，請勾選"尚未教過或僅簡單介紹"的選項。

每小題請勾選一項

尚未教過或僅簡單介紹

大部分在今年教過

大部分以前教過

C. 地球科學

- | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|
| a) 岩石、礦物、沙及土壤----- | ☐ | ☐ | ☐ | ☐ |
| b) 地表上的水(位置、型態及流動) ----- | ☐ | ☐ | ☐ | ☐ |
| c) 空氣(組成、存在的證據、用途、維持生命的重要性) ----- | ☐ | ☐ | ☐ | ☐ |
| d) 常見的地表特徵(如：山脈、平原、河流、沙漠)及
其人類土地利用之關係(如：農耕、灌溉、土地開發) ----- | ☐ | ☐ | ☐ | ☐ |
| e) 地球自然資源之使用與維護----- | ☐ | ☐ | ☐ | ☐ |
| f) 地球的水循環(河水由山入海，雲的形成及降雨) ----- | ☐ | ☐ | ☐ | ☐ |
| g) 每日的天氣或季節性氣候----- | ☐ | ☐ | ☐ | ☐ |
| h) 動、植物化石(如：年代、形成過程) ----- | ☐ | ☐ | ☐ | ☐ |
| i) 地球所處的太陽系(行星、太陽、月球) ----- | ☐ | ☐ | ☐ | ☐ |
| j) 地球沿地軸自轉(日、夜、影子) ----- | ☐ | ☐ | ☐ | ☐ |



37

您是否會給該學生做自然科學家庭作業？

請勾選是或否 ----- ☐ 是 ☐ 否

如果選答「否」，請跳至第40題 ➡

38

您通常多久給該班學生一次自然科學家庭作業？

請只勾選一項
 每節或幾乎每節課 ----- ☐
 約有一半的課 ----- ☐
 有些課 ----- ☐

39

您給該班學生的自然科學家庭作業，每次需要花多少時間完成？（請以普通程度學生所需要時間來作估計）

請只勾選一項
 少於15分鐘 ----- ☐
 15-30分鐘 ----- ☐
 31-60分鐘 ----- ☐
 61-90分鐘 ----- ☐
 超過90分鐘 ----- ☐

40

依您的看法，下列各項因素影響您該班自然科的教學之限制程度為何？

每小題請勾選一項
 很有影響 |
 有些影響 |
 一些影響 |
 沒有影響 |
 無法回答 |
 a) 學生的學業程度不同- ☐ -- ☐ -- ☐ -- ☐ -- ☐
 b) 學生的背景差異很大
 （如經濟、語言） -- ☐ -- ☐ -- ☐ -- ☐ -- ☐
 c) 有特殊需求的學生
 （如：聽障、視障、語障、身心障礙）
 ----- ☐ -- ☐ -- ☐ -- ☐ -- ☐
 d) 學生缺乏學習興趣 --- ☐ -- ☐ -- ☐ -- ☐ -- ☐
 e) 搗亂的學生 ----- ☐ -- ☐ -- ☐ -- ☐ -- ☐

41

過去兩年中，您是否曾參與和下列各項相關的研習活動？

每小題請勾選一項
 否
 是 |
 a) 自然科學學科內容 ----- ☐ -- ☐
 b) 自然科學教材教法 ----- ☐ -- ☐
 c) 自然科學課程 ----- ☐ -- ☐
 d) 將資訊科技融入自然科學課程中 ---- ☐ -- ☐
 e) 增進學生批判性思考或探究技能 ---- ☐ -- ☐
 f) 自然科學評量 ----- ☐ -- ☐

感謝您撥冗填答



TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

實 測

教師問卷

四年級

附錄八

出國人員出國報告書(第四年期部分)：

1. 參加 2008 NARST Annual International Conference 及參訪 NSF(National Science Foundation)
2. 參加 IERI 舉辦之 Assessment Designs, Item Response Theory and Proficiency Estimates in Large Scale Assessment
3. 參加 ACER 舉辦之 Data Analysis of Large-Scale Survey Data Workshop

出席國際學術會議心得報告

計畫編號	NSC 97-MOE-S-003-003
計畫名稱	國際數學與科學教育成就趨勢調查2007 (4/4) 甲
出國人員姓名 服務機關及職稱	任宗浩 國立台灣師範大學科學教育中心 助理研究員 李哲迪 國立台灣師範大學科學教育中心 助理研究員
會議時間地點	2009 年 2 月 13 日- 2 月 16 日 Princeton, New Jersey, USA
會議名稱	IERI Autumn Academy (Session II): Test Designs, Item Response Theory and Proficiency Estimates

一、 與會歷程

此工作坊為 IERI (IEA-ETS Research Institute) 所主辦。IERI 是兩個研究機構合作組成的組織：其中之一是主持 TIMSS 等國際調查研究的機構—IEA (International Association for the Evaluation of Educational Achievement)，另一個機構是負責 TOEFL、GRE、SAT 等許多美國大型測驗的中心—ETS (Educational Testing Service)。IERI 為推廣二次分析研究，鼓勵研究者使用學習成就調查資料庫，定期舉辦各種與研究設計和分析方法有關的工作坊。此次工作坊的主題是測驗設計、試題反應理論、以及能力估計。

工作坊分四天進行（議程如下表）。第一天介紹測驗之設計，主講人是 Eugenio J. Gonzalez 博士。其介紹的重點特別放在題本的設計上。第二天介紹試題反應理論在大型調查研究中的應用。主講人是 Kentaro Yamamoto 博士。第三天和第四天說明母群分布模型在能力估計中的應用。主講人是 Matthias von Davier 博士。在這個部份，講者介紹了一套由 ETS 所發展的軟體—DESI。此軟體可用以估計學生的能力，產生似真估計值 (plausible value)。工作坊進行的方式是先以理論介紹作為前導，隨後緊接著就是習作練習，前一節所學到的概念在後一節問題解決的過程中得以消化吸收。

Tuesday January 13

8:30 – 9:00	Registration
-------------	--------------

9:00 – 10:30	Welcome and Introductions Overview of Seminar and Agenda Review of Matrix Designs in Large Scale Assessment (Gonzalez)
10:30 – 10:45	Break
10:45 – 12:15	Review of Matrix Designs in Large Scale Assessment (Gonzalez)
12:15 – 1:45	Lunch
13:45 – 15:15	Matrix Designs in Large Scale Assessment (Work in Groups)
15:15 – 15:30	Break
15:30 – 17:00	Matrix Designs in Large Scale Assessment (Presentations)

Wednesday January 14

9:00 – 10:30	Review of Item Response Theory (IRT) Uses of IRT in Large Scale Assessment (Yamamoto)
10:30 – 10:45	Break
10:45 – 12:15	Item Response Theory Models (Yamamoto)
12:15 – 1:45	Lunch
13:45 – 15:15	Obtaining Item Difficulties and Person Scores (Yamamoto)
15:15 – 15:30	Break
15:30 – 17:00	Obtaining Item Difficulties and Person Scores (Yamamoto)
19:00 – 21:00	Group Dinner

Thursday January 15

9:00 – 10:30	Population Models in Large Scale Assessments (Von Davier)
10:30 – 10:45	Break
10:45 – 12:15	Population Models in Large Scale Assessments (Von Davier)
12:15 – 1:45	Lunch

13:45 – 15:15	Using DESI to obtain Person Scores (Von Davier)
15:15 – 15:30	Break
15:30 – 17:00	Using DESI to obtain Person Scores (Group Work)

Friday January 16

9:00 – 10:30	Putting it all together (Gonzalez, Von Davier, Yamamoto)
10:30 – 10:45	Break
10:45 – 12:15	Putting it all together (Gonzalez, Von Davier, Yamamoto)
12:15	Adjourn

二、 與會心得

整體而言，此次工作坊所介紹的內容頗為實用。以下以題本設計和能力估計為例說明之。

題本設計是大型調查研究必備的研究設計技術。爲了能準確地估計各國學生的能力，所需要的測驗試題不僅在題目數量上要夠多（以 TIMSS 為例，自然科學約爲 300 題左右），在涵蓋範圍上也要夠廣。但是學生的作答時間一般限制在兩個小時左右，因此每位學生能完成的題目數有限。爲了顧及測量的準確性和學生作答能力的限制，目前的作法是將試題分爲數種題本，題本間部分試題重複，每位學生僅作其中一種題本。由於試題數量減少，學生即可在兩小時內完成，又由於學生所做的試題有共同的部分，因此以整體學生而言，所做的試題數量就是足夠的。題本之設計除了要考量上述測驗設計的因素之外，還要考慮實務上遭遇的其他問題。例如：試題公布的問題。無論是 TIMSS、美國各州的成就測驗、或是我國的基本學力測驗，在當屆施測後，都需要公布部分試題。這些公布的試題在下一屆的題本中自然不能使用，而需以其他題目取代。這些公布試題應等量地分布在各種題本中。在本次工作坊中，講者對這類在實務上應注意的重點所做的提點是在其他地方難以學習到的。

爲了估計學生能力並表達測量誤差，TIMSS 和 PISA 等大型調查研究對學生能力的

估計均以似真值來表示。一個學生的能力由一組五個數值的似真值來表示。在產生似真值的標準程序上，除了考慮學生的答題反應之外，還考慮了母群的分布模型。所謂母群的分布模型是指在母群中各種特質的學生的分布情況。在估計學生能力時，利用學生答題反應之外的其他特質，運用條件機率的概念，即可更準確地推估學生的能力，減少測量誤差。在此次工作坊中，運用 DESI，講者展示了應用母群分布模型估計學生能力的具體方法。

不過，此次工作坊對於 IRT 在大型調查研究中的應用和似真值的估計說明得不夠深入。由於工作坊的時間有限，加上參加者程度不一，講師必須對內容有所取捨，這是可以理解的。

我國目前除參加 TIMSS、PIRLS、PISA、ICCS 等國際調查研究外，還發展了 TEPS、TASA 等大型調查研究。在規劃與執行大型調查研究時，對抽樣設計、試題發展、量尺建立、能力估計、二次分析統計方法等等課題，研究者不僅需要有所認識，還需要對這些主題有整體的瞭解；因為這些主題環環相扣，彼此有關，若無整體的瞭解，計畫難保順利執行。在研究設計和統計方法等相關技術的運用上，目前各研究團隊都還在起步階段。在這個階段，借重國外研究者的知識與經驗是提升我國研究能力的重要策略。類似的工作坊若能邀請 IEA 與 ETS 的研究者來台舉辦，必能使國內更多學者受惠。

三、 攜回資料名稱及內容

1. 各場次講者提供之簡報檔及習作資料。
2. DESI 程式軟體一套。

The IEA/ETS Research Institute



The IEA/ETS Research Institute (www.IERInstitute.org)

What is it?

- Collaborative agreement between IEA and ETS
- Focus on improving the science of large scale assessment
- Undertakes activities around three broad areas:
 - ✓ Research studies related to the development and implementation of large-scale assessments
 - ✓ Professional development and training
 - ✓ Dissemination of research findings and information gathered through large-scale assessments



The IEA/ETS Research Institute (www.IERInstitute.org)

1

Research Activities

- A "Virtual" Institute
- Research Agenda
- Make available consulting and advisory resources to researchers and national centers wishing to conduct policy or methodology related studies
 - ✓ Identify relevant policy issues
 - ✓ Identify analytical strategies most appropriate for providing and analyzing data to address these issues
 - ✓ Apply statistical techniques and other research methodologies



The IEA/ETS Research Institute (www.IERInstitute.org)

2

Research Activities

- Research Priorities:
 - ✓ Develop a more scientific approach to the development, use and interpretability of background questionnaires
 - ✓ Develop new constructs that extend the policy issues that might be addressed by these assessments
 - ✓ Improve the measurement of cognitive domains
 - ✓ Investigate effects of increased emphasis on the role of technology
 - ✓ Identify thematic issues for guiding secondary analyses



The IEA/ETS Research Institute (www.IERInstitute.org)

3

Training Activities

- Workshops and seminars
 - ✓ Thematic
 - ✓ Making use of specific data
- Internship and fellowship programs
- Professional exchanges



The IEA/ETS Research Institute (www.IERInstitute.org)

4

Dissemination Activities

- Sponsor a Monograph Series
 - ✓ Published annually
 - ✓ 5-7 papers
 - ✓ Peer reviewed
 - Receive favorable technical, substantive, and editorial review
- Sponsor other publications that contribute to the advancement of the field



The IEA/ETS Research Institute (www.IERInstitute.org)

5

Matrix Designs in Large Scale Assessment

Eugenio J. Gonzalez
November 2008



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

So many items, so little time...



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

The Assessment Design

- How do we deliver the items to the population of interest?



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

2

Some Questions to Ask

- What is the purpose of the assessment?
- How broad is the content area or set of skills covered?
- What scores will be reported?
- How much time is available to administer the item?
- How will the data be reported?
- What decisions will be made based on results?
- Number of administrations
 - ✓ Same year and over the years



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

3

Purpose of the Assessment

- Selection
 - ✓ Have required set of skills
- Grading
 - ✓ Have mastered contents of course
- Diagnosis
 - ✓ Decide on individual treatment
- Describe population
 - ✓ Mastery of set of skills or knowledge



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

4

Broadness of the Assessment

- Contents covered at grade, or up to that grade?
- How many content areas?
- What skills are measured?
- How long does it take to measure those skills?
- How many items are necessary



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

5

Scales to be Reported

- How many scales?
- How many subscales?
- Who gets scores on what?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

6

Time Available for Administration

- How long does it take to perform individual tasks?
- How long will students last before fatigue sets in?
- School schedule structure
- What will teachers and school principals tolerate?
- Resources available?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

7

Reporting the Results

- Overall scale scores
- Subscale scores
- Percent correct or scale scores?
- Percents passing benchmarks
- Item by item
- Passing / no passing
- Other?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

8

Decisions Based on Results

- Describe population?
- Describe individuals? School performance?
- Individual consequences?
- Group consequences?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

9

Number of Administrations

- How many administration will take place?
- An on-going program? Or once in a lifetime?
- Is there an item release policy?
- Risks of disclosure/exposure of the items?
- Refreshing the assessment?
- (When) Will new component(s) be introduced?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

10

Assessment Designs in LSA

- Depend on answers to previous questions
- Except for very simple cases there is no unique solution
- The design is always a compromise between what is desired and what is possible given specific circumstances and resources
 - ✓ Need to weigh pros and cons of each feature vs. choosing a different one
 - ✓ Make an overall judgment of the adequacy of the design(s)



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

11

Some of the Challenges...

- Domain is usually very broad
- Limited testing time (physical & psychological)
- Challenges:
 - ✓ Need to administer the items in a "sensible" design
 - ✓ Need to summarize performance on the items
 - ✓ Need to adjust for unreliability of estimates



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

12

How do Matrix Designs Help?

- Impossible to test everyone on everything
 - ✓ Too many people
 - ✓ Too many items
 - ✓ Too expensive
- Some students are tested on some things
- In large scale assessments we tend to sample items and students
- Calculations need to be adjusted using sampling design information



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

13

Arranging the Items

- Items are the individual questions or tasks that make up the assessment
- Blocks are subsets of items in the assessment
 - ✓ A subset could be one, some, or all the items in the assessment
 - ✓ Blocks are generally exhaustive and mutually exclusive
 - Every item belongs to a block
 - Each item belongs to only one block



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

14

Some Concepts

- The "assessment" constitutes the entirety of the items
- The "blocks" are subsets of items (also called clusters)
- The "form" is a particular set of blocks delivered to the students in a particular order
 - ✓ In paper and pencil applications we use booklet or form
 - ✓ In computer delivered assessments there is no booklet...



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

15

Some Choices of Designs

- Complete Matrix Designs
 - ✓ Everyone takes everything
 - ✓ Could have blocks in different order
 - ✓ One form or multiple forms
- Incomplete Matrix Design
 - ✓ Everyone takes part of the assessment
 - Dynamic or Static
 - Balanced or unbalanced



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

16

Complete Matrix Design

- All students take all the items in the same order
 - ✓ IEA Reading Literacy 1999 and 2001
 - One form, 2 blocks (A and B)
 - ✓ Advantages? Disadvantages?
- All students take all the items but in varying forms
 - ✓ Random or semi random: an unknown number of forms
 - ✓ Static: there is a fixed set of forms
 - ✓ Advantages? Disadvantages?
- Some Examples



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

17

Incomplete Matrix Designs

- ✓ Each student takes only part of the assessment
 - Dynamic: based on performance
 - Static: pre-printed or predetermined sets of items
- ✓ Advantages? Disadvantages?
- ✓ Some examples



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

18

Some Studies

- Two major organizations:
 - ✓ International Association for the Evaluation of Educational Achievement (IEA)
 - Conducts TIMSS & PIRLS
 - Major activity is comparative educational studies, research and training
 - ✓ Organization for Economic Cooperation and Development (OECD)
 - Conducts PISA
 - Data collection in education and other fields
 - Collects indicators to inform governments



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

19

The Studies

- TIMSS
 - ✓ Trends in International Mathematics and Science Study
- PIRLS
 - ✓ Progress of International Reading Literacy Study
- PISA
 - ✓ Program of International Student Assessment



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

20

Overview of Studies

	TIMSS	PIRLS	PISA
Subjects	Mathematics and Science	Reading literacy	Science Literacy, Mathematics, Reading
Grades/age	2003: 4 th /8 th	4 th	15-years old
Participating Jurisdictions	2003: 49	2006: 46	2006: 57



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

21

Overview of Studies

	TIMSS	PIRLS	PISA
Cycle	4 years	5 years	3 years
Last release	TIMSS 2003	PIRLS 2006	PISA 2006
Test Booklets (last cycle)	12 booklets	12 booklets + Reader	13 booklets
Testing time	90 minutes	80 minutes	120 minutes
Questionnaire	Principals, Teachers, Students	Principals, Teachers, Students (Parents)	Principals, Students (Parents)



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

22

TIMSS Assessment Design

- 90 minute assessment at grade 8
- Rotated block assessment design: blocks of items were rotated among 12 booklets.
- No student answered all items
- Students encountered booklets with both mathematics and science items
- Each booklet contained both trend items (1995, 1999) and new items.



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

23

TIMSS Achievement Scales

- Overall Mathematics
 - ✓ At Grade 8: Number; Algebra; Measurement; Geometry; Data
 - ✓ At Grade 4: Number; Patterns, Equations, and Relationships; Measurement; Geometry; Data
- Overall Science
 - ✓ At Grade 8: Life Science; Chemistry; Physics; Earth Science; Environmental Science
 - ✓ At Grade 4: Life Science; Physical Science; Earth Science
- Also cognitive domain scales



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

24

The TIMSS 2003 Test Design

Book	Pos 1	Pos 2	Pos 3	Pos 4	Pos 5	Pos 6
1	M01	M02	S06	S07	M05	M07
2	M02	M03	S05	S08	M06	M08
3	M03	M04	S04	S09	M13	M11
4	M04	M05	S03	S10	M14	M12
5	M05	M06	S02	S11	M09	M13
6	M06	M01	S01	S12	M10	M14
7	S01	S02	M06	M07	S05	S07
8	S02	S03	M05	M08	S06	S08
9	S03	S04	M04	M09	S13	S11
10	S04	S05	M03	M10	S14	S12
11	S05	S06	M02	M11	S09	S13
12	S06	S01	M01	M12	S10	S14



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

25

What is PIRLS?

- International study of reading literacy at 4th grade
- Assessment of student proficiency in reading comprehension
- Extensive collection of data on context for learning to read



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

26

PIRLS 2006 Test Design

- 10 passages; 5 literary & 5 informational
- 126 items; 167 score points
- Multiple-choice and constructed-response questions
- PIRLS Reader

PIRLS 2006 Booklets														
	1	2	3	4	5	6	7	8	9	10	11	12	R	
Part 1	L1	L2	L3	L4	I1	I2	I3	I4	L1	I2	L3	I4	L5	
Part 2	L2	L3	L4	I1	I2	I3	I4	L1	I1	L2	I3	L4	I5	



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

27

What is PISA?

- Program for International Student Assessment
- The main questions are:
 - ✓ Are students well prepared to meet the challenges of the future?
 - ✓ Are they able to analyze, reason, and communicate their ideas effectively?
 - ✓ Do they have the capacity to continue learning throughout life?



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

28

What is PISA?

- Measure how well 15-year-old approaching the end of compulsory schooling are prepared to meet the challenges of today's knowledge societies
- Does not focus on mastery of school curriculum, but on the ability to use knowledge and skills to meet "real-life challenges"



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

29

PISA Test Design

- Administered every 3 years
- Originally 3 major domains: Mathematics, Science and Reading
 - ✓ In 2003 included Problem Solving
- Every 9 years major emphasis on a domain
- In between the 9 years minor emphasis on a domain



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

30

PISA Test Design

	2000	2003	2006	2009
Reading	MAJOR	Minor	Minor	MAJOR
Mathematics	Minor	MAJOR	Minor	Minor
Science	Minor	Minor	MAJOR	Minor



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

31

PISA 2006 Test Design

PISA 2006 Booklets				
Form	Block 1	Block 2	Block 3	Block 4
1	S1	S2	S4	S7
2	S2	S3	M3	R1
3	S3	S4	M4	M1
4	S4	M3	S5	M2
5	S5	S6	S7	S3
6	S6	R2	R1	S4
7	S7	R1	M2	M4
8	M1	M2	S2	S6
9	M2	S1	S3	R2
10	M3	M4	S6	S1
11	M4	S5	R2	S2
12	R1	M1	S1	S5
13	R2	S7	M1	M3



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

32

Comments? Questions?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

Analysis with IRT

IERI Academy Winter 2009
Center of Global Assessment
ETS
Kentaro Yamamoto

Presentation Outline

- Classical Test Theory (CTT)
- Item Response Theory (IRT)
- Common IRT Models
- Parameter Estimations
- Features of IRT model
- Application of IRT for test development
- Consequence of assessment features on IRT modeling
- Weighting
- Booklet design
- Non-response
- Scoring accuracy

Establishing the Comparability

- Instrumentation
 - BQ, Cognitive, Instruction, Manual
- Sampling, Target Population, Sampling Frame
- Weighting
- Operation
- **Scoring**
- **Scaling (Non-response, Common parameters)**
- **Population modeling**
- **Reporting**

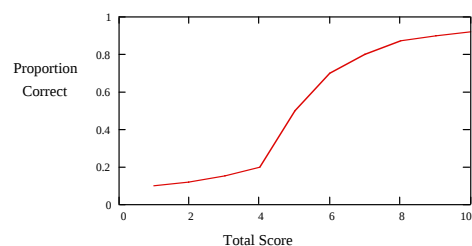
Moh's Scale of Mineral Hardness

	Sample	A	B	C	D	E	F
1	Talc	O	O	O	O	O	O
2	Gypsum	O	O	O	O	O	O
3	Calcite	X	O	O	O	O	O
4	Fluorite	X	X	O	O	O	O
5	Apatite	X	X	O	O	O	O
6	Orthoclase	X	X	X	O	O	O
7	Quartz	X	X	X	X	O	O
8	Topaz	X	X	X	X	O	O
9	Corundum	X	X	X	X	X	O
10	Diamond	X	X	X	X	X	X
		2.5	3.5	5.5	6.5	8.5	9.5

Typical Math Test

Student	Items											Total
	1	2	3	4	5	6	7	8	9	10	11	
A	O	O	O	O	O	O	O	O	O	X	O	10
B	O	O	O	O	X	O	O	O	O	O	X	9
C	O	O	O	O	O	O	X	O	X	X	O	8
D	O	X	O	O	O	X	O	O	O	O	X	8
E	O	O	O	O	X	O	O	X	X	O	X	7
F	X	O	O	X	O	O	X	O	O	X	X	6
G	O	O	X	X	O	X	O	X	X	X	X	4
H	O	O	X	O	X	X	X	X	X	X	X	2
Total	7	7	6	6	5	5	5	5	4	3	2	

Common Observation



Classical Test Theory

$$\text{Observed Score} = \text{True Score} + \text{Error Score}$$

$$O = T + e$$

Observed Score: test score

(can be standardized and/or transformed into various forms)

True Score: hypothetical mean of an infinite number of tests taken by an examinee when practice effects are disregarded, i.e., expectation of the scores

Error: amount of discrepancy between examinees' true scores and observed scores for population

Classical Test Theory

- Offers a diverse collection of techniques and formulas for item analysis, test development, scoring, scaling, and equating, as well as the evaluation of test reliability, validity and fairness
- Has a long history of theoretical and practical development with well developed body of knowledge
- Very useful for evaluating respondents skills and analysis of test scores
- Used often to produce parallel test forms

Limitations of CTT

- Sample Dependent
 - Statistics (e.g., difficulty, discrimination) that describe test items depend upon group of respondents who take the items
- Test form Dependent
 - Observed & true scores that describe how respondents perform depend upon the collection of items given to the students

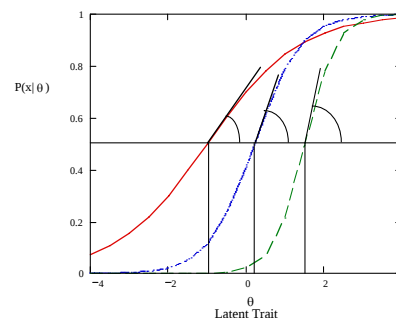
Limitations of CTT

- Same errors of measurement for all examinees regardless of examinees ability levels
 - Measurement error is population based
- Not able to match test forms for projected ability distribution
- Strictly parallel test do not exist
- Requires equating of
 - Item statistics for use in test assembly
 - Test scores for use across administrations

Item Response Theory

- Provides a consistent framework for thinking about and resolving testing problems
- Developed in 1960s to addresses some of the limitations of CTT, and offers a more flexible set of techniques
 - Estimate of standard error of measurement for each examinee
 - IRT pattern scoring does not require equating
 - Automated test assembly & CAT
- Reminds us that our testing goal is to measure an unobservable or latent construct, and that it is critical to consider the relationship of observable item responses to the latent construct

Item Characteristic Curve



Item Response Theory

- Offers a variety of probability models relating observable item responses, x_i , to the latent trait or ability, θ , and the item's measurement properties, $\underline{\beta}$ (item difficulty and discrimination, e.g.)

$$P(x_i = k | \theta, \underline{\beta}), \begin{cases} k = 0 \text{ or } 1 \text{ for dichotomous items} \\ k = 0, 1, \dots, K \text{ for polytomous items} \end{cases}$$

Advantage of IRT

- Because it makes stronger assumptions about the behavior of people and items, it allows us to solve problems that cannot be solved by classical test theory.
- The use of IRT allows us to answer many questions about a test before we ever administer it.

Advantage of IRT

- Evaluate the effectiveness of a test at different levels of ability
- Evaluate the ability of respondents who took different set of items
- Evaluate a set of items on a same scale established by a different set
- Design tests to measure best at specific ability level
- Develop new tests and investigate them without administering them
- Develop item statistics that do not change when the group of examinees change

Common IRT Models

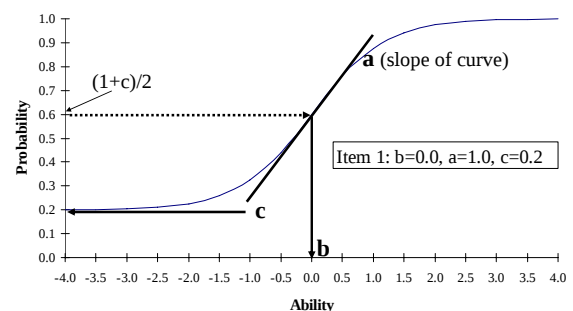
- Dichotomous IRT Models for Multiple Choice Items:
 - one-parameter logistic (1PL) model
 - two-parameter logistic (2PL) model
 - three-parameter logistic (3PL) model – most common
- Polytomous IRT Models for Multiple Responses:
 - graded response model
 - partial credit model

3PL Model

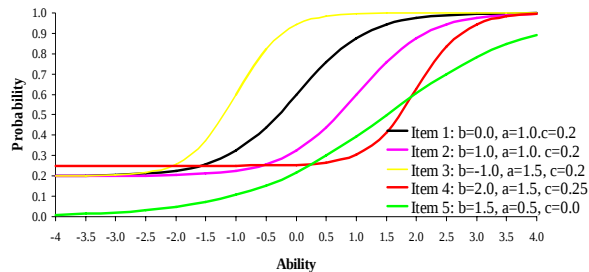
$$P(x_{ij} = 1 | \theta_j) = c_i + \frac{1 - c_i}{1.0 + \exp(-Da_i(\theta_j - b_i))}$$

- where
 - θ is student j ability
 - a_i is discriminating power for item i
 - b_i is item difficulty for item i , and
 - c_i is the "pseudo-guessing" parameter for item i

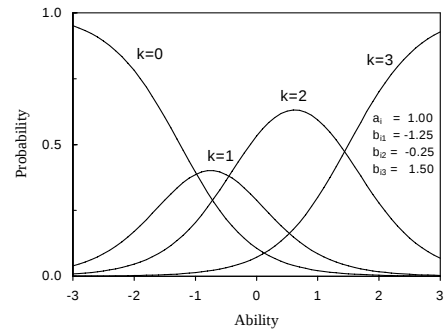
IRT Parameter Interpretation



Five ICCs



Graded Response Model: Category Response Functions (for item scores 0,1,2, and 3)



Common IRT Assumptions

- Unidimensionality: A single latent variable is sufficient to explain all responses to items.
 - Local Independence: At a given level of X , responses to all items are statistically independent.

$$P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n | \theta) = \prod_{i=1}^n P(U_i = u_i | \theta)$$

- Model-Data Fit
 - Appropriate use of an IRT model requires reasonable satisfaction of its assumptions.

Review of MML for Single Population

2PL IRT
$$P(x_{ij} = 1) = \frac{1}{1.0 + \exp(-Da_i(\theta_j - b_i))}$$

$$P(\underline{x}_j | \theta_j, \underline{a}, \underline{b}) = \prod_{i=1}^n P_i(\theta_j)^{x_{ij}} (1 - P_i(\theta_j))^{1-x_{ij}}$$

Integrate θ
$$P(\underline{x}_j | \underline{a}, \underline{b}) = \int_{\theta} P(\underline{x}_j | \theta, \underline{a}, \underline{b}) \cdot f(\theta) d\theta$$

Maximize likelihood
$$L(\underline{a}, \underline{b} | \underline{X}) = \prod_j L(\underline{a}, \underline{b} | \underline{x}_j)$$

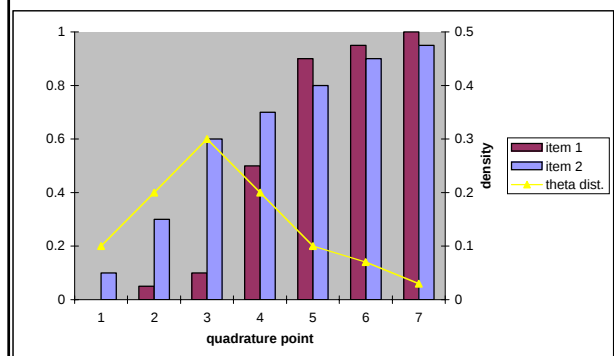
22

EM steps of Item calibration

- During E-step, posterior distribution as well as proportions correct for each items at each quadrature points are estimated based on the prior.
- During M-step likelihood function is maximized with respect to the model parameters.
- IRT linking is based on either proficiency distributions or the item parameters.

23

What E-step produces



Extend to Multiple Population

Integrate θ | group g $P(x_g | \alpha_g, g) = \int P(x_g | \theta, \alpha_g) f_g(\theta) d\theta$

Sum over groups $P(X | \alpha) = \sum_g \int P(x_g | \theta, \alpha_g) f_g(\theta) d\theta$

Maximize likelihood $L(\alpha | X) = \prod_j L(\alpha | x_j)$

25

Evaluate Fit of IRT Parameters

χ^2

• MD

• RMSD

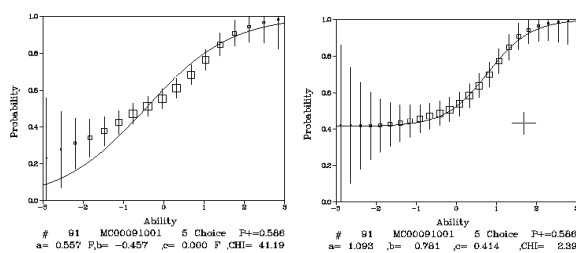
• Impact on ability estimate

$$MD_g = \int (\hat{P}_g(\theta) - P(\theta)) f_g(\theta) d\theta$$

$$RMSD_g = \sqrt{\int (\hat{P}_g(\theta) - P(\theta))^2 f_g(\theta) d\theta}$$

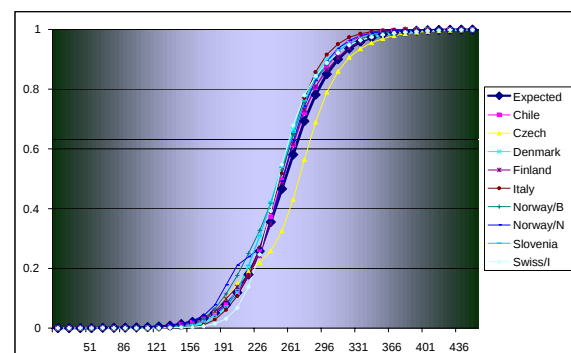
26

Assessing Model Fit: 1 & 3PL



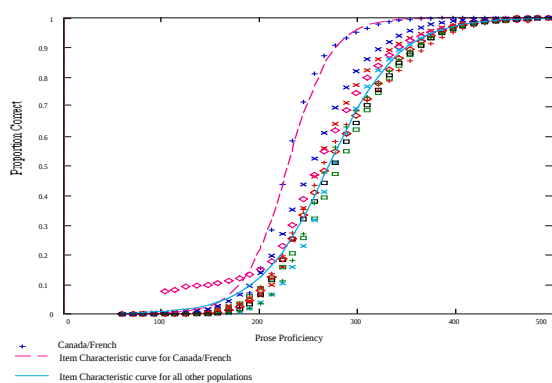
28

IRT Response Curve of Document 7-8



28

IRT Response Curve of Q- ITEM



29

Application of IRT Models to Test Development

- Item information functions
- Test information functions
- Standard error of measurement

Item Information

- Item information shows the contribution of an item to measurement precision with respect to ability parameter
- Important for test assembly, linking, virtual linking, assuring desired accuracy, so on

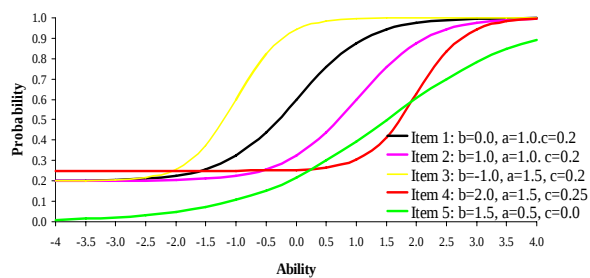
Item Information

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

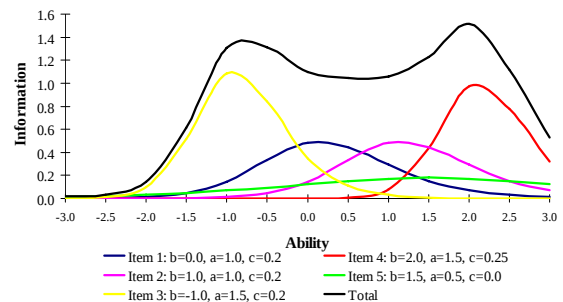
where

- $P_i(\theta)$ is the item response function,
- $Q_i(\theta) = 1 - P_i(\theta)$
- $P'_i(\theta)$ is the derivative of $P_i(\theta)$ with respect to θ

Item Characteristic Curves



Item Information Functions



Questions: Which abilities do these items measure best?
For what level is this test most appropriate?

Test Information

- Test information function is the sum of item information of the items in the test:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

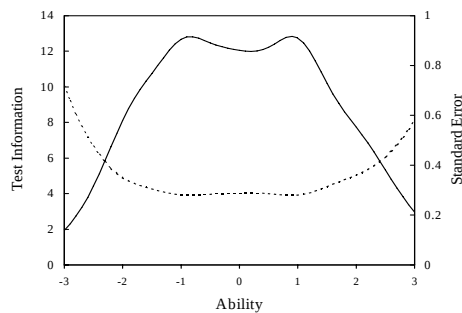
where n is the number of items in the test.

Standard Error

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

- Assesses measurement precision for each examinee level
- Used to calculate reliability

TIF & SE



Weights

TIMSS/PIRLS

$$FW = W_{\text{school}} * Ad_{\text{school}} * W_{\text{class}} * W_{\text{student}} * Ad_{\text{student}}$$

PISA

$$FW = Ad_{\text{subject}} * W_{\text{school}} * Ad_{\text{school}} * W_{\text{student}} * Ad_{\text{student}}$$

IALS/ALLS

$$FW = W_{\text{psu}} * W_{\text{household}} * Ad_{\text{household}} * W_{\text{respondent}} * Ad_{\text{respondent}} * Ad_{\text{post stratification}}$$

Booklet design

TIMSS: Bib-spiral of blocks of mixed subjects of Math(5) and Science(6), ~80 items,

PISA: Subject specific blocks, Reading, Math, Science.

IALS/ALL: Bib-spiral of 3 domains of Literacy, appended spiral design

39

Booklet Design TIMSS

- In 1999 TIMSS, 26 clusters of items were bib spiraled to produce 8 booklets

Booklet	Clusters																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	1	3	5					6										4	7						
2	2		1	3	5				6														4	7		
3	2			1	3	5				6										4	7					
4	2				1	3	5				6										4	7				
5	2	5				1	3	5				6										4	7			
6	2		5				1	3					6										4	7		
7	2	3		5										6										4	7	
8	2	1													6	3	5									

A(12m), B-H(12m), I-R(22m), S-V(10m), W-Z(10m)

40

Booklet Design 2003 PISA

Booklet	Cluster1	Cluster2	Cluster3	Cluster4
1	M1	M2	M4	R1
2	M2	M3	M5	R2
3	M3	M4	M6	PS1
4	M4	M5	M7	PS2
5	M5	M6	S1	M1
6	M6	M7	S2	M2
7	M7	S1	R1	M3
8	S1	S2	R2	M4
9	S2	R1	PS1	M5
10	R1	R2	PS2	M6
11	R2	PS1	M1	M7
12	PS1	PS2	M2	S1
13	PS2	M1	M3	S2

M1-M7,S1-S2,R1-R2,PS1-PS2: 30 min

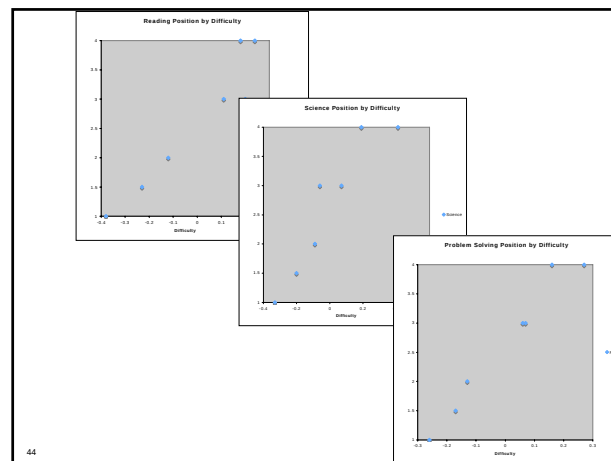
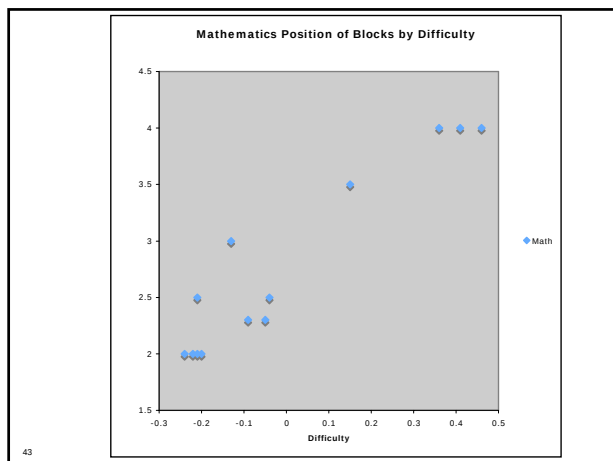
41

Booklet Design 2003 PISA

	M1	M2	M3	M4	M5	M6	M7	S1	S2	R1	R2	PS1	PS2
1	1	2		3						4			
2		1	2	3							4		
3			1	2	3							4	
4				1	2	3							4
5	4				1	2	3						
6		4				1	2	3					
7			4				1	2	3				
8				4				1	2	3			
9					4				1	2	3		
10						4				1	2	3	
11	3						4				1	2	3
12		3						4				1	2
13	2		3						4				1

M1-M7,S1-S2,R1-R2,PS1-PS2: 30 min

42



IALS: Average Percent Correct; Block by Block Order

Block Order	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Total
First	72	78	74	72	77	83	77	76
Second	78	76	77	78	72	83	79	78
Last	74	77	77	77	77	83	72	77

ALL :Average Proportion Correct; Block by Block Order

Block Order	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Total
First	62.6	64.8	68.9	66.2	64.8	63.1	60.1	48.7	62.4
Second	61.6	65.0	68.3	64.8	65.0	63.3	59.9	5.01	62.3

Booklet Design IALS/ALL

	b1	b2	b3	b4	b5	b6	b7	b8
b1		X	X		X			X
b2			X	X		X	X	
b3	X			X	X			X
b4	X	X				X	X	
b5		X		X		X		
b6	X		X		X			
b7	X		X					X
b8		X		X			X	

1-4: Prose&Document 5,6: Numeracy 7,8: Problem Solving

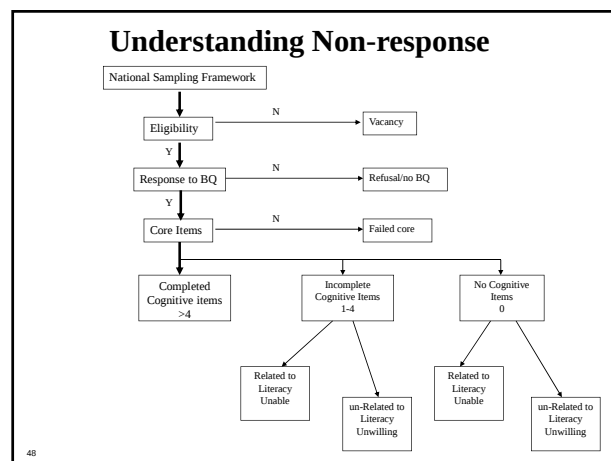
46

Booklet Design IALS/ALL

	B1	B2	B3	B4	B5	B6	B7	B8
1	1	2						
2	1	2						
3		1	2					
4		1		2				
5			1	2				
6	2		1		2			
7	2			1				
8		2			1			
9	1					2		
10		1				2	2	
11			1		2			
12				1		2		
13		2			1			
14				2	1			
15	2					1		
16			2				1	
17					1	2		
18					2	1		
19	1							2
20		1					2	
21			1				2	2
22				1				1
23	2						1	
24		2	2					1
25				2				1
26					2			
27						1	2	
28						2	1	

1-4: Prose&Document 5,6: Numeracy 7,8: Problem Solving

47



Scoring Open Ended Questions

TIMSS: 25-40% open ended, extensive training, sample booklets scoring.

PISA: 30% open ended, extensive training, sample booklets scoring.

IALS/ALLS: 100% open ended, extensive training, exchange booklets across countries to sample booklets scoring, correct scoring if necessary.

49

IRT Modeling/Scaling

TIMSS: moved from Rasch(95) to 3pl+PC(99 on)

Common item parameters for all countries

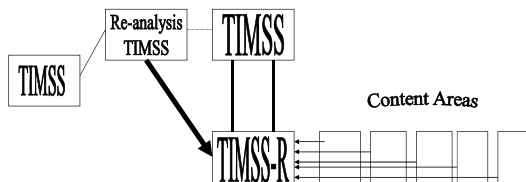
PISA: Multi dimensional Rasch, limited 2pl, Common item parameters for all countries

IALS/ALL: 2pl IRT, Subset of items may be different from common item parameters for

50

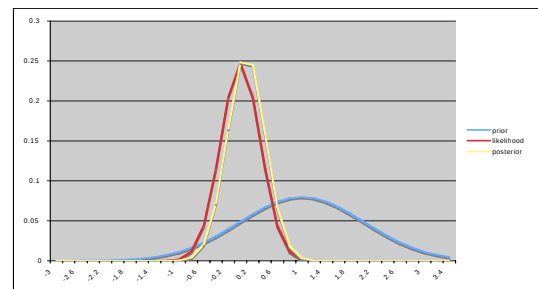
Trend linkings Example

- **Re-analysis of 1995 TIMSS**
 - To address 1PL-3PL, and re-center to mean of 500, and sd. of 100 based on grade 8 only
- **Common scale of 1995 TIMSS-1999TIMSS-R**
 - Trend analysis
- **Content area of each domain 1999TIMSS-R**
 - Profile of content areas by setting the distributions to TIMSS-R



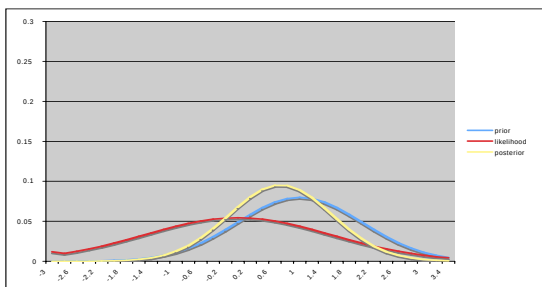
51

Impact of Prior Distribution on Proficiency Estimate 1



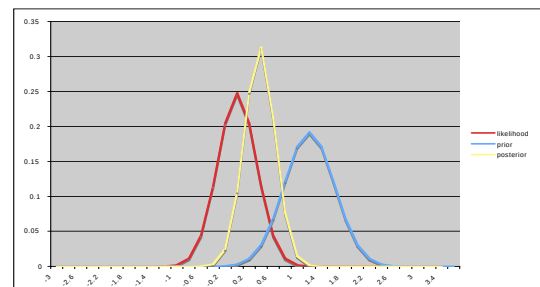
52

Impact of Prior Distribution on Proficiency Estimate 2



53

Impact of Prior Distribution on Proficiency Estimate 3



ETS Listening Learning Leading

Modeling International Large Scale Survey Assessment Data

IERI Fall Academy 2008
Matthias von Davier
Educational Testing Service

On international large scale educational surveys

- International large scale assessments
- Assessment designs in large scale assessments
- Background data in international assessments
- Deriving skill distributions in subgroups of interest
- Research on international surveys (the IERI ad)
- Outlook

ETS Listening Learning Leading

(International) Large Scale Assessments

Some key principles:

- Common assessment framework
- Common scale(s) across countries & languages
- Coordinated administrations
- Repeated, linked assessments
- Group level reporting
- Curriculum (coverage?) or life-skills based
- International comparisons

ETS Listening Learning Leading

(International) Large Scale Educational Assessments

The mother of all Large Scale Assessments:
National Assessment of Educational Progress

- NAEP (..., 2008, ...), government mandated since 1960ies

International Large Scale Assessments

- TIMSS (1995, 1999, 2003, 2007, 2011, ...)
- PIRLS (1992, 2001, 2011, ...)
- PISA (2000, 2003, 2006, 2009, ...)
- IALS ('94, '96, '98), ALL (2003)
- PIAAC (2011, ...)

ETS Listening Learning Leading

For example: TIMSS

TIMSS (1995, 1999, 2003, 2007, 2011, ...)

- Trends in Mathematics and Science Study
- An IEA study (IEA? - see next slide...)
- Currently 60-some countries participating
- International Study Center (ISC), Boston College
- ETS provides psychometric consulting and statistical software for ISC's analyses & reporting

ETS Listening Learning Leading

IEA IEA

Page 1 of 3

IEA
International Association for the Evaluation of Educational Achievement

IEA

From the IEA Headquarters
Headquarters & Branches
Mission Statement
Council Members & Staff
Large Scale
IEA Publications
IEA Council Assembly
Standing Committee
Executive Board
Data Processing and Research Center
IEA-ETS Research Institute
Global Partnership

STUDIES
Learning on IEA Study
Current Studies
Completed Studies
IEA Studies Evaluation
IEA Studies Database and Data Analysis

IEA-ETS Academy: 11-19 May 2008 ••• PIRLS 2006 results released ••• Page 1 of 3

CCR 2006
International Civic and Citizenship Education Study

The purpose of the International Civic and Citizenship Education Study (ICCS) is to investigate the ways in which young people are prepared to undertake their roles as citizens.

Current Studies
Learning

TIMSS Advanced 2008
Replication of TIMSS 1995 Advanced Mathematics and Physics

First administered in 10 countries in 1995, TIMSS Advanced assesses school/learning objectives with special preparation in advanced mathematics and physics. Countries that...

ETS 2008 REAP RELEASED ON
PIRLS 2006 REAP
IEA ADVANCEMENT STUDY EVAL
THE 2008 IEA
PIRLS AND TIMSS
4
LATEST PUBLIC
PIRLS 2006
PIRLS 2006

<http://www.iesl.nl/>

4/25/2008
ETS Listening Learning Leading

For example: PISA

PISA (2000, 2003, 2006, 2009, ...)

- Programme for International Student Assessment
- An OECD study, 67 countries in 2009...
- Australian Council of Educational Research (ACER) is conducting the study for OECD
- ETS provides consulting on technical issues, ETS representative heads the reading literacy expert group, ...



7

Liberalism Learning Leading

- Programme for International Student Assessment
- An OECD study, 67 countries in 2009...
- Australian Council of Educational Research (ACER) is conducting the study for OECD
- ETS provides consulting on technical issues, ETS representative heads the reading literacy expert group, ...



For example: PIAAC

PIAAC (2011, ...)

- Programme for the International Assessment of Adult Competencies (PIAAC)
- An OECD study, 20-some countries in 2011...
- TIMSS, PISA, NAEP are school based assessments, and sample students (or classes) within schools
- PIAAC (like IALS, ALL) is a household survey, and samples adults 16-65 year old from households



9

Listening Learning Leading

- Programme for the International Assessment of Adult Competencies (PIAAC)
- An OECD study, 20-some countries in 2011...
- TIMSS, PISA, NAEP are school based assessments, and sample students (or classes) within schools
- PIAAC (like IALS, ALL) is a household survey, and samples adults 16-65 year old from households



Assessment designs in large scale assessments

- All domains are covered broadly, the number of sub-domains and items is determined during framework development
- A large number of items is developed and organized into blocks, which are combined into test booklets.
- Test booklets include an incomplete, but balanced subset of all possible combinations of the number of blocks administered
- -> Balanced incomplete block (BIB) design

11

Listening. Learning. Leading.

- | Exhibit 11: TIMSS 2007 Student Achievement Booklet Design –
Fourth and Eighth Grades | | | | |
|---|-------------------|-----|--------|-----|
| Student Achievement Booklet | Assessment Blocks | | | |
| | Part 1 | | Part 2 | |
| Booklet 1 | M01 | M02 | S01 | S02 |
| Booklet 2 | S02 | S03 | M03 | M03 |
| Booklet 3 | M03 | M04 | S03 | S04 |
| Booklet 4 | S04 | S05 | M04 | M05 |
| Booklet 5 | M05 | M06 | S05 | S06 |
| Booklet 6 | S06 | S07 | M06 | M07 |
| Booklet 7 | M07 | M08 | S07 | S08 |
| Booklet 8 | S08 | S09 | M08 | M09 |
| Booklet 9 | M09 | M10 | S09 | S10 |
| Booklet 10 | S10 | S11 | M10 | M11 |
| Booklet 11 | M11 | M12 | S11 | S12 |
| Booklet 12 | S12 | S13 | M12 | M13 |
| Booklet 13 | M13 | M14 | S13 | S14 |
| Booklet 14 | S14 | S01 | M14 | M01 |

Assessment Blocks				
Student Achievement Baseline	Part 1		Part 2	
Baseline 1	M01	M02	S01	S02
Baseline 2	S02	S03	M02	M03
Baseline 3	M03	M04	S03	S04
Baseline 4	S04	S05	M04	M05
Baseline 5	M05	M06	S05	S06
Baseline 6	S06	S07	M06	M07
Baseline 7	M07	M08	S07	S08
Baseline 8	S08	S09	M08	M09
Baseline 9	M09	M10	S09	S10
Baseline 10	S10	S11	M10	M11
Baseline 11	M11	M12	S11	S12
Baseline 12	S12	S13	M12	M13
Baseline 13	M13	M14	S13	S14
Baseline 14	S14	S01	M14	M01

LSA frame summary

- Students respond to 3-6 short blocks of items, blocks appear in multiple locations within multiple booklets
- Multiple test forms have to be linked, blocks from previous assessments are combined with new blocks
- Trends over time often have to be reported
- Differences between countries in ability distributions, but no DIF is a basis for comparability
- Sample items (released old form items) have to be provided to stakeholders.

N9. Which of the boxes X, Y, or Z has the LEAST mass?



- A. X
- B. Y
- C. Z
- D. All three boxes have the same mass.

N-9

Multiple LSA components

In addition to the cognitive items administered, students also fill out questionnaires, as do their teachers or principals

Students are asked about a variety of issues related to their academic achievement, their leisure activities, etc.

Teachers may be asked about their instructional practices, and their assessment of the school environment

School principals may be asked about school infrastructure, and teacher professional development

Background Data in Skill Surveys

Background data serve (at least) two purposes in international surveys of skills:

- BV's are utilized in modeling the population distribution of skills. BV's help reducing uncertainty as well as biases of group comparisons
- Background variables (BV) are central in understanding differences between higher and lower skilled groups of the target population

OECD Programme for International Student Assessment 2006

OECD
PISA

Canada

STUDENT QUESTIONNAIRE

PISA SCHOOL ID [] [] [] [] [] []
PISA STUDENT ID [] [] [] [] [] []

Background Data for Modeling

Background variables (BV) are needed as auxiliary information in a population model used to reduce uncertainty of measurement and to provide unbiased imputations of skill variables:

- BV are selected and preprocessed to be integrated in a latent regression model (mixed effects model, multilevel IRT model)
- Conditional distributions of skills given BV are estimated, and imputations (plausible values) are drawn for reporting group statistics and variance estimation

Background Data as Major Reporting Variables

Background variables (BV) are central to reporting, and are used to derive indicators for reporting skills in subgroups of interest to stakeholders and policy makers.

- BV may directly assess the group definition of interest (gender, ethnicity)
- BV may be aggregated to indicator variables such as attitudes towards technology, reading behaviors, etc.
- BV may be proxies to variables we cannot directly assess or we are now allowed to ask in some or all countries (SES, gross income)



19

Listening Learning Leading

Challenges of using Background Data

Background variables are selected based on experts knowledge of relevant differences. In international surveys, this selection is faced with multiple challenges:

- Comparability across countries:
 - A variable that is quite predictive of skills in one country may or may not be so in other countries
 - The response formats used can be subject to cultural differences in response sets (e.g. avoiding extreme responses vs. tendency to avoid middle/undecided responses)
- Relationships between questions, or collections of questions used to derive indicators, may vary across countries.
- Some background variables very useful as a proxy to SES in one country may not vary much in some other countries (number of bathrooms in a house in the USA vs. Western Europe)



20

Listening Learning Leading

Other issues in LSA, not covered today...

- Translation into multiple target languages, translation verification
- Target population, definition, inclusions, exclusions, sampling of students within educational institutions
- Scoring constructed response items, rater agreement, cross country comparability of rating standards
- Weighting, non-response adjustments, replacement, response rate, design effect...



21

Listening Learning Leading

Putting it all together:

There are many pieces to put together:

- Cognitive domain task responses
- Complex sample of item responses
- Countries provide a representative sample as the basis of international comparisons
- Assume we have clean data, received in time from all participating countries:

How are the different sources of information coming together, and what do the results look like?



22

Listening Learning Leading

Modeling large scale assessment data

- Reported are student proficiency distributions by countries and in subgroups within country
- IRT based methods are used due to the level of complexity of the data
- A multilevel IRT model with predictors based on the background data is used to analyze the data
 - Often called conditioning or latent regression model
- Reporting uses imputations (plausible values) to generate distributions of proficiencies, given reporting (major background) variables.



23

Listening Learning Leading

Latent regression IRT

The model LR-IRT represents is a latent regression model or a multi-level, better: 2-level, IRT model:

- The first level is the measurement model, a multidimensional, simple-structure IRT model
- The second level is a linear regression of the (multidimensional) latent trait θ on covariates, referred to as background variables in large scale assessments



24

Listening Learning Leading

LR-IRT Measurement model, 1

The LR-IRT measurement model is

- A composition of standard one - dimensional IRT models for multiple scales.
- Each item measures only one scale
- That means simple structure
- Which implies a factored likelihood



25

Listening Learning Leading

LR-IRT Measurement model, 2

- For dichotomous multiple choice items, often the 3PL is used
- For dichotomous constructed response items, the 2PL should be used
- For polytomous items, the generalized partial credit model (GPCM) is used
- One can use the Rasch model as well (Conquest, PISA & TIMSS 1995)



26

Listening Learning Leading

LR-IRT Measurement model, 3

- IRT estimation is carried out before the latent regression is estimated
- -> item parameters are considered as fixed and known constants in LR-IRT
- Adams, Wilson and Wu show that impact is minimal for the Rasch model case
- Patz & Junker talk about this as adopting a *divide and conquer* strategy



27

Listening Learning Leading

LR-IRT Estimation

We use the (EM) algorithm, a standard tool used a lot for IRT and other latent variable model estimation problems

- E-step: Estimate posterior means and posterior (co-)variances
- M-step: Maximum likelihood estimates of the latent regression parameters



28

Listening Learning Leading

LR-IRT Estimation, EM-cycle

The E-Step produces preliminary estimates of the posterior means and variances given previous regression parameters

The M-step produces preliminary estimates of the regression parameters given the previous E-steps estimates of posterior moments

This cycle is repeated until parameters converge or a predetermined maximum of cycles is reached



29

Listening Learning Leading

Drawing Plausible Values

- NAEP and other survey assessments report subgroup estimates, not individual scores
- Covariates are used to incorporate reporting variables and other auxiliary information into a comprehensive model where they help avoiding biased group estimates
- NAEP uses multiple imputations derived from the LR-IRT model estimates to generate subgroup means, variances and other statistics



30

Listening Learning Leading

Plausible Values (PVs)

- Plausible values (PV's) are multiple imputations drawn from examinee posterior distributions
- Plausible values together with sampling weights are also used to estimate variances of the reported statistics
- This is done outside of LR-IRT, LR-IRT just generates the PVs using the regression estimates relating proficiency to background data



31

Listening. Learning. Leading.

Summary

- Survey assessments collect student proficiency data using sparse responses from multiple booklets
- In addition, multiple other sources of information are available on student background
- Latent regression models integrate IRT based proficiency estimation with population models using background data
- Reporting is based on group-level statistics aggregated using major background variables like gender, ethnicity, school type, parents education...



32

Listening. Learning. Leading.

Latent Regressions in DESI / MGROUP & Drawing Plausible Values

IERI Fall Academy 2008
Matthias von Davier
ETS
Research & Development

ETS Confidential & Proprietary

Overview

- Latent regression in a nutshell
 - i.e., DESI / MGROUP in a nutshell
- Measurement model, Population model
- Estimation
- Drawing plausible values
- Outlook

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP in a nutshell

The model estimated by DESI is a latent regression model or a multi-level, better: 2-level, IRT model:

- The first level is the measurement model, a multidimensional, simple-structure IRT model
- The second level is a linear regression of the (multidimensional) latent trait θ on covariates, referred to as background variables in NAEP

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

Let $x_1, \dots, x_I$ denote item responses, θ the latent trait, and let the $z_1, \dots, z_K$ denote the covariates:

- The IRT model donates:
 $P(x_1 \dots x_I | \theta) = \prod_{i=1}^I P(x_i | \theta)$
- The latent linear regression donates:
 $f(\theta | z_1, \dots, z_K) = \phi(\theta | \mu(z_1 \dots z_K), \Sigma)$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

The marginal probability of observed responses $x_1, \dots, x_I$ given the $z_1, \dots, z_K$ covariates is then

$$P(x_1 \dots x_I | z_1 \dots z_K) = \int_{\theta} \left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_1 \dots z_K), \Sigma) d\theta$$

The $z_1, \dots, z_K$ determine the prior mean $\mu(z_1 \dots z_K)$ of the ability distribution, Σ denotes the common (co-)variance (matrix), $\phi(\cdot)$ denotes the (multivariate) normal density.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

Bayes' theorem is applied to find the posterior distribution of θ given observed responses $x_1, \dots, x_I$ and the $z_1, \dots, z_K$ covariates

$$P(\theta | x_1 \dots x_I, z_1 \dots z_K) = \frac{\left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_1 \dots z_K), \Sigma)}{P(x_1 \dots x_I | z_1 \dots z_K)}$$

The posterior distribution of θ is central to many subsequent calculations, it is used to calculate posterior means and variances (and finally PV's)

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI Measurement model, 1

The DESI measurement model is

- A composition of standard one - dimensional IRT models for multiple scales.
- Each item measures only one scale
- That means simple structure
- Which implies a factored likelihood

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

7

DESI Measurement model, 2

- For dichotomous multiple choice items, often the 3PL is used
- For dichotomous constructed response items, the 2PL should be used
- For polytomous items, the generalized partial credit model (GPCM) is used
- One can use the Rasch model as well (Conquest, PISA, and TIMSS in 1995)

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

8

DESI Measurement model, 3

- IRT estimation is carried out before the DESI latent regression is estimated
- -> item parameters are considered as fixed and known constants in DESI, not as estimated quantities
- Adams, Wilson and Wu show that impact is minimal for the Rasch model case
- Patz & Junker talk about adopting a divide and conquer strategy

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

9

This is roughly DESI Phase 1

- Read in item parameters from previous IRT estimation, use as fixed constants
- Read in observed responses and covariates for each examinee
- Calculate likelihood $L(x_1..x_I|\theta) \sim P(x_1..x_I|\theta)$ for each examinee, store for next phase
- Calculate covariance matrix of covariates, $KOV(z_1..z_K)$ and store for next phase

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

10

DESI Population model, 1

- The population model describes the relationship between covariates (background variables) and the ability distribution.
- In DESI it is assumed that a latent regression can be used to describe this relation

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

11

DESI Population model, 2

The regression is latent since the dependent variable is not directly observed

Recall that we only have:

$$P(\theta | x_1..x_I, z_1..z_K) = \frac{\left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_1..z_K), \Sigma)}{P(x_1..x_I | z_1..z_K)}$$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

12

DESI Population model, 3

Instead of an observed dependent variable, we use the posterior mean and posterior variance in the regression estimation

$$E(\theta^q | x_1 \dots x_I, z_1 \dots z_K) = \int_{\theta} \theta^q P(\theta | x_1 \dots x_I, z_1 \dots z_K) d\theta$$

$$V(\theta | x_1 \dots x_I, z_1 \dots z_K) = E(\theta^2 | x_1 \dots z_K) - (E(\theta | x_1 \dots z_K))^2$$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

13

DESI Population model, 4

For the regression in question, we have the usual system that relates a linear combination of the covariates for each examinee n to the posterior mean of this examinee

$$E(\theta | x_{n1} \dots x_{nI}, z_{n1} \dots z_{nK}) = \sum_{k=1}^K \beta_k z_{nk} + \varepsilon_n$$

The β_k are regression coefficients that have to be estimated, which is roughly what Phase II of DESI does.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

14

DESI Estimation

DESI uses an Estimation – Maxi-mization (EM) algorithm, a standard tool used a lot for IRT and other latent variable model estimation problems

- E-step: Estimate posterior means and posterior (co-)variances
- M-step: Maximum likelihood estimates of the latent regression parameters

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

15

DESI Estimation, E-step

E-step estimates posterior means and posterior (co-)variances:

- Calculation of posterior moments requires potentially multidimensional (numerical) integration
- There are many ways to do that, approximate solutions, smart integration methods, not so smart brute force methods or Monte- Carlo methods
- More about different E-step versions next time

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

16

DESI Estimation, M-step

The M-step in DESI uses posterior means and covariates to estimate regression parameters. It is assumed that we can use the OLS estimator.

- OLS estimator is $\beta = (Z^T Z)^{-1} Z^T \theta$, with posterior mean vector (matrix) θ
- The Σ estimator is the sum of the regression error variance $V(\varepsilon)$ and the average posterior variance $E(V(\theta | x_1 \dots z_k))$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

17

DESI Estimation, EM-cycle

The E-Step produces preliminary estimates of the posterior means and variances given previous regression parameters

The M-step produces preliminary estimates of the regression parameters given the previous E-steps estimates of posterior moments

This cycle is repeated until parameters converge or a predetermined maximum of cycles is reached

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

18

EM is roughly DESI Phase II

The EM cycles can be time consuming, but modern hardware allow samples of 200,000 students or more with 500 background variables and 5 math scales:

- Computational cost is not an issue any more, large samples of students are fine
- More of an issue are small student samples with large numbers of background variables, numerical instabilities can result.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

19

Drawing Plausible Values

- NAEP and other survey assessments report subgroup estimates, not individual scores
- Covariates are used to incorporate reporting variables and other auxiliary information into a comprehensive model where they help avoiding biased group estimates
- NAEP uses multiple imputations derived from the DESI model estimates to generate subgroup means, variances and other statistics

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

20

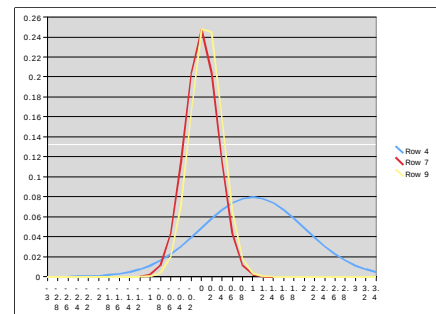
Plausible Values ~ Phase III

- Plausible values (PV's) are multiple imputations drawn from examinee posterior distributions
- Plausible values together with sampling weights are also used to estimate variances of the reported statistics
- This is done outside of DESI, DESI just generates the PV's using the posterior moments we talked about

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

21

Recall $\text{Prior} \times \text{Likelihood} = \text{Posterior}$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

22

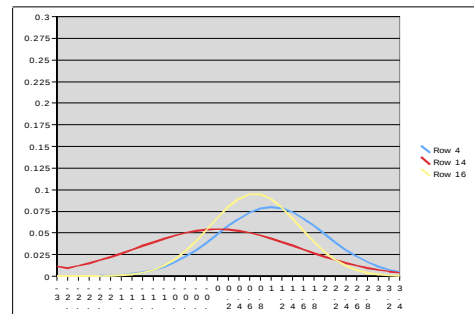
Testing versus Survey Assessments

- Individual testing provides a narrow likelihood based on many items, provides a lot of information
 - prior does not matter much
 - population needed neither needed nor wanted
- Survey assessments use fewer items, the likelihood is broader, represents less information
 - priors matter, covariates of proficiency are useful
 - population model providing priors is essential

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

23

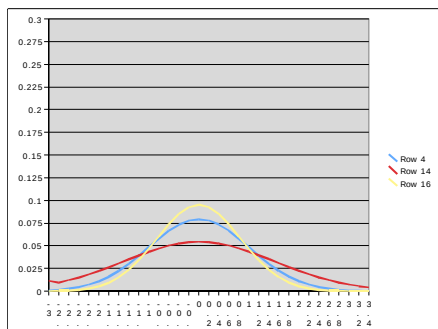
Ed. Surveys: Prior Information, $M=1.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

24

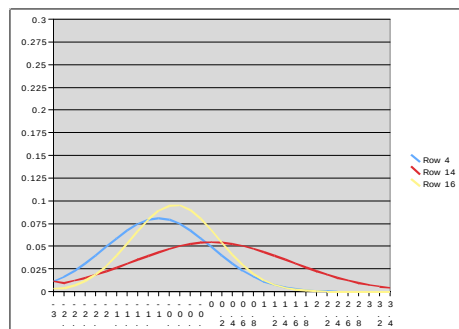
Ed. Surveys: Prior Information, $M=0.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

25

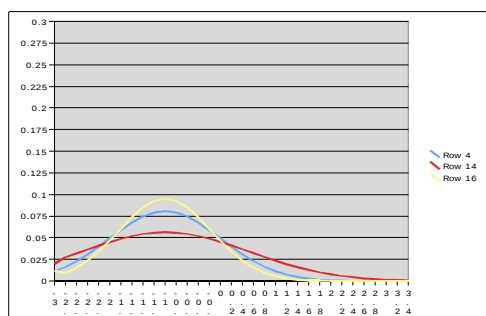
Ed. Surveys: Prior Information, $M=-1.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

26

Prior Information, $M=-1.0$, and Likelihood $M=-1.0$ (below average Joe)



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

27

!!! Priors are estimated, not fixed !!!

- Many average-Joe cases go into the prior
- Easy(?) to see in multiple-group models
- In latent regressions, the average effect of background variable on posterior is estimated
- Likelihood and prior do not need to agree
 - If not, posterior moves away, given information
 - If in agreement, prior*likelihood works like a charm
- 'borrowing information' from similar cases is the key concept

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

28

Plausible Values ~ Phase III

- *DESI provides estimates of subgroup means based on PVs for user defined grouping variables*
- *PVs are a means to an end, not optimal for anything, but very good for many types of analyses*
- *Many background variable effects are represented in PVs, allows analyses using these without bias*

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

29

Summary

- Latent regression model extends IRT
- Allows the estimation of effects of background variables
- Provides average effects based on sample correlations between θ and BQ
- Used for prior distributions, borrowing information from similar cases

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

30

出席國際學術會議心得報告

計畫編號	NSC 97-MOE-S-003-003
計畫名稱	國際數學與科學教育成就趨勢調查2007 (4/4) 甲
出國人員姓名 服務機關及職稱	任宗浩 國立台灣師範大學科學教育中心 助理研究員 李哲迪 國立台灣師範大學科學教育中心 助理研究員
會議時間地點	2009 年 5 月 26 日- 5 月 28 日 Singapore
會議名稱	Analysis of Large Scale Survey Data

一、 與會歷程

此工作坊為 ACER (Australian Council for Educational Research) 所主辦。ACER 是澳洲非常重要的教育評鑑機構，同時也是負責統籌執行 PISA (The Programme for International Student Assessment) 的單位。此工作坊之目的在介紹大型調查資料的統計分析方法。工作坊的講者是 ACER 的資深研究員—Wolfram Schulz 博士。工作坊分三天進行（議程如下表）。第一天的主題是取樣誤差。第二天的主題是運用似真值進行資料分析。第三天的主題是階層線性模式分析。工作坊的每個場次都是先介紹概念和理論，然後實做分析，交錯進行。

Tuesday 26 May, 2009

09.30 – 11.00 Introduction and Overview

Presentation: Sampling and weighting

11:00 – 11:30 Coffee break

11:30 – 13:00 Presentation: Estimation of sampling variance

Practice session: Calculating sampling variance

13:00 – 14.00 Lunch break

14.00 – 15.00 Presentation: Using SPSS for the estimation of sampling variance

Practice session: Learning how to use SPSS module

15:00 – 15:15 Short coffee break

15:15 – 16.00 Practice session: Reproducing tables from report in international studies

Wednesday 27 May, 2009

09.30 – 10.30	Presentation: Scaling methodologies used in international studies Practice session: Estimation of combined measurement and sampling error
11:00 – 11:30	Coffee break
11:30 – 13:00	Practice session: Analysis of plausible values
13:00 – 14.00	Lunch break
14.00 – 15.00	Presentation: Regression analysis
15:00 – 15:15	Short coffee break
15:15 – 16.00	Practice session: Estimating regression models with large scale survey data

Thursday 28 May, 2009

09.30 – 11.00	Concepts: Multi-level modelling Introduction to software packages for multi-level modelling
11:00 – 11:30	Coffee break
11:30 – 13:00	Practice session: Estimating multi-level models with large scale survey data
13:00 – 14.00	Lunch break
14:00 – 15.00	Practice session: Comparing multi-level and single-level regression results
15:00 – 15:15	Short coffee break
15.15 – 16.00	Conclusion and open discussion

二、 與會心得

此工作坊最大的特色是 ACER 所開發的統計分析軟體—Replicates。此軟體是專為大型調查資料分析而設計的。在運用大型調查研究所蒐集的資料進行統計推論時，所有統計量的標準誤都必須考慮測量誤差和抽樣誤差。以 TIMSS 和 PISA 的資料為例，其測量誤差以五個似真值來體現，而抽樣誤差則要透過 Jackknife 的方法來估計。為了取得測量誤差和抽樣誤差的估計，相同的統計估計的計算要重複執行將近 80 次。在實際的操作上，需要撰寫巨集語法來完成重複的計算。Replicates 的功能就是在產生 SPSS 的巨集語法，以計算諸如平均值、百分比、相關係數、迴歸方程的參數等等統計量的標準誤。一般研究者對巨集語法的撰寫並不熟悉，就算熟悉該語法，每次分析都要撰寫巨

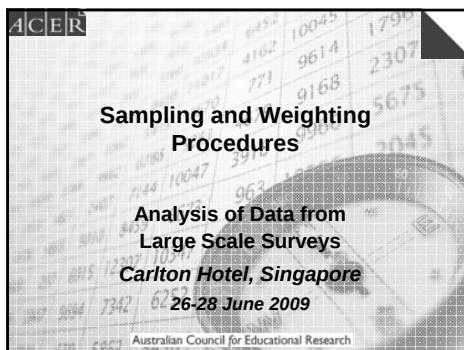
集程式也是很方便。有了此一整合在 SPSS 當中的外掛程式，在分析上就極為便利。

雖然 Replicates 很方便，但是它主要是針對 PISA 的資料而設計的，內定的變數名稱的命名規則是 PISA 資料庫的命名規則。如果要以 Replicates 來分析 TIMSS 的資料，則需要修改它所產生的巨集程式檔。不過，IEA 也發展了一套類似的軟體，稱為 IDB Analyzer。這兩套軟體可在 ACER 和 IEA 的官方網站上直接下載。比較兩套軟體，Replicates 較有彈性，考慮的分析項目較為周全。

綜觀國內研究，研究者進行大型調查資料二次分析時，一般皆無考慮測量誤差和抽樣誤差，統計量的標準誤都被低估了。在此情況下，統計分析的計算結果雖然顯示有顯著差異，但由於真實的標準誤可能很大，因此其實並沒有顯著。學者目前的作法或許是因為手邊沒有方便的統計分析套件所致，此工作坊所介紹的 Replicates 若能引介給國內學者，相信對國內二次分析此研究領域的發展必有極大的助益。

三、 攜回資料名稱及內容

1. 各場次講者提供之簡報檔及習作資料。
2. Replicates (V 6.2.1) 程式軟體一套。



Topics of this presentation

- Sampling in large-scale data surveys
- Computation of sampling weights
- How to apply replication methods

Sample versus Census - 1

- Surveys usually draw a sample because a census:
 - Requires unavailable budgets
 - Time consuming and thus incompatible with deadlines
 - Does not necessarily bring additional and required information

Sample versus Census - 2

- Several ways of drawing a sample
- Two major principles:
 - Avoiding bias in the selection
 - Randomness
 - Sampling frame covers the population adequately
 - Ensuring the varying probabilities of selection are reflected $\Rightarrow$ weights
 - Achieving maximum precision, given a set of constraints (time, funds, and so on)

Selecting a simple random sample

- An SRS gives an equal probability of selection for each n selected from population N which is

$$p_i = \frac{n}{N}$$

- and each selected n has a weight of

$$w_i = \frac{1}{p_i} = \frac{N}{n} = \frac{400}{40} = 10$$

Using weights with SRS

- Since the weights are the same for all selected members of the sample, weighted and unweighted means are identical:

$$\hat{\mu}_{(X)} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \frac{w_i \sum_{i=1}^n x_i}{w_i \sum_{i=1}^n 1} = \frac{\sum_{i=1}^n x_i}{n}$$

Using weights with SRS

- The sum of the weights always rebuilds the size of the population

$$\sum_{i=1}^n w_i = \sum_{i=1}^n \frac{N}{n} = N$$

Student surveys

- Simple random sampling generally too expensive and often impossible.
 - Often no central database for sampling of individual students
 - Use of SRS is impractical for survey administration.
 - No link of students to classrooms or schools available.
- Therefore, two-stage designs common, where schools are selected at first stage and students or intact classrooms at the second.

Two stage sample

- n schools are selected from a population of N schools
 - $\Rightarrow$ what is the probability of a particular school to be drawn?
- Within school i , n_i students or n_i classrooms are drawn from a population of N_i students or N_i classroom
 - $\Rightarrow$ what is the probability of a particular student in school i to be drawn?

An example of a two stage sample

- Let us imagine:
 - a population of 10 schools
 - Within any school, there are 40 students
 - So the educational system counts 400 students
- 4 schools are drawn according to a SRS procedure
- 10 students within the 4 selected schools are selected according to a SRS

Two-stage probabilities

- Four schools selected out of 10

$$p_{1_i} = \frac{n_{sc}}{N_{sc}} = \frac{4}{10} = 0.4$$

- 10 students selected within each school

$$p_{2_{ij}} = \frac{n_i}{N_i} = \frac{10}{40} = 0.25$$

Final weights (SRS two-stage)

- The final selection probability is

$$p_{ij} = p_{1_i} p_{2_{ij}} = \frac{n_{sc} n_i}{N_{sc} N_i}$$

- and the final total weight is the product of school and student weight or the inverse of the selection probability:

$$w_{ij} = \frac{1}{p_{ij}} = \frac{1}{0.1} = 10$$

Weights summary

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	40							
2	40	0.4	2.5	0.25	4	0.1	10	100
3	40							
4	40							
5	40	0.4	2.5	0.25	4	0.1	10	100
6	40							
7	40	0.4	2.5	0.25	4	0.1	10	100
8	40							
9	40							
10	40	0.4	2.5	0.25	4	0.1	10	100
Total			10					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10							
2	15	0.4	2.5	0.66	1.5	0.27	3.75	37.5
3	20							
4	25							
5	30	0.4	2.5	0.33	3	0.13	7.5	75
6	35							
7	40	0.4	2.5	0.25	4	0.1	10	100
8	45							
9	80							
10	100	0.4	2.5	0.1	10	0.04	25	250
Total	400		10					462.5

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10	0.4	2.5	1	1	0.4	4	40
2	15	0.4	2.5	0.66	1.5	0.27	3.75	37.5
3	20	0.4	2.5	0.5	2	0.2	5	50
4	25	0.4	2.5	0.4	2.5	0.16	6.25	62.5
Total			10					190

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
7	40	0.4	2.5	0.250	4	0.10	10.00	100.0
8	45	0.4	2.5	0.222	4.5	0.08	11.25	112.5
9	80	0.4	2.5	0.125	8	0.05	20.00	200.0
10	100	0.4	2.5	0.100	10	0.04	25.00	250.0
Total			10					662.5

From SRS to PPS Sampling

- As school sizes differ across schools, then:
 - student weights will vary across schools.
 - In smaller schools, probability of selection is higher.
 - In larger schools, probability of selection is lower.
 - Very large uncertainty on the student population estimates $\Rightarrow$ population estimates will have a large sampling variance

Sampling with school probabilities proportional to size (PPS)

- Larger schools get a higher probability of selection than smaller schools...
- ...but students in larger schools have a smaller within-school probability of being selected than students in small schools, and...
- the school and within school weights differ between schools, but final student weights do not vary ('self-weighted sample').

School weights with PPS sampling

- Probability of a school to be selected is equal to the ratio of the school size multiplied by the number of schools to be sampled and divided by the total number of students in the population:

$$p_{1_i} = \frac{N_i * n_{sc}}{N}$$

PPS sample of schools

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10							
2	15							
3	20	0.2	5.00	0.500	2.0	0.1	10	100
4	25							
5	30							
6	35							
7	40	0.4	2.50	0.250	4.0	0.1	10	100
8	45							
9	80	0.8	1.25	0.125	8.0	0.1	10	100
10	100	1	1.00	0.100	10.0	0.1	10	100
Total	400		9.75					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10	0.1	10	1	1	0.1	10	100
2	15	0.15	6.7	0.67	1.5	0.1	10	100
3	20	0.2	5	0.5	2	0.1	10	100
4	25	0.25	4	0.4	2.5	0.1	10	100
Total			25.67					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
7	40	0.4	2.5	0.25	4	0.1	10	100
8	45	0.45	2.2	0.2	4.5	0.1	10	100
9	80	0.8	1.25	0.125	8	0.1	10	100
10	100	1	1	0.1	10	0.1	10	100
Total			6.97					400

PPS sample of schools

- With a PPS sample of school:
 - Self weighted sample of students
 - Rebuild the student population
- But
 - Large variability of the school population estimate
- ⇒ Use of a systematic procedure

PSS Sampling example (sampling interval 100, start 75.2)

School label	School size	From student number	To student number	Part of the sample
1	10	1	10	No
2	15	11	25	No
3	20	26	45	No
4	25	46	70	No
5	30	71	100	Yes
6	35	101	135	No
7	40	136	175	No
8	45	176	220	Yes
9	80	221	300	Yes
10	100	301	400	Yes

PPS Sampling - summary

- More efficient way of sampling
 - Reduction of schools
 - Reduction of variation in weights
- Weights may still differ because of ...
 - Over- or under-sampling of some strata in the population
 - Lack of accuracy in measures used for selection of sample
 - Adjustments to weights for school and student non-response

Possible weight components

- School base weight
- School weight trimming factor (to avoid extremely large or small school weights)
- Student or Classroom base weight
- School non-response adjustment
- Student non-response adjustment
- Final student or classroom weight trimming factor (to avoid final student weights more than four times the median)

School-level weights

- The **School base weight** (for PPS samples) is the population size over the number of sampled schools multiplied with the school size:

$$BW_{sc}^i = \frac{N_{pop}}{n_{sc} * N_{non}}$$

- The **school non-response adjustment** is the number of sampled schools over the number of participating schools (in PISA: schools weighted with size):

$$Sc_{adj} = \frac{n_{sc}}{n_p}$$

Student-level weights

- The **Student base weight** is the population size over the number of sampled students multiplied with the school size:

$$BW_{st}^i = \frac{N_{st}^i}{n_{st}^i}$$

- The **student-level non-response adjustment** is the number of sampled (and eligible) students within a school over the number of participating students:

$$St_{adj}^i = \frac{n_{st}^i}{n_p^i}$$

Final sampling weight

- The final student weight (in simple designs) is equal to multiplying school and student base weights and the non-response adjustments:

$$W_i = BW_{sc}^i * Sc_{adj} * BW_{st}^i * St_{adj}^i$$

- Final student weights are often (but not always) the same within schools!
- Population size can be estimated by summing students weights within and across schools

Summary of Weighting

- One-level SRS sampling usually not appropriate for student surveys.
- PPS sampling more efficient than Two-level SRS.
- Weights do vary across students/schools.
- Generally, analysis of large-scale survey data should use sampling weights!!!

Questions?

Practice Session - weighting

- Generated data with 998 schools and 99,479 students in population
- Variables are
 - SCORE: Achievement score with mean of 500 and SD of 100
 - GENDER of students
 - SES: Socio-economic status index with mean of 0 and SD of 1
 - PROG: General vs. vocational study program (1, 0)

Sampling design

- PPS sample
- Explicit stratification: PROG
- Implicit stratification: High/low SES schools
- 50 schools per explicit stratum
- 30 students randomly selected per school
- Schools sorted by enrolment

Your task

- Please compute the sampling weights for this sample
 - WGT_EXMPL_DATA.SAV (96 schools)
- Student-level weights can be computed at school level
- Only participating schools in data!

Variables in datafile

- POPSIZE: Population size of each explicit stratum
- NSCHOOL: Number of schools in sample
- ENROL: Number of enrolled students at each school
- STUDSMP: Student sample size within schools
- SPAR: Number of participating students within schools

Weights (within explicit strata)

- School base weight:
 - POPSIZE / (50 * ENROL)
- School non-response adjustment
 - 50 / PSCHOOL
- Student base weight
 - ENROL / STUDSMP
- Student non-response adjustment
 - STUDSMP / SPAR
- Final student weights
 - Product of all four weights

```

* VIEW WEIGHTS with SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Tables Window Help

COMPUTE SCL_BW = FREQ100 / (50 * ENROL) .
COMPUTE SCL_AP = 50 / FREQ100 .
COMPUTE STD_BW = ENROL / STDENR .
COMPUTE STD_AP = STDENR / SPAR .
COMPUTE TOTAL_W = SCL_BW * SCL_AP * STD_BW * STD_AP .
EXECUTE .

WEIGHT
BY STD_W .
FREQUENCIES
VARIABLES=PROG
/ORDER= ANALYSIS .

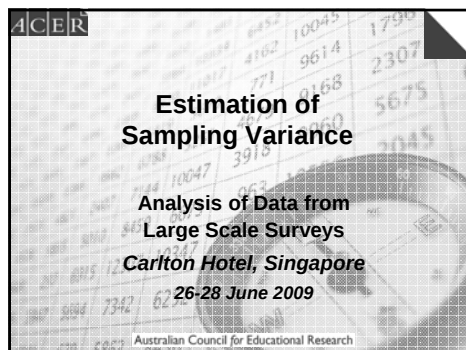
```

Check of results

- Compute SUMWGT
 - Product of Final weight and participating students (SPAR)
- Weight by SUMWGT
 - Frequency of PROG
 - Compare with table:

PROG

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	42022	42.2	42.2	42.2
	1.00	57457	57.8	57.8	100.0
Total		99479	100.0	100.0	



Topics of this presentation

- Error components in measurement
- Implications of cluster sampling for the estimation of sampling variance
- Methods to estimate sampling variance with cluster sample data

Error components associated with measurement

- Sampling error
 - Random (-> estimation of sampling error)
 - Systematic (e.g. non-response bias)
- Measurement error
 - Random (-> reliability, multiple imputation)
 - Systematic (e.g. unfavourable test conditions)

Simple Random Sample (SRS)

- A sample is random if all the individuals in the population have a known probability to be selected
- The sample will be simple if the individuals are all selected independently and every individual has the same chance of being selected

Sampling variance

- In the case of a SRS, the sampling variance is:
 - proportional to the variance
 - inversely proportional to the sample size
- The standard error can be obtained from the variance in the population and the sample size
- However, the variance in the population is (typically) not known

Calculation of Standard Errors

- As we don't know the variance in the population, the SE for a mean of θ as obtained from a sample is calculated as

$$SE_{\hat{\mu}} = \frac{\hat{\sigma}_{\hat{\mu}}}{\sqrt{n}}$$
- similarly, the SE for a percentage P is calculated as

$$SE_{\hat{p}} = \sqrt{\frac{PQ}{n}}$$

Drawing a sample of students

- As already mentioned, for practical, conceptual and financial reasons, SRS is not appropriate
- Cluster sampling provides a more adequate solution:
 - At the first stage schools are drawn from a sampling frame,
 - then classes and or students are selected randomly within school

Cluster sampling (1)

Cluster sampling (2)

- The variance in a cluster sample consists of *Between-school* and *Within-school* variance
- In an infinite population of clusters and an infinite population within clusters, and if the sample sizes m are equal within l sampled clusters, the variance is:

$$V(\bar{y}) = \frac{\sigma^2_{\text{between}}}{l} + \frac{\sigma^2_{\text{within}}}{l * m}$$

Cluster sampling (3)

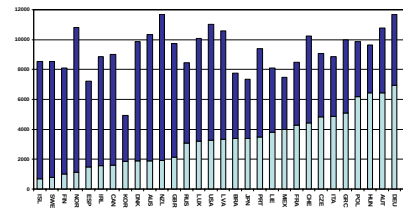
Comparison with a SRS:

- For example, 150 schools with 35 students per schools. Expected STD= 100

$$\sigma^2_{(u)} = \frac{4000}{150} + \frac{6000}{150 * 35} = 26.667 + 1.143 = 27.8$$

$$\sigma^2_{(u)} = \frac{0}{150} + \frac{10000}{5250} = 1.90$$

School and Within school variance (Reading Pisa 2000)



How to estimate sampling variance for complex designs?

- Design so complex that there are no formulae, even for a mean
 - Multilevel analysis (tends to overestimate the sampling variance for stratified samples)
 - Taylor series linearization (requires data on stratification, e.g. STATA)
 - Re-sampling methods
 - Bootstrap methods.
 - Replication methods (e.g. WESVAR, AM but can also be programmed in SPSS or SAS).

Types of replication methods

- Jackknifing
 - JK1 (only recommended with no information on sampling design)
 - JK2 (TIMSS 1995, 1999, 2003, PIRLS, CivEd 1999)
- Balanced Repeated Replication (BRR)
 - BRR
 - Fay's Method (applied in PISA)

Jackknife for unstratified sample

Replicate	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
School 1	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 2	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 3	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 4	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11
School 5	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11
School 6	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11
School 7	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11
School 8	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11
School 9	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11
School 10	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00

Jackknife for stratified sample

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1		2	1	1	1	1	1	1	1	1
1	2		0	1	1	1	1	1	1	1	1
2	3		1	0	1	1	1	1	1	1	1
2	4		1	2	1	1	1	1	1	1	1
3	5		1	1	2	1	1	1	1	1	1
3	6		1	1	0	1	1	1	1	1	1
4	7		1	1	1	0	1	1	1	1	1
4	8		1	1	1	2	1	1	1	1	1
5	9		1	1	1	1	2	1	1	1	1
5	10		1	1	1	1	0	1	1	1	1
6	11		1	1	1	1	1	2	1	1	1
6	12		1	1	1	1	1	0	1	1	1
7	13		1	1	1	1	1	1	0	1	1
7	14		1	1	1	1	1	1	2	1	1
8	15		1	1	1	1	1	1	1	0	1
8	16		1	1	1	1	1	1	1	2	1
9	17		1	1	1	1	1	1	1	1	0
9	18		1	1	1	1	1	1	1	2	1
10	19		1	1	1	1	1	1	1	1	2
10	20		1	1	1	1	1	1	1	1	0

Balanced Repeated Replication (BRR)

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1	2	0	0	2	0	0	0	2	2	0
1	2	0	2	0	2	2	0	0	0	0	2
2	3	2	2	0	0	2	0	0	0	2	2
2	4	0	0	2	2	0	2	2	2	0	0
3	5	2	0	2	0	0	2	0	0	2	2
3	6	0	2	0	2	2	0	2	2	0	0
4	7	2	2	0	2	0	0	2	0	0	2
4	8	0	0	2	0	2	2	0	2	2	0
5	9	2	2	0	0	2	0	0	2	0	2
5	10	0	0	0	2	0	2	2	0	2	2
6	11	2	2	2	2	0	2	0	0	2	0
6	12	0	0	0	0	2	2	0	2	2	2
7	13	2	0	2	2	2	0	2	0	0	2
7	14	0	2	0	0	2	0	2	2	0	2
8	15	2	0	0	2	2	0	2	0	0	2
8	16	0	2	2	0	0	2	0	2	2	0
9	17	2	0	0	0	2	2	0	2	0	2
9	18	0	2	2	2	0	0	0	2	0	2
10	19	2	2	0	0	0	2	2	2	0	0
10	20	0	0	2	2	2	0	0	2	0	2

Fay's variant of BRR

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5	0.5	1.5
1	2	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5	1.5	0.5
2	3	1.5	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5	0.5
2	4	0.5	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5	1.5
3	5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5
3	6	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5
4	7	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5	1.5	1.5
4	8	0.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5	0.5	0.5
5	9	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5	1.5
5	10	0.5	0.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5	0.5
6	11	1.5	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5
6	12	0.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	1.5
7	13	1.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5
7	14	0.5	1.5	0.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5
8	15	1.5	0.5	0.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5
8	16	0.5	1.5	1.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5
9	17	1.5	0.5	0.5	0.5	1.5	1.5	0.5	1.5	0.5	1.5
9	18	0.5	1.5	1.5	1.5	0.5	0.5	1.5	1.5	1.5	0.5
10	19	1.5	1.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5
10	20	0.5	0.5	1.5	1.5	1.5	0.5	0.5	1.5	1.5	1.5

Replication Theory

- Variance estimators generally take the form of

$$Var_{rep}(\theta) = c \sum_{k=1}^G [\theta_k - \bar{\theta}]^2$$
- θ is the parameter of interest
- G is the total number of replicates with k denoting the k -th replicate
- c is a constant that depends on the method
 - For JK1, $c = (G-1)/G$
 - For JK2, $c = 1$
 - for BRR, $c = 1/G$
 - for Fay's BRR, $c = 1/(G(1-k^2))$

Computing sampling variance in PISA

- In Pisa
 - Fay's method
 - 80 replicate weights in data file
 - Perturbation factor k is 0.5
 - $c = 1/(80(1-0.5)^2) = 0.05$

$$\sigma_{(\hat{\theta})}^2 = \frac{1}{G(1-k)^2} \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{80(1-0.5)^2} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{20} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2$$

- To compute sampling errors, one needs to estimate once with the full PISA weight and 80 times with each replicate weight!

Computing sampling variance in CIVED, TIMSS or PIRLS

- In IEA studies
 - Jackknife (JK2) method
 - 75 replicate weights in data file
 - $c = 1$
- To compute sampling errors, one needs to estimate once with the full PISA weight and 75 times with each replicate weight!

How can we achieve this?

- Special software
 - WESVAR (Westat)
 - AM (AIR)
- Macros (requires syntax handling)
 - SPSS
 - SAS
- IEA Database Analyser
- SPSS Analysis module
 - Based on SPSS macros but menu-driven

Questions?

Example for practice - 1

- Compute Standard errors based on example with Jackknife (JK2) replication (generated data)
- Please open SPSS file
 - REPLICATES_EXAMPLE.sav
- Variables in File
 - Mean of achievement scores for each replicate weights (SCORE)
 - Mean of achievement scores with full sample weight (SCORE_FW)

Example for practice - 2

- Compute squared difference between scores with replicate weights and full sample weight
- Compute sum and multiply with correct constant (for JK2 = 1)

ACER

Scaling methodology

Analysis of Data from
Large Scale Surveys
Carlton Hotel, Singapore
26-28 June 2009

Australian Council for Educational Research

Rationale for IRT scaling of data

- Summarising data instead of dealing with many single items
- Raw scores or percent correct sample-dependent
- Makes equating possible and can deal with rotated test forms

The 'Rasch model'

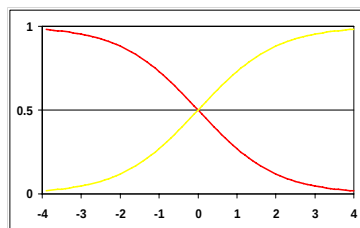
- Models the probability to respond correctly to an item as

$$P_i(X_{ni} = 1) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

- Likewise, the probability of NOT responding correctly is modelled as

$$P(X_{ni} = 0) = \frac{1}{1 + \exp(\theta_n - \delta_i)}$$

IRT curves



Some examples

Student ability	Item Difficulty	Probability of success
-2	-2	0.50
-1	-1	0.50
0	0	0.50
1	1	0.50
2	2	0.50

Some examples

Student ability	Item Difficulty	Probability of success
-2	-3	0.73
-1	-2	0.73
0	-1	0.73
1	0	0.73
2	1	0.73

Student ability	Item Difficulty	Probability of success
-2	-1	0.27
-1	0	0.27
0	1	0.27
1	2	0.27
2	3	0.27

Some examples

Student ability	Item Difficulty	Probability of success
-2	-4	0.88
-1	-3	0.88
0	-2	0.88
1	-1	0.88
2	0	0.88

Student ability	Item Difficulty	Probability of success
-2	0	0.12
-1	1	0.12
0	2	0.12
1	3	0.12
2	4	0.12

Anchoring of tests

- The Rasch model allows to derive comparable scores for two different assessment via the inclusion of "anchor" items.
- Anchoring can be used across time (e.g. repeated surveys) but also across different population assessments (e.g. grades)

Example of anchoring between grades 4 and 8

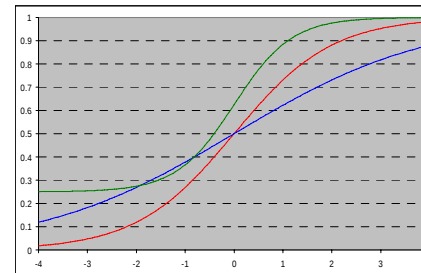


3 Parameter model

- Three-parameter models have additional parameters for item discrimination and guessing:

$$P(X_{ij} = 1 | \theta_i, a_i, b_i, c_i) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta_i - b_i))}$$

Comparisons of the ICC curves

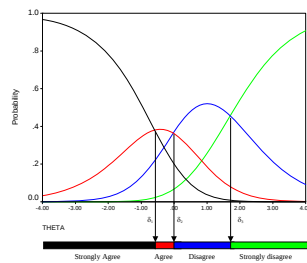


IRT models for polytomous items

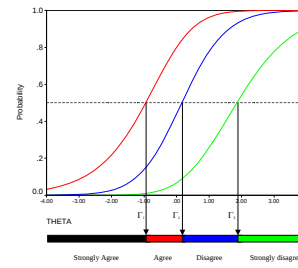
- Extension of Rasch model with additional step parameters τ
- Rating scale model* has equal step parameters across items
- Partial credit model* has different step parameters for each item

$$P_{x_i}(\theta) = \frac{\exp \sum_{j=0}^{x_i} (\theta_n - (\delta_i + \tau_{ij}))}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^h (\theta_n - (\delta_i + \tau_{ij}))} \quad x_i = 0, 1, 2, \dots, m_i$$

Partial Credit Model – probabilities



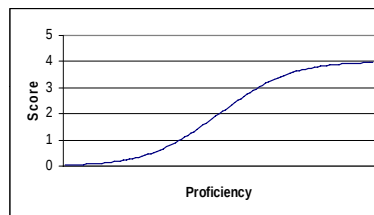
Partial Credit Model – threshold curves



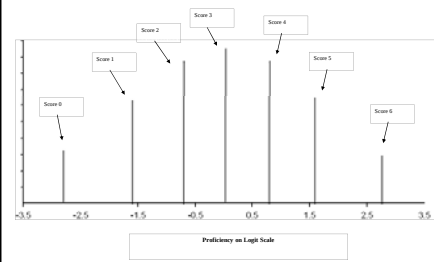
How might we impute a reasonable proficiency value?

- Choose the *proficiency* that makes the score most likely
 - Maximum Likelihood
 - Weighted Likelihood
- Choose the most likely *proficiency* for the score
 - empirical Bayes
- Choose a selection of likely proficiencies for the score
 - Multiple imputations (plausible values)

Maximum Likelihood vs. Raw Score



The Resulting Proficiency Distribution



Characteristics of MLE

- Unbiased at individual level with sufficient information BUT biased towards ends of ability scale.
- Arbitrary treatment of perfects and zeroes required
- Discrete scale & measurement error leads to bias in population parameter estimates

Weighted Likelihood Estimation

- WLE maximizes the weighted likelihood, wL , where w is the weight function.
- Solve for each case

$$\sum_{i=1}^k \left[r_i + \frac{J}{2I} \right] - \sum_{j=1}^k \frac{\exp(\theta_a - \delta_j)}{1 + \exp(\theta_a - \delta_j)} = 0$$

- $\frac{-J}{2I^2}$ being the bias for the ML estimates
- with I being the Information function: $\sum_{i=1}^k \frac{p_{vi}'^2}{p_{vi}(1-p_{vi})}$
- and $J = \sum_{i=1}^k \frac{p_{vi}' p_{vi}''}{p_{vi}(1-p_{vi})}$

Characteristics of WLE

- Less biased than MLE
- Provides estimates for perfect and zero scores
- BUT discrete scale & measurement error leads to bias in population parameter estimates**

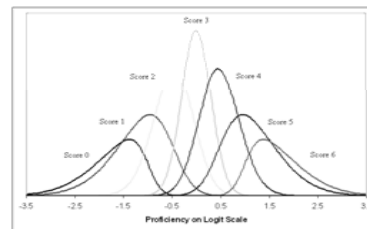
Plausible Values

- What are plausible values?
- Why do we use them?
- How to analyse plausible values?

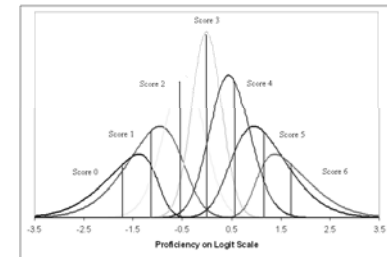
Purpose of educational tests

- Measure particular students**
(minimise measurement error of individual estimates)
- Assess populations**
(minimise error when generalising to the population)

Posterior distributions for test scores on 6 dichotomous items



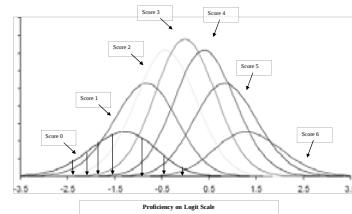
Empirical Bayes - EAPs



Characteristics of EAPs

- Biased at the individual level but unbiased population means (NOT variances)
- Discrete scale, bias & measurement error leads to bias in population parameter estimates
- Requires assumptions about the distribution of proficiency in the population

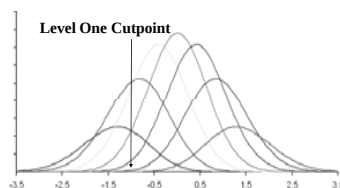
Plausible Values



Characteristics of Plausible Values

- Not fair at the student level
- Produces unbiased population parameter estimates
– if assumptions of scaling are reasonable
- Requires assumptions about the distribution of proficiency

Estimating percentages below benchmark with Plausible Values



The proportion of plausible values less than the cut-point will be a superior estimator to the EAP, MLE or WLE based values

Methodology of PVs

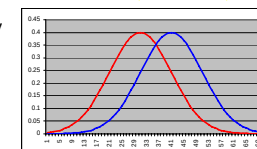
- Mathematically computing posterior distributions around test scores
- Drawing 5 random values for each assessed individual from the posterior distribution for that individual

What is conditioning?

- Assuming normal posterior distribution: $N(\mu, \sigma^2)$

- Model sub-populations: $N(\mu + \alpha X, \sigma^2)$

X=0 for boy
X=1 for girl



$$N(\mu + \alpha X + \beta Y + \gamma Z + \dots, \sigma^2)$$

Multi-dimensional scaling

- 4 scales for major domain => 4 abilities for each person, $\theta_1, \theta_2, \theta_3$ and θ_4
- Covariance between the abilities can be estimated.
- It is assumed that only one ability is drawn upon for answering a particular item.

Conditioning Variables

- Plausible values should only be analysed with data that were included in the conditioning (otherwise, results may be biased).
- Aim: Maximise information included in the conditioning, that is use as many variables as possible.
- To reduce number of conditioning variables, factor scores from principal component analysis can be used.
- Use school means of performance takes between-school variation into account (no inclusion of school questionnaire data needed)

Plausible values

- Model with conditioning variables will improve precision of prediction of ability (population estimates ONLY)
- Multidimensional models improves estimates, through using information about correlation between dimensions.
- Conditioning provides unbiased estimates for modelled parameters.
- For tests with more than one dimension multi-dimensional scaling is recommended.

Results from analysis with simulated data

- Simulation used to test bias in WLE, EAP and PV estimates
- Generating item responses
 - True abilities are known
 - True relationships between abilities and background variables are known
 - Within- and between school variances are known
- Simulated item responses were used to estimate WLEs, EAPs and PVs with ConQuest

Outcomes of simulation studies - 1

- MEAN:
WLE, EAP and PV unbiased
- SD and PERCENTILES:
WLE overestimated
EAP underestimated
PV unbiased

Outcomes of simulation studies - 2

- CORRELATION COEFFICIENT:
WLE underestimated
EAP-C overestimated
PV-C unbiased
- REGRESSION COEFFICIENT:
WLE, EAP-C, PV-C unbiased
- But RESIDUAL VARIANCE:
WLE overestimated
EAP-C underestimated
PV-C unbiased

Calculating of measurement error

- In PISA, TIMSS or PIRLS data files, there are five plausible values for each cognitive domain
- Using five plausible values enable researchers to calculate estimates of the measurement error

How to analyse PVs - 1

- Estimated mean is the AVERAGE of the mean for each PV

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M \hat{\mu}_i$$

- Sampling variance is the AVERAGE of the sampling variance for each PV

$$\hat{\sigma}_{(\hat{\mu})}^2 = \frac{1}{M} \sum_{i=1}^M \hat{\sigma}_{(\hat{\mu}_i)}^2$$

How to analyse PVs - 2

- Measurement variance computed as:

$$\hat{\sigma}_{(PV)}^2 = \frac{1}{M-1} \sum_{i=1}^M (\hat{\mu}_i - \hat{\mu})^2$$

- Total standard error computed from measurement and sampling variance as:

$$\hat{\sigma}_{(\hat{\mu}_{PV})} = \sqrt{\sigma_{(\mu)}^2 + (1 + \frac{1}{M})\sigma_{(PV)}^2}$$

How to analyse PVs - 3

- μ can be replaced by any statistic for instance:
 - SD
 - Percentile
 - Correlation coefficient
 - Regression coefficient
 - R-square
 - etc.

Steps for estimating both sampling and measurement error

- Compute statistic for each PV for fully weighted sample
- Compute statistics for each PV for 80 replicate samples
- Compute sampling error (based on previous steps)
- Compute measurement error
- Combine error variances to calculate standard error

Questions?

Practice 1

- Compute combined standard error for Science achievement in Australia (PISA 2006)
 - [Example_PVs.xls](#)

Steps

- Compute overall results as average of five plausible values
- Compute overall sampling variance from average squared SEs for each PV
- Compute measurement variance
 - Sum of squared deviations from average times $1/(M-1) = 0.25$
- Compute combined standard error
 - Square root of overall sampling variance + measurement variance times 1.2

Practice 2

- Reproducing parts of PISA 2003 report, table 2.5c, page 351 (Maths performance)
 - Means (PV)
 - Standard deviation (PV)
 - Mean by male/female (PV)
 - Difference male/female (PV)
 - Percentiles (PV)

Practice 3

- Reproducing TIMSS 2003 report, page 16, exhibit 1.1 (Maths performance)
 - Means (PV)
 - Mean by male/female (PV)
 - Difference by male/female (PV)

Practice 4

- Reproducing PISA 2003 table 2.5a, page 349 (Maths proficiency levels)
 - percentages for each proficiency level (PV)

Practice 5

- Reproducing PISA 2003 report, table 4.2e, page 387
 - Recode FAMSTRUC into single-parent vs. others
 - Group percentages
 - Maths performance by groups
 - Difference by groups

PISA reporting of achievement by quartiles of questionnaire indices

- In PISA reports achievement means by quartiles of questionnaire indices
- To ensure equal distribution random error is added to questionnaire index scores
- No exact replication possible
- SPSS module provides tool for producing results

Practice 6

- Reproducing parts of PISA 2003 international report, table 3.7, page 368 (variable MATHEFF, MATHS)
 - Means
 - Performance by quartile

Effect Sizes

- Instead of comparing absolute differences researcher may want to report standardised differences:

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

- SPSS replicates enables researchers to estimate standard errors for single variables and plausible values

Practice 7

- Use one PISA 2006 country data set (e.g. Hong Kong data) to compute effect sizes plus standard errors for
 - Gender (make 0, 1 variable)
 - Reading

Relative Risk - 1

- Measure of association between an antecedent factor and an outcome factor based on 4X4 table
- Calculation requires dummy variables (coded 0 and 1).
- For example, the outcome variable might be reaching a performance benchmark (yes=1) and the antecedent a student's gender (female=1)

Relative Risk Ratio - 2

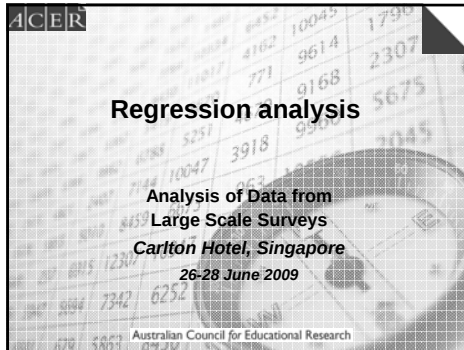
		Outcome measure		Total
		Yes	No	
Antecedent measure	Yes	p_{11}	p_{12}	$p_{1.}$
	No	p_{21}	p_{22}	$p_{2.}$
Total		$p_{.1}$	$p_{.2}$	p

$$RR = \frac{(p_{11} / p_{1.})}{(p_{21} / p_{2.})}$$

- The resulting value how much more likely a respondent would fall into the category given the antecedent

Practice 8

- Use one PISA 2006 country data set (e.g. Hong Kong data) and create dummy variables for each plausible value (Reading) for the proficiency level 3 and above ($>$ value 480.2)
- Create five gender dummies (G1 to G5) with male as 1 and female as 0
- Save dataset and use SPSS Replicates Module



Regression is...

- Explaining the distribution of a dependent variable Y with...
- the distribution of one or more independent variables $X_1, X_2, \dots, X_n$

Types of regression models

- Linear regression
- Non-linear regression
- Logistical regression
- Time Series Regression
- Multi-level Regression (linear and logistical)

Regression can be...

- Bivariate with just one predictor
- Multiple with various predictors
- Path analysis, when there is more than one dependent variables
- Non-recursive, when two variables are both dependent and independent with respect to each other...

The linear OLS (Ordinary Least Squares) regression model

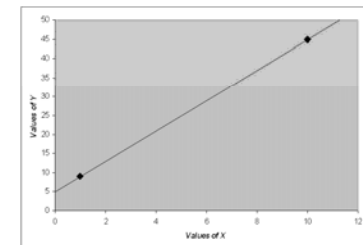
- Regression equation

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- Error term

$$\varepsilon_i \sim N(0, \sigma^2)$$

Graphical display of regression equation



Regression equation

$$\hat{Y} = 5 + 4X$$

- This equation can be graphically represented by a straight line
- A equation with 2 variables (X, Y) can be represented in a two-dimensional space

Linear regression is based on:

- Means: $\bar{X} = \frac{\sum X}{N}$
- Variances: $\text{var}(x) = \frac{\sum X - \bar{X}}{N - 1}$
- Covariances: $\text{cov}(x, y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1}$

Regression Parameters

- Coefficient: $b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$
- Intercept: $a = \bar{Y} - b\bar{X}$
- Predicted Y: $\hat{Y} = a + bX$
- Residual: $e = \hat{Y} - Y$

Partitioning of Variance

Y consists of $Y = \bar{Y} + (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$

Var(Y) is $\sum y^2 = ss_{reg} + ss_{res}$

Explained variance: $r_{xy}^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}$

Unexplained variance: $ss_{res} = (1 - r_{xy}^2) s_y^2$

Tests of significance (simple random samples)

- F-Test for regression $F = \frac{ss_{reg}/k}{ss_{res}/N - k - 1}$
- t-test for parameters $t = \frac{b}{s_b}$
- SE for parameter $s_b = \sqrt{\frac{ss_{res}/N - k - 1}{\sum x^2}}$

Standardised Coefficient

- Regression coefficients can be standardised as: $\beta^* = b \frac{s_x}{s_y}$
- In bivariate regression this is equal to $r_{xy} = \frac{\text{cov}(X, Y)}{s_x s_y}$
- Or a regression between standardised variables $z_y = \beta^* z_x$

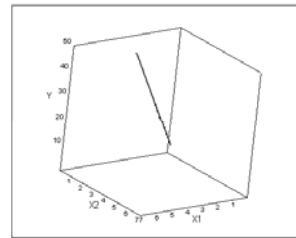
Multiple Regression

- Equation becomes $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + e_i$
- Partial coefficients with standardised scores can be simply calculated as

$$\beta_1^* = \frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2}$$

- and $\beta_2^* = \frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2}$

Equation with two predictors: $Y = 5 + 2X_1 + 3X_2$



How to judge whether a model is good?

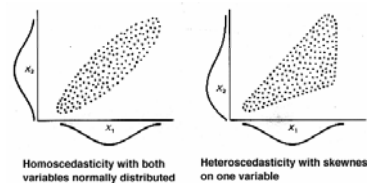
- Explained variance proportion as measures by R Square.
- Size of regression coefficients.
- Significance tests (F-test for model, t-tests for parameters).
- Inclusion of all relevant variables (Theory!).
- Is method appropriate?

Assumptions of linear regression

- Linearity of relationship
- Continuous nature of (at least) the dependent variable
- Normality of distributions
- Homoscedasticity
- Absence of multicollinearity (independent variables not being highly correlated with each other)

Heteroscedasticity

- Heteroscedasticity is caused by non-normality of one of the variables, an indirect relationship between variables, or to the effect of a data transformation.



Analysis of residuals

- Usually done with standardised residuals $z = \frac{Y - \hat{Y}}{s_{y.x}}$
- Detect outliers (extreme values can distort estimates)
- Check random and even distribution

Multicollinearity - 1

- Means that one or more independent variables are linear combination of other independent variables
- This can distort the estimation of coefficients
- Can lead to large standard errors for the coefficients

Multicollinearity - 2

- No 'scientific' rules
- Check correlation between independent variables
- Use multicollinearity statistics (VIF, Tolerance)
- Deletion of variable may lead to misspecification of models
- Sometimes variables can be combined (for examples Mother's and Father's education)

Use of Dummy Variables

- Dichotomous or Categorical variables can be recoded into dummies with (0,1).
- In a bivariate case this leads to the intercept=mean(0), slope=mean(1).
- Sometimes interaction effects can be modelled (different slopes for different categories).

Use of standardised coefficients

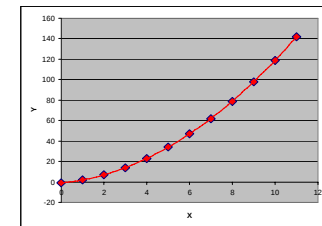
- Often thought to be 'easier' to interpret.
- Standardisation depends on variances of independent variables.
- Unstandardised coefficient are generally preferable (can be translated directly).
- Standardised coefficients cannot always be compared across populations.

Curvilinear relationships

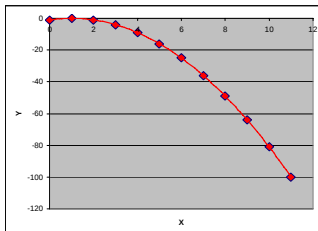
- Often relationships are not linear, but curvilinear (e.g. homework & performance)
- This can be modelled by adding a quadratic term to the equation:

$$Y_i = \alpha + \beta_1 X + \beta_2 X^2 + e_i$$

$$Y = -1 + 2X + 1X^2$$



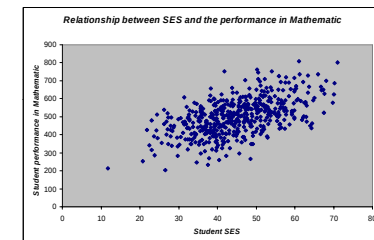
$$Y = -1 + 2X - 1X^2$$



When does linear regression modelling become inappropriate?

- When the dependent variable is dichotomous or polytomous (>> Logistical Regression).
- When data are sequential over time and variables are 'auto correlated' (>> Time Series Analysis).
- When context effects need to be analysed and slopes are different across higher level units (>> Multi-level Analysis).

Example: SES and performance

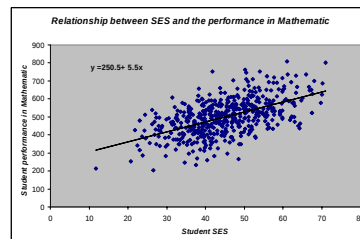


SES and performance

$$Y_i = \alpha + \beta X_i$$

- $Y_i = 250.5 + 5.5(\text{SES})$

SES and Maths performance



Observed, predicted and residual scores

Students	SES	Observed score	Predicted Score	Residual
1	49	463	520	-57
2	53	384	542	-158
3	51	579	531	+48
4	42	404	481.5	-77.5
5	42	282	481.5	-199.5

Questions?

Practice 1

- Estimate regression model for GENDATA_sample.sav
 - Predictors: Gender, SES, PROG
- First using normal syntax
- Then with SPSS module to compute correct standard errors

Practice 2

- Reproducing PISA 2003 Table 4.3a, page 392
 - Linear regression model (PV) Maths on ESCS
 - Quadratic regression model (PV) Maths on ESCS
 - Mean, standard deviation and skew of ESCS
 - Percentage of students with ESCS < -1.

Practice 3

- Reproducing Figure 4.9, page 176
 - Graph of gradient based on quadratic regression
- Use EXCEL spreadsheet [SLOPE.XLS](#)

附錄八

出國人員出國報告書(第四年期部分)：

1. 參加 2008 NARST Annual International Conference 及參訪 NSF(National Science Foundation)
2. 參加 IERI 舉辦之 Assessment Designs, Item Response Theory and Proficiency Estimates in Large Scale Assessment
3. 參加 ACER 舉辦之 Data Analysis of Large-Scale Survey Data Workshop

出席國際學術會議心得報告

計畫編號	NSC 97-MOE-S-003-003
計畫名稱	國際數學與科學教育成就趨勢調查2007 (4/4) 甲
出國人員姓名 服務機關及職稱	任宗浩 國立台灣師範大學科學教育中心 助理研究員 李哲迪 國立台灣師範大學科學教育中心 助理研究員
會議時間地點	2009 年 2 月 13 日- 2 月 16 日 Princeton, New Jersey, USA
會議名稱	IERI Autumn Academy (Session II): Test Designs, Item Response Theory and Proficiency Estimates

一、與會歷程

此工作坊為 IERI (IEA-ETS Research Institute) 所主辦。IERI 是兩個研究機構合作組成的組織：其中之一是主持 TIMSS 等國際調查研究的機構—IEA (International Association for the Evaluation of Educational Achievement)，另一個機構是負責 TOEFL、GRE、SAT 等許多美國大型測驗的中心—ETS (Educational Testing Service)。IERI 為推廣二次分析研究，鼓勵研究者使用學習成就調查資料庫，定期舉辦各種與研究設計和分析方法有關的工作坊。此次工作坊的主題是測驗設計、試題反應理論、以及能力估計。

工作坊分四天進行（議程如下表）。第一天介紹測驗之設計，主講人是 Eugenio J. Gonzalez 博士。其介紹的重點特別放在題本的設計上。第二天介紹試題反應理論在大型調查研究中的應用。主講人是 Kentaro Yamamoto 博士。第三天和第四天說明母群分布模型在能力估計中的應用。主講人是 Matthias von Davier 博士。在這個部份，講者介紹了一套由 ETS 所發展的軟體—DESI。此軟體可用以估計學生的能力，產生似真估計值 (plausible value)。工作坊進行的方式是先以理論介紹作為前導，隨後緊接著就是習作練習，前一節所學到的概念在後一節問題解決的過程中得以消化吸收。

Tuesday January 13

8:30 – 9:00	Registration
-------------	--------------

9:00 – 10:30	Welcome and Introductions Overview of Seminar and Agenda Review of Matrix Designs in Large Scale Assessment (Gonzalez)
10:30 – 10:45	Break
10:45 – 12:15	Review of Matrix Designs in Large Scale Assessment (Gonzalez)
12:15 – 1:45	Lunch
13:45 – 15:15	Matrix Designs in Large Scale Assessment (Work in Groups)
15:15 – 15:30	Break
15:30 – 17:00	Matrix Designs in Large Scale Assessment (Presentations)

Wednesday January 14

9:00 – 10:30	Review of Item Response Theory (IRT) Uses of IRT in Large Scale Assessment (Yamamoto)
10:30 – 10:45	Break
10:45 – 12:15	Item Response Theory Models (Yamamoto)
12:15 – 1:45	Lunch
13:45 – 15:15	Obtaining Item Difficulties and Person Scores (Yamamoto)
15:15 – 15:30	Break
15:30 – 17:00	Obtaining Item Difficulties and Person Scores (Yamamoto)
19:00 – 21:00	Group Dinner

Thursday January 15

9:00 – 10:30	Population Models in Large Scale Assessments (Von Davier)
10:30 – 10:45	Break
10:45 – 12:15	Population Models in Large Scale Assessments (Von Davier)
12:15 – 1:45	Lunch

13:45 – 15:15	Using DESI to obtain Person Scores (Von Davier)
15:15 – 15:30	Break
15:30 – 17:00	Using DESI to obtain Person Scores (Group Work)

Friday January 16

9:00 – 10:30	Putting it all together (Gonzalez, Von Davier, Yamamoto)
10:30 – 10:45	Break
10:45 – 12:15	Putting it all together (Gonzalez, Von Davier, Yamamoto)
12:15	Adjourn

二、 與會心得

整體而言，此次工作坊所介紹的內容頗為實用。以下以題本設計和能力估計為例說明之。

題本設計是大型調查研究必備的研究設計技術。爲了能準確地估計各國學生的能力，所需要的測驗試題不僅在題目數量上要夠多（以 TIMSS 為例，自然科學約爲 300 題左右），在涵蓋範圍上也要夠廣。但是學生的作答時間一般限制在兩個小時左右，因此每位學生能完成的題目數有限。爲了顧及測量的準確性和學生作答能力的限制，目前的作法是將試題分爲數種題本，題本間部分試題重複，每位學生僅作其中一種題本。由於試題數量減少，學生即可在兩小時內完成，又由於學生所做的試題有共同的部分，因此以整體學生而言，所做的試題數量就是足夠的。題本之設計除了要考量上述測驗設計的因素之外，還要考慮實務上遭遇的其他問題。例如：試題公布的問題。無論是 TIMSS、美國各州的成就測驗、或是我國的基本學力測驗，在當屆施測後，都需要公布部分試題。這些公布的試題在下一屆的題本中自然不能使用，而需以其他題目取代。這些公布試題應等量地分布在各種題本中。在本次工作坊中，講者對這類在實務上應注意的重點所做的提點是在其他地方難以學習到的。

爲了估計學生能力並表達測量誤差，TIMSS 和 PISA 等大型調查研究對學生能力的

估計均以似真值來表示。一個學生的能力由一組五個數值的似真值來表示。在產生似真值的標準程序上，除了考慮學生的答題反應之外，還考慮了母群的分布模型。所謂母群的分布模型是指在母群中各種特質的學生的分布情況。在估計學生能力時，利用學生答題反應之外的其他特質，運用條件機率的概念，即可更準確地推估學生的能力，減少測量誤差。在此次工作坊中，運用 DESI，講者展示了應用母群分布模型估計學生能力的具體方法。

不過，此次工作坊對於 IRT 在大型調查研究中的應用和似真值的估計說明得不夠深入。由於工作坊的時間有限，加上參加者程度不一，講師必須對內容有所取捨，這是可以理解的。

我國目前除參加 TIMSS、PIRLS、PISA、ICCS 等國際調查研究外，還發展了 TEPS、TASA 等大型調查研究。在規劃與執行大型調查研究時，對抽樣設計、試題發展、量尺建立、能力估計、二次分析統計方法等等課題，研究者不僅需要有所認識，還需要對這些主題有整體的瞭解；因為這些主題環環相扣，彼此有關，若無整體的瞭解，計畫難保順利執行。在研究設計和統計方法等相關技術的運用上，目前各研究團隊都還在起步階段。在這個階段，借重國外研究者的知識與經驗是提升我國研究能力的重要策略。類似的工作坊若能邀請 IEA 與 ETS 的研究者來台舉辦，必能使國內更多學者受惠。

三、 攜回資料名稱及內容

1. 各場次講者提供之簡報檔及習作資料。
2. DESI 程式軟體一套。

The IEA/ETS Research Institute



The IEA/ETS Research Institute (www.IERInstitute.org)

What is it?

- Collaborative agreement between IEA and ETS
- Focus on improving the science of large scale assessment
- Undertakes activities around three broad areas:
 - ✓ Research studies related to the development and implementation of large-scale assessments
 - ✓ Professional development and training
 - ✓ Dissemination of research findings and information gathered through large-scale assessments



The IEA/ETS Research Institute (www.IERInstitute.org)

1

Research Activities

- A "Virtual" Institute
- Research Agenda
- Make available consulting and advisory resources to researchers and national centers wishing to conduct policy or methodology related studies
 - ✓ Identify relevant policy issues
 - ✓ Identify analytical strategies most appropriate for providing and analyzing data to address these issues
 - ✓ Apply statistical techniques and other research methodologies



The IEA/ETS Research Institute (www.IERInstitute.org)

2

Research Activities

- Research Priorities:
 - ✓ Develop a more scientific approach to the development, use and interpretability of background questionnaires
 - ✓ Develop new constructs that extend the policy issues that might be addressed by these assessments
 - ✓ Improve the measurement of cognitive domains
 - ✓ Investigate effects of increased emphasis on the role of technology
 - ✓ Identify thematic issues for guiding secondary analyses



The IEA/ETS Research Institute (www.IERInstitute.org)

3

Training Activities

- Workshops and seminars
 - ✓ Thematic
 - ✓ Making use of specific data
- Internship and fellowship programs
- Professional exchanges



The IEA/ETS Research Institute (www.IERInstitute.org)

4

Dissemination Activities

- Sponsor a Monograph Series
 - ✓ Published annually
 - ✓ 5-7 papers
 - ✓ Peer reviewed
 - Receive favorable technical, substantive, and editorial review
- Sponsor other publications that contribute to the advancement of the field



The IEA/ETS Research Institute (www.IERInstitute.org)

5

Matrix Designs in Large Scale Assessment

Eugenio J. Gonzalez
November 2008



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

So many items, so little time...



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

The Assessment Design

- How do we deliver the items to the population of interest?



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

2

Some Questions to Ask

- What is the purpose of the assessment?
- How broad is the content area or set of skills covered?
- What scores will be reported?
- How much time is available to administer the item?
- How will the data be reported?
- What decisions will be made based on results?
- Number of administrations
 - ✓ Same year and over the years



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

3

Purpose of the Assessment

- Selection
 - ✓ Have required set of skills
- Grading
 - ✓ Have mastered contents of course
- Diagnosis
 - ✓ Decide on individual treatment
- Describe population
 - ✓ Mastery of set of skills or knowledge



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

4

Broadness of the Assessment

- Contents covered at grade, or up to that grade?
- How many content areas?
- What skills are measured?
- How long does it take to measure those skills?
- How many items are necessary



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

5

Scales to be Reported

- How many scales?
- How many subscales?
- Who gets scores on what?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

6

Time Available for Administration

- How long does it take to perform individual tasks?
- How long will students last before fatigue sets in?
- School schedule structure
- What will teachers and school principals tolerate?
- Resources available?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

7

Reporting the Results

- Overall scale scores
- Subscale scores
- Percent correct or scale scores?
- Percents passing benchmarks
- Item by item
- Passing / no passing
- Other?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

8

Decisions Based on Results

- Describe population?
- Describe individuals? School performance?
- Individual consequences?
- Group consequences?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

9

Number of Administrations

- How many administration will take place?
- An on-going program? Or once in a lifetime?
- Is there an item release policy?
- Risks of disclosure/exposure of the items?
- Refreshing the assessment?
- (When) Will new component(s) be introduced?



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

10

Assessment Designs in LSA

- Depend on answers to previous questions
- Except for very simple cases there is no unique solution
- The design is always a compromise between what is desired and what is possible given specific circumstances and resources
 - ✓ Need to weigh pros and cons of each feature vs. choosing a different one
 - ✓ Make an overall judgment of the adequacy of the design(s)



Copyright © IEA-ETS Research Institute (www.IERInstitute.org)

11

Some of the Challenges...

- Domain is usually very broad
- Limited testing time (physical & psychological)
- Challenges:
 - ✓ Need to administer the items in a "sensible" design
 - ✓ Need to summarize performance on the items
 - ✓ Need to adjust for unreliability of estimates



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

12

How do Matrix Designs Help?

- Impossible to test everyone on everything
 - ✓ Too many people
 - ✓ Too many items
 - ✓ Too expensive
- Some students are tested on some things
- In large scale assessments we tend to sample items and students
- Calculations need to be adjusted using sampling design information



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

13

Arranging the Items

- Items are the individual questions or tasks that make up the assessment
- Blocks are subsets of items in the assessment
 - ✓ A subset could be one, some, or all the items in the assessment
 - ✓ Blocks are generally exhaustive and mutually exclusive
 - Every item belongs to a block
 - Each item belongs to only one block



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

14

Some Concepts

- The "assessment" constitutes the entirety of the items
- The "blocks" are subsets of items (also called clusters)
- The "form" is a particular set of blocks delivered to the students in a particular order
 - ✓ In paper and pencil applications we use booklet or form
 - ✓ In computer delivered assessments there is no booklet...



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

15

Some Choices of Designs

- Complete Matrix Designs
 - ✓ Everyone takes everything
 - ✓ Could have blocks in different order
 - ✓ One form or multiple forms
- Incomplete Matrix Design
 - ✓ Everyone takes part of the assessment
 - Dynamic or Static
 - Balanced or unbalanced



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

16

Complete Matrix Design

- All students take all the items in the same order
 - ✓ IEA Reading Literacy 1999 and 2001
 - One form, 2 blocks (A and B)
 - ✓ Advantages? Disadvantages?
- All students take all the items but in varying forms
 - ✓ Random or semi random: an unknown number of forms
 - ✓ Static: there is a fixed set of forms
 - ✓ Advantages? Disadvantages?
- Some Examples



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

17

Incomplete Matrix Designs

- ✓ Each student takes only part of the assessment
 - Dynamic: based on performance
 - Static: pre-printed or predetermined sets of items
- ✓ Advantages? Disadvantages?
- ✓ Some examples



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

18

Some Studies

- Two major organizations:
 - ✓ International Association for the Evaluation of Educational Achievement (IEA)
 - Conducts TIMSS & PIRLS
 - Major activity is comparative educational studies, research and training
 - ✓ Organization for Economic Cooperation and Development (OECD)
 - Conducts PISA
 - Data collection in education and other fields
 - Collects indicators to inform governments



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

19

The Studies

- TIMSS
 - ✓ Trends in International Mathematics and Science Study
- PIRLS
 - ✓ Progress of International Reading Literacy Study
- PISA
 - ✓ Program of International Student Assessment



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

20

Overview of Studies

	TIMSS	PIRLS	PISA
Subjects	Mathematics and Science	Reading literacy	Science Literacy, Mathematics, Reading
Grades/age	2003: 4 th /8 th	4 th	15-years old
Participating Jurisdictions	2003: 49	2006: 46	2006: 57



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

21

Overview of Studies

	TIMSS	PIRLS	PISA
Cycle	4 years	5 years	3 years
Last release	TIMSS 2003	PIRLS 2006	PISA 2006
Test Booklets (last cycle)	12 booklets	12 booklets + Reader	13 booklets
Testing time	90 minutes	80 minutes	120 minutes
Questionnaire	Principals, Teachers, Students	Principals, Teachers, Students (Parents)	Principals, Students (Parents)



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

22

TIMSS Assessment Design

- 90 minute assessment at grade 8
- Rotated block assessment design: blocks of items were rotated among 12 booklets.
- No student answered all items
- Students encountered booklets with both mathematics and science items
- Each booklet contained both trend items (1995, 1999) and new items.



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

23

TIMSS Achievement Scales

- Overall Mathematics
 - ✓ At Grade 8: Number; Algebra; Measurement; Geometry; Data
 - ✓ At Grade 4: Number; Patterns, Equations, and Relationships; Measurement; Geometry; Data
- Overall Science
 - ✓ At Grade 8: Life Science; Chemistry; Physics; Earth Science; Environmental Science
 - ✓ At Grade 4: Life Science; Physical Science; Earth Science
- Also cognitive domain scales



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

24

The TIMSS 2003 Test Design

Book	Pos 1	Pos 2	Pos 3	Pos 4	Pos 5	Pos 6
1	M01	M02	S06	S07	M05	M07
2	M02	M03	S05	S08	M06	M08
3	M03	M04	S04	S09	M13	M11
4	M04	M05	S03	S10	M14	M12
5	M05	M06	S02	S11	M09	M13
6	M06	M01	S01	S12	M10	M14
7	S01	S02	M06	M07	S05	S07
8	S02	S03	M05	M08	S06	S08
9	S03	S04	M04	M09	S13	S11
10	S04	S05	M03	M10	S14	S12
11	S05	S06	M02	M11	S09	S13
12	S06	S01	M01	M12	S10	S14



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

25

What is PIRLS?

- International study of reading literacy at 4th grade
- Assessment of student proficiency in reading comprehension
- Extensive collection of data on context for learning to read



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

26

PIRLS 2006 Test Design

- 10 passages; 5 literary & 5 informational
- 126 items; 167 score points
- Multiple-choice and constructed-response questions
- PIRLS Reader

PIRLS 2006 Booklets													
	1	2	3	4	5	6	7	8	9	10	11	12	R
Part 1	L1	L2	L3	L4	I1	I2	I3	I4	L1	I2	L3	I4	L5
Part 2	L2	L3	L4	I1	I2	I3	I4	L1	I1	L2	I3	L4	I5



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

27

What is PISA?

- Program for International Student Assessment
- The main questions are:
 - ✓ Are students well prepared to meet the challenges of the future?
 - ✓ Are they able to analyze, reason, and communicate their ideas effectively?
 - ✓ Do they have the capacity to continue learning throughout life?



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

28

What is PISA?

- Measure how well 15-year-old approaching the end of compulsory schooling are prepared to meet the challenges of today's knowledge societies
- Does not focus on mastery of school curriculum, but on the ability to use knowledge and skills to meet "real-life challenges"



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

29

PISA Test Design

- Administered every 3 years
- Originally 3 major domains: Mathematics, Science and Reading
 - ✓ In 2003 included Problem Solving
- Every 9 years major emphasis on a domain
- In between the 9 years minor emphasis on a domain



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

30

PISA Test Design

	2000	2003	2006	2009
Reading	MAJOR	Minor	Minor	MAJOR
Mathematics	Minor	MAJOR	Minor	Minor
Science	Minor	Minor	MAJOR	Minor



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

31

PISA 2006 Test Design

PISA 2006 Booklets				
Form	Block 1	Block 2	Block 3	Block 4
1	S1	S2	S4	S7
2	S2	S3	M3	R1
3	S3	S4	M4	M1
4	S4	M3	S5	M2
5	S5	S6	S7	S3
6	S6	R2	R1	S4
7	S7	R1	M2	M4
8	M1	M2	S2	S6
9	M2	S1	S3	R2
10	M3	M4	S6	S1
11	M4	S5	R2	S2
12	R1	M1	S1	S5
13	R2	S7	M1	M3



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

32

Comments? Questions?



Copyright © IEA-ETS Research Institute (www.IERIInstitute.org)

Analysis with IRT

IERI Academy Winter 2009
Center of Global Assessment
ETS
Kentaro Yamamoto

Presentation Outline

- Classical Test Theory (CTT)
- Item Response Theory (IRT)
- Common IRT Models
- Parameter Estimations
- Features of IRT model
- Application of IRT for test development
- Consequence of assessment features on IRT modeling
- Weighting
- Booklet design
- Non-response
- Scoring accuracy

Establishing the Comparability

- Instrumentation
 - BQ, Cognitive, Instruction, Manual
- Sampling, Target Population, Sampling Frame
- Weighting
- Operation
- **Scoring**
- **Scaling (Non-response, Common parameters)**
- **Population modeling**
- **Reporting**

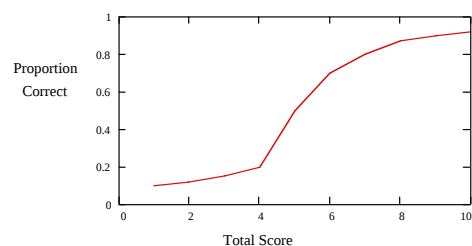
Moh's Scale of Mineral Hardness

	Sample	A	B	C	D	E	F
1	Talc	O	O	O	O	O	O
2	Gypsum	O	O	O	O	O	O
3	Calcite	X	O	O	O	O	O
4	Fluorite	X	X	O	O	O	O
5	Apatite	X	X	O	O	O	O
6	Orthoclase	X	X	X	O	O	O
7	Quartz	X	X	X	X	O	O
8	Topaz	X	X	X	X	O	O
9	Corundum	X	X	X	X	X	O
10	Diamond	X	X	X	X	X	X
		2.5	3.5	5.5	6.5	8.5	9.5

Typical Math Test

Student	Items											Total
	1	2	3	4	5	6	7	8	9	10	11	
A	O	O	O	O	O	O	O	O	O	X	O	10
B	O	O	O	O	X	O	O	O	O	O	X	9
C	O	O	O	O	O	O	X	O	X	X	O	8
D	O	X	O	O	O	X	O	O	O	O	X	8
E	O	O	O	O	X	O	O	X	X	O	X	7
F	X	O	O	X	O	O	X	O	O	X	X	6
G	O	O	X	X	O	X	O	X	X	X	X	4
H	O	O	X	O	X	X	X	X	X	X	X	2
Total	7	7	6	6	5	5	5	5	4	3	2	

Common Observation



Classical Test Theory

$$\text{Observed Score} = \text{True Score} + \text{Error Score}$$

$$O = T + e$$

Observed Score: test score

(can be standardized and/or transformed into various forms)

True Score: hypothetical mean of an infinite number of tests taken by an examinee when practice effects are disregarded, i.e., expectation of the scores

Error: amount of discrepancy between examinees' true scores and observed scores for population

Classical Test Theory

- Offers a diverse collection of techniques and formulas for item analysis, test development, scoring, scaling, and equating, as well as the evaluation of test reliability, validity and fairness
- Has a long history of theoretical and practical development with well developed body of knowledge
- Very useful for evaluating respondents skills and analysis of test scores
- Used often to produce parallel test forms

Limitations of CTT

- Sample Dependent
 - Statistics (e.g., difficulty, discrimination) that describe test items depend upon group of respondents who take the items
- Test form Dependent
 - Observed & true scores that describe how respondents perform depend upon the collection of items given to the students

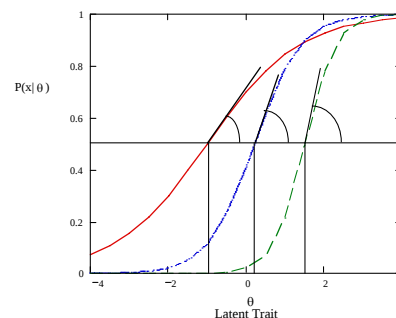
Limitations of CTT

- Same errors of measurement for all examinees regardless of examinees ability levels
 - Measurement error is population based
- Not able to match test forms for projected ability distribution
- Strictly parallel test do not exist
- Requires equating of
 - Item statistics for use in test assembly
 - Test scores for use across administrations

Item Response Theory

- Provides a consistent framework for thinking about and resolving testing problems
- Developed in 1960s to addresses some of the limitations of CTT, and offers a more flexible set of techniques
 - Estimate of standard error of measurement for each examinee
 - IRT pattern scoring does not require equating
 - Automated test assembly & CAT
- Reminds us that our testing goal is to measure an unobservable or latent construct, and that it is critical to consider the relationship of observable item responses to the latent construct

Item Characteristic Curve



Item Response Theory

- Offers a variety of probability models relating observable item responses, x_i , to the latent trait or ability, θ , and the item's measurement properties, $\underline{\beta}$ (item difficulty and discrimination, e.g.)

$$P(x_i = k | \theta, \underline{\beta}), \begin{cases} k = 0 \text{ or } 1 \text{ for dichotomous items} \\ k = 0, 1, \dots, K \text{ for polytomous items} \end{cases}$$

Advantage of IRT

- Because it makes stronger assumptions about the behavior of people and items, it allows us to solve problems that cannot be solved by classical test theory.
- The use of IRT allows us to answer many questions about a test before we ever administer it.

Advantage of IRT

- Evaluate the effectiveness of a test at different levels of ability
- Evaluate the ability of respondents who took different set of items
- Evaluate a set of items on a same scale established by a different set
- Design tests to measure best at specific ability level
- Develop new tests and investigate them without administering them
- Develop item statistics that do not change when the group of examinees change

Common IRT Models

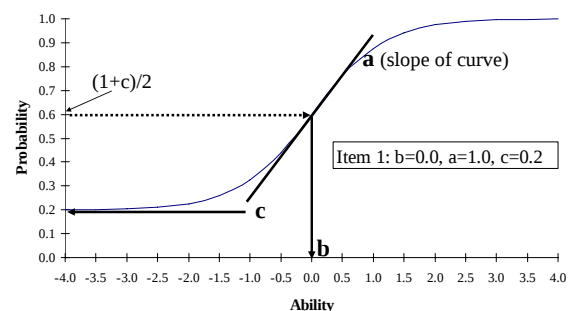
- Dichotomous IRT Models for Multiple Choice Items:
 - one-parameter logistic (1PL) model
 - two-parameter logistic (2PL) model
 - three-parameter logistic (3PL) model – most common
- Polytomous IRT Models for Multiple Responses:
 - graded response model
 - partial credit model

3PL Model

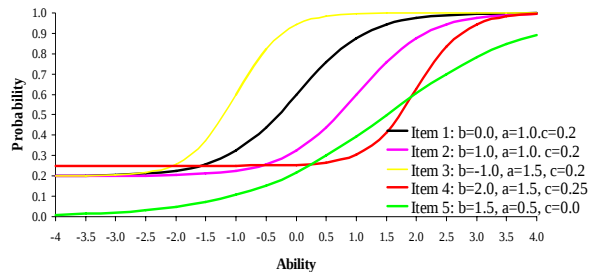
$$P(x_{ij} = 1 | \theta_j) = c_i + \frac{1 - c_i}{1.0 + \exp(-Da_i(\theta_j - b_i))}$$

- where
 - θ is student j ability
 - a_i is discriminating power for item i
 - b_i is item difficulty for item i , and
 - c_i is the "pseudo-guessing" parameter for item i

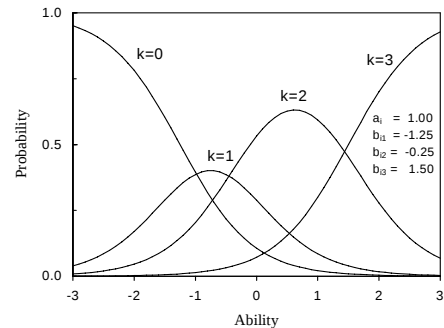
IRT Parameter Interpretation



Five ICCs



Graded Response Model: Category Response Functions (for item scores 0,1,2, and 3)



Common IRT Assumptions

- Unidimensionality: A single latent variable is sufficient to explain all responses to items.
 - Local Independence: At a given level of X , responses to all items are statistically independent.

$$P(U_1 = u_1, U_2 = u_2, \dots, U_n = u_n | \theta) = \prod_{i=1}^n P(U_i = u_i | \theta)$$

- Model-Data Fit
 - Appropriate use of an IRT model requires reasonable satisfaction of its assumptions.

Review of MML for Single Population

2PL IRT
$$P(x_{ij} = 1) = \frac{1}{1.0 + \exp(-Da_i(\theta_j - b_i))}$$

$$P(\underline{x}_j | \theta_j, \underline{a}, \underline{b}) = \prod_{i=1}^n P_i(\theta_j)^{x_{ij}} (1 - P_i(\theta_j))^{1-x_{ij}}$$

Integrate θ
$$P(\underline{x}_j | \underline{a}, \underline{b}) = \int_{\theta} P(\underline{x}_j | \theta, \underline{a}, \underline{b}) \cdot f(\theta) d\theta$$

Maximize likelihood
$$L(\underline{a}, \underline{b} | \underline{X}) = \prod_j L(\underline{a}, \underline{b} | \underline{x}_j)$$

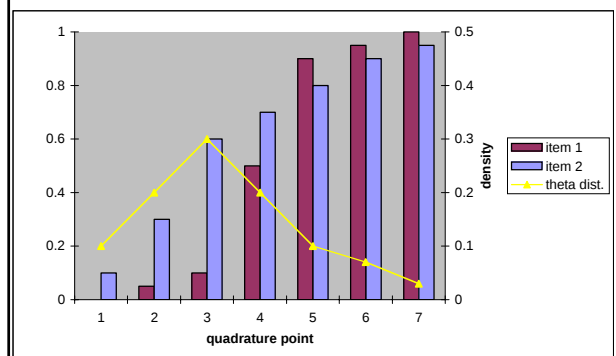
22

EM steps of Item calibration

- During E-step, posterior distribution as well as proportions correct for each items at each quadrature points are estimated based on the prior.
- During M-step likelihood function is maximized with respect to the model parameters.
- IRT linking is based on either proficiency distributions or the item parameters.

23

What E-step produces



Extend to Multiple Population

Integrate θ | group g $P(x_g | \alpha_g, g) = \int P(x_g | \theta, \alpha_g) f_g(\theta) d\theta$

Sum over groups $P(X | \alpha) = \sum_g \int P(x_g | \theta, \alpha_g) f_g(\theta) d\theta$

Maximize likelihood $L(\alpha | X) = \prod_j L(\alpha | x_j)$

25

Evaluate Fit of IRT Parameters

□ χ^2

• MD

• RMSD

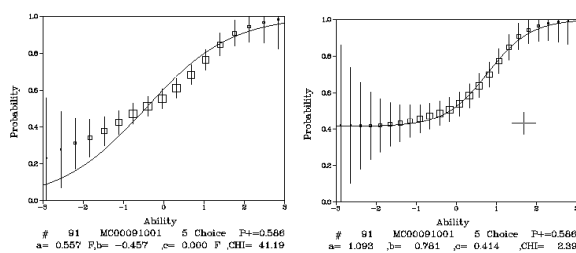
• Impact on ability estimate

$$MD_g = \int (\hat{P}_g(\theta) - P(\theta)) f_g(\theta) d\theta$$

$$RMSD_g = \sqrt{\int (\hat{P}_g(\theta) - P(\theta))^2 f_g(\theta) d\theta}$$

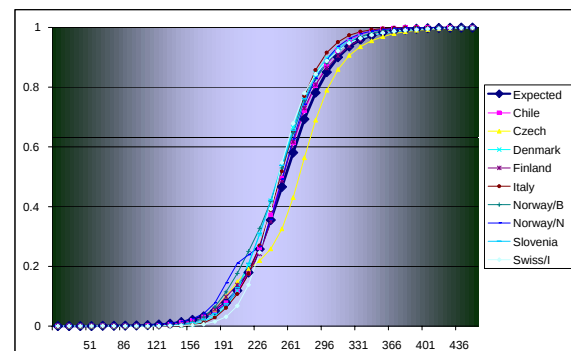
26

Assessing Model Fit: 1 & 3PL



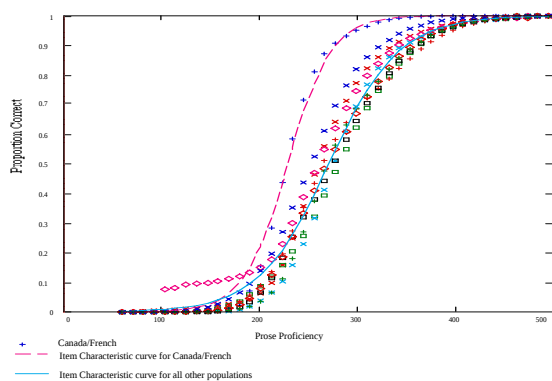
28

IRT Response Curve of Document 7-8



28

IRT Response Curve of Q- ITEM



29

Application of IRT Models to Test Development

- Item information functions
- Test information functions
- Standard error of measurement

Item Information

- Item information shows the contribution of an item to measurement precision with respect to ability parameter
- Important for test assembly, linking, virtual linking, assuring desired accuracy, so on

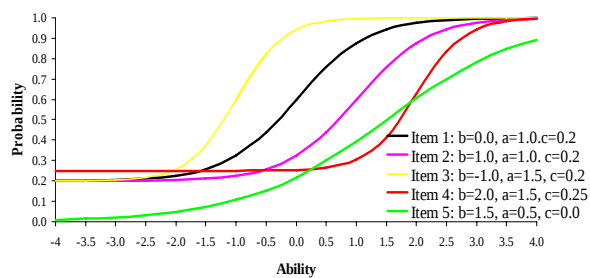
Item Information

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

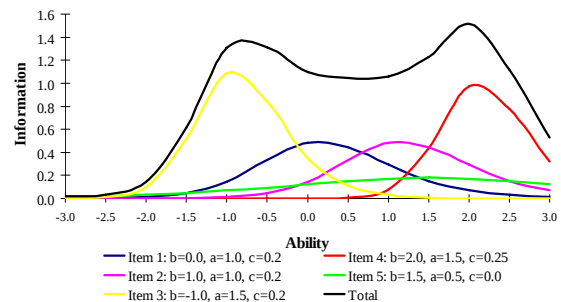
where

- $P_i(\theta)$ is the item response function,
- $Q_i(\theta) = 1 - P_i(\theta)$
- $P'_i(\theta)$ is the derivative of $P_i(\theta)$ with respect to θ

Item Characteristic Curves



Item Information Functions



Questions: Which abilities do these items measure best?
For what level is this test most appropriate?

Test Information

- Test information function is the sum of item information of the items in the test:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

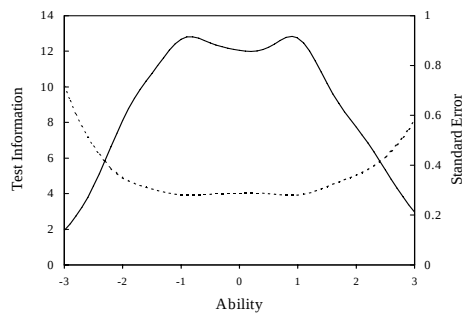
where n is the number of items in the test.

Standard Error

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

- Assesses measurement precision for each examinee level
- Used to calculate reliability

TIF & SE



Weights

TIMSS/PIRLS

$$FW = W_{\text{school}} * Ad_{\text{school}} * W_{\text{class}} * W_{\text{student}} * Ad_{\text{student}}$$

PISA

$$FW = Ad_{\text{subject}} * W_{\text{school}} * Ad_{\text{school}} * W_{\text{student}} * Ad_{\text{student}}$$

IALS/ALLS

$$FW = W_{\text{psu}} * W_{\text{household}} * Ad_{\text{household}} * W_{\text{respondent}} * Ad_{\text{respondent}} * Ad_{\text{post stratification}}$$

Booklet design

TIMSS: Bib-spiral of blocks of mixed subjects of Math(5) and Science(6), ~80 items,

PISA: Subject specific blocks, Reading, Math, Science.

IALS/ALL: Bib-spiral of 3 domains of Literacy, appended spiral design

39

Booklet Design TIMSS

- In 1999 TIMSS, 26 clusters of items were bib spiraled to produce 8 booklets

Booklet	Clusters																									
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	1	3	5					6										4	7						
2	2		1	3	5				6														4	7		
3	2			1	3	5				6										4	7					
4	2				1	3	5				6										4	7				
5	2	5				1	3	5				6										4	7			
6	2		5				1	3					6										4	7		
7	2	3		5										6									4	7		
8	2	1													6	3	5									

A(12m), B-H(12m), I-R(22m), S-V(10m), W-Z(10m)

40

Booklet Design 2003 PISA

Booklet	Cluster1	Cluster2	Cluster3	Cluster4
1	M1	M2	M4	R1
2	M2	M3	M5	R2
3	M3	M4	M6	PS1
4	M4	M5	M7	PS2
5	M5	M6	S1	M1
6	M6	M7	S2	M2
7	M7	S1	R1	M3
8	S1	S2	R2	M4
9	S2	R1	PS1	M5
10	R1	R2	PS2	M6
11	R2	PS1	M1	M7
12	PS1	PS2	M2	S1
13	PS2	M1	M3	S2

M1-M7,S1-S2,R1-R2,PS1-PS2: 30 min

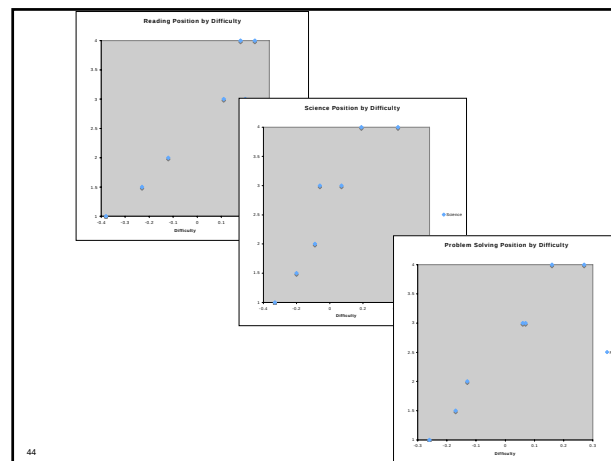
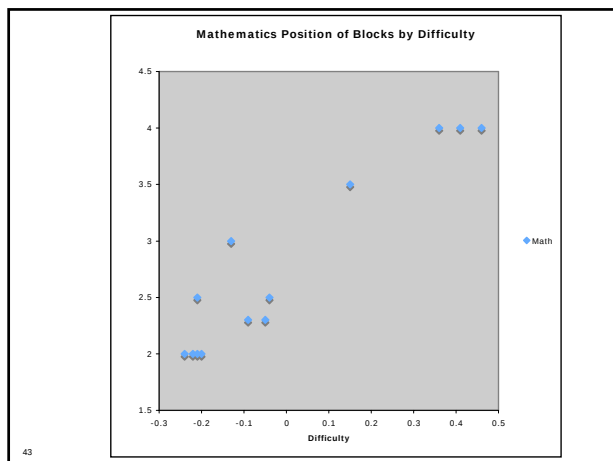
41

Booklet Design 2003 PISA

	M1	M2	M3	M4	M5	M6	M7	S1	S2	R1	R2	PS1	PS2
1	1	2		3						4			
2		1	2	3							4		
3			1	2	3							4	
4				1	2	3							4
5	4				1	2	3						
6		4				1	2	3					
7			4				1	2	3				
8				4				1	2	3			
9					4				1	2	3		
10						4				1	2	3	
11	3						4				1	2	3
12		3						4				1	2
13	2		3						4				1

M1-M7,S1-S2,R1-R2,PS1-PS2: 30 min

42



IALS: Average Percent Correct; Block by Block Order

Block Order	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Total
First	72	78	74	72	77	83	77	76
Second	78	76	77	78	72	83	79	78
Last	74	77	77	77	77	83	72	77

ALL :Average Proportion Correct; Block by Block Order

Block Order	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Total
First	62.6	64.8	68.9	66.2	64.8	63.1	60.1	48.7	62.4
Second	61.6	65.0	68.3	64.8	65.0	63.3	59.9	5.01	62.3

Booklet Design IALS/ALL

	b1	b2	b3	b4	b5	b6	b7	b8
b1		X	X		X			X
b2			X	X		X	X	
b3	X			X	X			X
b4	X	X				X	X	
b5		X		X		X		
b6	X		X		X			
b7	X		X					X
b8		X		X			X	

1-4: Prose&Document 5,6: Numeracy 7,8: Problem Solving

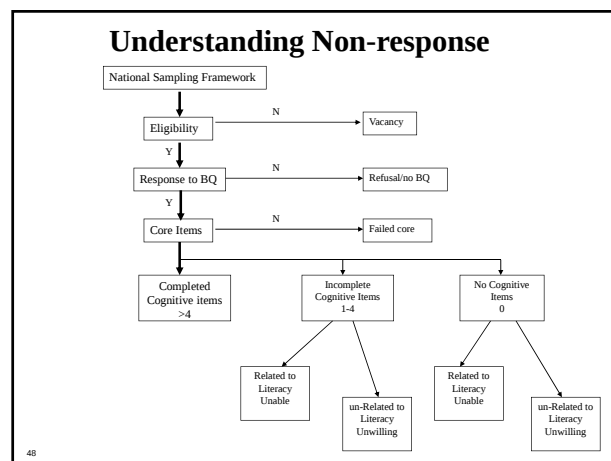
46

Booklet Design IALS/ALL

	B1	B2	B3	B4	B5	B6	B7	B8
1	1	2						
2	1	2						
3		1	2					
4		1		2				
5			1	2				
6	2		1	2				
7	2			1				
8		2			1			
9	1					2		
10		1				2	2	
11			1		2			
12				1		2		
13		2			1			
14				2	1			
15	2						1	
16			2				1	
17					1	2		
18					2	1		
19	1							2
20		1					2	
21			1				2	2
22				1				1
23	2							1
24		2	2					1
25				2				1
26					2			
27							1	2
28							2	1

1-4: Prose&Document 5,6: Numeracy 7,8: Problem Solving

47



Scoring Open Ended Questions

TIMSS: 25-40% open ended, extensive training, sample booklets scoring.

PISA: 30% open ended, extensive training, sample booklets scoring.

IALS/ALLS: 100% open ended, extensive training, exchange booklets across countries to sample booklets scoring, correct scoring if necessary.

49

IRT Modeling/Scaling

TIMSS: moved from Rasch(95) to 3pl+PC(99 on)

Common item parameters for all countries

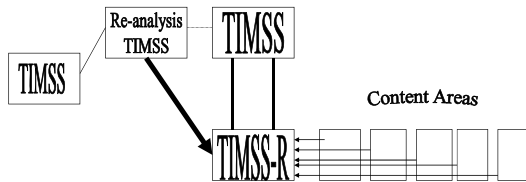
PISA: Multi dimensional Rasch, limited 2pl, Common item parameters for all countries

IALS/ALL: 2pl IRT, Subset of items may be different from common item parameters for

50

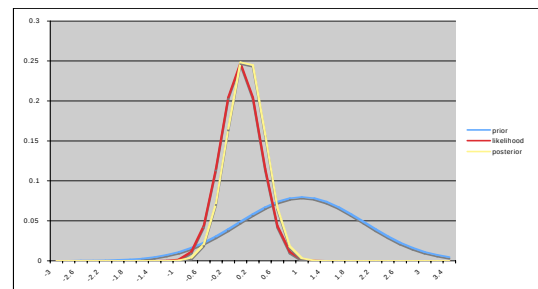
Trend linkings Example

- **Re-analysis of 1995 TIMSS**
 - To address 1PL-3PL, and re-center to mean of 500, and sd. of 100 based on grade 8 only
- **Common scale of 1995 TIMSS-1999TIMSS-R**
 - Trend analysis
- **Content area of each domain 1999TIMSS-R**
 - Profile of content areas by setting the distributions to TIMSS-R



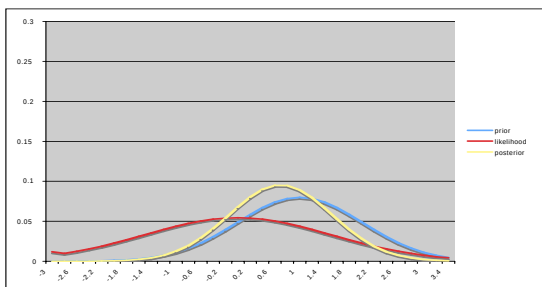
51

Impact of Prior Distribution on Proficiency Estimate 1



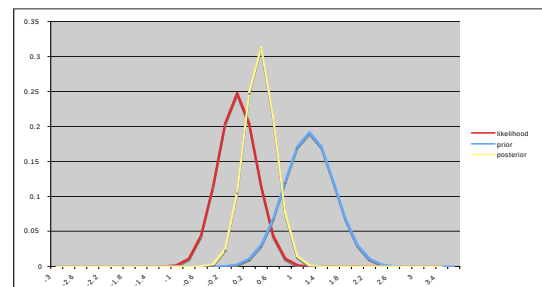
52

Impact of Prior Distribution on Proficiency Estimate 2



53

Impact of Prior Distribution on Proficiency Estimate 3



ETS Listening Learning Leading

Modeling International Large Scale Survey Assessment Data

IERI Fall Academy 2008
Matthias von Davier
Educational Testing Service

On international large scale educational surveys

- International large scale assessments
- Assessment designs in large scale assessments
- Background data in international assessments
- Deriving skill distributions in subgroups of interest
- Research on international surveys (the IERI ad)
- Outlook

ETS Listening Learning Leading

(International) Large Scale Assessments

Some key principles:

- Common assessment framework
- Common scale(s) across countries & languages
- Coordinated administrations
- Repeated, linked assessments
- Group level reporting
- Curriculum (coverage?) or life-skills based
- International comparisons

ETS Listening Learning Leading

(International) Large Scale Educational Assessments

The mother of all Large Scale Assessments:
National Assessment of Educational Progress

- NAEP (..., 2008, ...), government mandated since 1960ies

International Large Scale Assessments

- TIMSS (1995, 1999, 2003, 2007, 2011, ...)
- PIRLS (1992, 2001, 2011, ...)
- PISA (2000, 2003, 2006, 2009, ...)
- IALS ('94, '96, '98), ALL (2003)
- PIAAC (2011, ...)

ETS Listening Learning Leading

For example: TIMSS

TIMSS (1995, 1999, 2003, 2007, 2011, ...)

- Trends in Mathematics and Science Study
- An IEA study (IEA? - see next slide...)
- Currently 60-some countries participating
- International Study Center (ISC), Boston College
- ETS provides psychometric consulting and statistical software for ISC's analyses & reporting

ETS Listening Learning Leading

IEA IEA

Page 1 of 3

IEA
International Association for the Evaluation of Educational Achievement

IEA

From the IEA Headquarters
Headquarters & Branches
Mission Statement
Council Members & Staff
Large Scale
IEA Publications
IEA Council Assembly
Standing Committee
Executive Board
Data Processing and Research Center
IEA-ETS Research Institute
Global Partnership

STUDIES
Learning on IEA Study
Current Studies
Completed Studies
IEA Studies Evaluation
IEA Studies Database and Data Analysis

IEA-ETS Academy: 11-19 May 2008
PISA 2006 results released
Page 1 of 3

IEA-ETS Academy: 11-19 May 2008
PISA 2006 results released
Page 1 of 3

The purpose of the International Civic and Citizenship Education Study (ICCS) is to investigate the ways in which young people are prepared to undertake their roles as citizens.

Current Studies
Learning
TIMSS Advanced 2008
Replication of TIMSS 1995 Advanced Mathematics and Physics
First administered in 10 countries in 1995, TIMSS Advanced assesses school/learning objectives with special preparation in advanced mathematics and physics. Countries that...

4/25/2008
Listening Learning Leading

For example: PISA

PISA (2000, 2003, 2006, 2009, ...)

- Programme for International Student Assessment
- An OECD study, 67 countries in 2009...
- Australian Council of Educational Research (ACER) is conducting the study for OECD
- ETS provides consulting on technical issues, ETS representative heads the reading literacy expert group, ...

PISA Page 1 of 3

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
OECD

Search

OECD Programme for International Student Assessment (PISA)

PISA (Home)

- What PISA is
- What PISA measures
- What PISA produces
- Participating Countries
- Contact

Are students well prepared for future challenges? Can they analyse, reason and communicate effectively? Do they have the capacity to continue learning throughout life? The OECD Programme for International Student Assessment (PISA) answers these questions and more, through its surveys of 15-year-olds in the principal industrialised countries. Every three years, it assesses how far students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society.

What's new

PISA 2006 results
24-Dec-2007
Learn about the most comprehensive international picture of science learning today. For all data and a comprehensive analysis click here.

PISA brochure
23-Feb-2007
This brochure on the OECD's Programme for International Students Assessment describes what the programme has achieved so far and sets out plans for the project up until 2015.

Don't Miss

- PISA 2006 PISA
- PISA brochure
- Why is PISA
- PISA auf Deutsch
- PISA en Français
- PISA en español
- Directorate for Education
- How are students doing in science today?
- How are students doing in science today?
- PISA 2006 results - Analysis and Data
- Volume 1 - Reading
- Volume 2 - Science
- Volume 3 - Mathematics
- Volume 4 - Framework
- Volume 5 - Technical Report
- Volume 6 - Policy Report
- Volume 7 - Policy Report
- Volume 8 - Policy Report
- Volume 9 - Policy Report
- Volume 10 - Policy Report
- Volume 11 - Policy Report
- Volume 12 - Policy Report
- Volume 13 - Policy Report
- Volume 14 - Policy Report
- Volume 15 - Policy Report
- Volume 16 - Policy Report
- Volume 17 - Policy Report
- Volume 18 - Policy Report
- Volume 19 - Policy Report
- Volume 20 - Policy Report
- Volume 21 - Policy Report
- Volume 22 - Policy Report
- Volume 23 - Policy Report
- Volume 24 - Policy Report
- Volume 25 - Policy Report
- Volume 26 - Policy Report
- Volume 27 - Policy Report
- Volume 28 - Policy Report
- Volume 29 - Policy Report
- Volume 30 - Policy Report
- Volume 31 - Policy Report
- Volume 32 - Policy Report
- Volume 33 - Policy Report
- Volume 34 - Policy Report
- Volume 35 - Policy Report
- Volume 36 - Policy Report
- Volume 37 - Policy Report
- Volume 38 - Policy Report
- Volume 39 - Policy Report
- Volume 40 - Policy Report
- Volume 41 - Policy Report
- Volume 42 - Policy Report
- Volume 43 - Policy Report
- Volume 44 - Policy Report
- Volume 45 - Policy Report
- Volume 46 - Policy Report
- Volume 47 - Policy Report
- Volume 48 - Policy Report
- Volume 49 - Policy Report
- Volume 50 - Policy Report
- Volume 51 - Policy Report
- Volume 52 - Policy Report
- Volume 53 - Policy Report
- Volume 54 - Policy Report
- Volume 55 - Policy Report
- Volume 56 - Policy Report
- Volume 57 - Policy Report
- Volume 58 - Policy Report
- Volume 59 - Policy Report
- Volume 60 - Policy Report
- Volume 61 - Policy Report
- Volume 62 - Policy Report
- Volume 63 - Policy Report
- Volume 64 - Policy Report
- Volume 65 - Policy Report
- Volume 66 - Policy Report
- Volume 67 - Policy Report
- Volume 68 - Policy Report
- Volume 69 - Policy Report
- Volume 70 - Policy Report
- Volume 71 - Policy Report
- Volume 72 - Policy Report
- Volume 73 - Policy Report
- Volume 74 - Policy Report
- Volume 75 - Policy Report
- Volume 76 - Policy Report
- Volume 77 - Policy Report
- Volume 78 - Policy Report
- Volume 79 - Policy Report
- Volume 80 - Policy Report
- Volume 81 - Policy Report
- Volume 82 - Policy Report
- Volume 83 - Policy Report
- Volume 84 - Policy Report
- Volume 85 - Policy Report
- Volume 86 - Policy Report
- Volume 87 - Policy Report
- Volume 88 - Policy Report
- Volume 89 - Policy Report
- Volume 90 - Policy Report
- Volume 91 - Policy Report
- Volume 92 - Policy Report
- Volume 93 - Policy Report
- Volume 94 - Policy Report
- Volume 95 - Policy Report
- Volume 96 - Policy Report
- Volume 97 - Policy Report
- Volume 98 - Policy Report
- Volume 99 - Policy Report
- Volume 100 - Policy Report

http://www.pisa.oecd.org/pages/0,2967,en_32252551_32252551_1_1_1_1_00.html

4/25/2008

ETS Listening Learning Leading 8

For example: PIAAC

PIAAC (2011, ...)

- Programme for the International Assessment of Adult Competencies (PIAAC)
- An OECD study, 20-some countries in 2011...
- TIMSS, PISA, NAEP are school based assessments, and sample students (or classes) within schools
- PIAAC (like IALS, ALL) is a household survey, and samples adults 16-65 year old from households

Programme for the International Assessment of Adult Competencies (PIAAC) Page 1 of 2

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
OECD

Ministère de l'Éducation, du Travail et des Affaires sociales
Direction de l'Emploi, du Travail et des Affaires sociales

Statistiques, Politiques de l'Emploi - Programme for the International Assessment of Adult Competencies (PIAAC)

Programme for the International Assessment of Adult Competencies (PIAAC)

Background: This paper: www.oecd.org/employment/naep/. The Programme for the International Assessment of Adult Competencies (PIAAC) aims at developing a strategy to address the supply and demand of competencies that would:

- identify and measure differences between individuals and countries in competencies believed to underlie both personal and societal success;
- assess the impact of these competencies on social and economic outcomes at individual and aggregate levels;
- gauge the performance of education and training systems in generating required competencies; and
- help to clarify the policy levers that could contribute to enhancing competencies.

The papers listed here are background documents that have been used to help develop such a strategy:

- Draft Strategy Paper: Policy Objectives, Strategic Options and Cost Implications
- "What Do Policy Makers Need to Know About the Skills of Young People and the

http://www.oecd.org/document/576,3343,46_2649_33927_34474617_1_1_1_1_00.html

4/25/2008

ETS Listening Learning Leading 10

Assessment designs in large scale assessments

- All domains are covered broadly, the number of sub-domains and items is determined during framework development
- A large number of items is developed and organized into blocks, which are combined into test booklets.
- Test booklets include an incomplete, but balanced subset of all possible combinations of the number of blocks administered
- -> Balanced incomplete block (BIB) design

Exhibit 11: TIMSS 2007 Student Achievement Booklet Design - Fourth and Eighth Grades

Student Achievement Booklet	Assessment Blocks			
	Part 1		Part 2	
Booklet 1	M01	M02	S01	S02
Booklet 2	S02	S03	M02	M03
Booklet 3	M03	M04	S03	S04
Booklet 4	S04	S05	M04	M05
Booklet 5	M05	M06	S05	S06
Booklet 6	S06	S07	M06	M07
Booklet 7	M07	M08	S07	S08
Booklet 8	S08	S09	M08	M09
Booklet 9	M09	M10	S09	S10
Booklet 10	S10	S11	M10	M11
Booklet 11	M11	M12	S11	S12
Booklet 12	S12	S13	M12	M13
Booklet 13	M13	M14	S13	S14
Booklet 14	S14	S15	M14	M15

LSA frame summary

- Students respond to 3-6 short blocks of items, blocks appear in multiple locations within multiple booklets
- Multiple test forms have to be linked, blocks from previous assessments are combined with new blocks
- Trends over time often have to be reported
- Differences between countries in ability distributions, but no DIF is a basis for comparability
- Sample items (released old form items) have to be provided to stakeholders.

N9. Which of the boxes X, Y, or Z has the LEAST mass?



- A. X
- B. Y
- C. Z
- D. All three boxes have the same mass.

Multiple LSA components

In addition to the cognitive items administered, students also fill out questionnaires, as do their teachers or principals

Students are asked about a variety of issues related to their academic achievement, their leisure activities, etc.

Teachers may be asked about their instructional practices, and their assessment of the school environment

School principals may be asked about school infrastructure, and teacher professional development

Background Data in Skill Surveys

Background data serve (at least) two purposes in international surveys of skills:

- BV's are utilized in modeling the population distribution of skills. BV's help reducing uncertainty as well as biases of group comparisons
- Background variables (BV) are central in understanding differences between higher and lower skilled groups of the target population

OECD Programme for International Student Assessment 2006

OECD
PISA

Canada

STUDENT QUESTIONNAIRE

PISA SCHOOL ID [] [] [] [] [] []
PISA STUDENT ID [] [] [] [] [] []

Background Data for Modeling

Background variables (BV) are needed as auxiliary information in a population model used to reduce uncertainty of measurement and to provide unbiased imputations of skill variables:

- BV are selected and preprocessed to be integrated in a latent regression model (mixed effects model, multilevel IRT model)
- Conditional distributions of skills given BV are estimated, and imputations (plausible values) are drawn for reporting group statistics and variance estimation

Background Data as Major Reporting Variables

Background variables (BV) are central to reporting, and are used to derive indicators for reporting skills in subgroups of interest to stakeholders and policy makers.

- BV may directly assess the group definition of interest (gender, ethnicity)
- BV may be aggregated to indicator variables such as attitudes towards technology, reading behaviors, etc.
- BV may be proxies to variables we cannot directly assess or we are now allowed to ask in some or all countries (SES, gross income)



19

Listening Learning Leading

Challenges of using Background Data

Background variables are selected based on experts knowledge of relevant differences. In international surveys, this selection is faced with multiple challenges:

- Comparability across countries:
 - A variable that is quite predictive of skills in one country may or may not be so in other countries
 - The response formats used can be subject to cultural differences in response sets (e.g. avoiding extreme responses vs. tendency to avoid middle/undecided responses)
- Relationships between questions, or collections of questions used to derive indicators, may vary across countries.
- Some background variables very useful as a proxy to SES in one country may not vary much in some other countries (number of bathrooms in a house in the USA vs. Western Europe)



20

Listening Learning Leading

Other issues in LSA, not covered today...

- Translation into multiple target languages, translation verification
- Target population, definition, inclusions, exclusions, sampling of students within educational institutions
- Scoring constructed response items, rater agreement, cross country comparability of rating standards
- Weighting, non-response adjustments, replacement, response rate, design effect...



21

Listening Learning Leading

Putting it all together:

There are many pieces to put together:

- Cognitive domain task responses
- Complex sample of item responses
- Countries provide a representative sample as the basis of international comparisons
- Assume we have clean data, received in time from all participating countries:

How are the different sources of information coming together, and what do the results look like?



22

Listening Learning Leading

Modeling large scale assessment data

- Reported are student proficiency distributions by countries and in subgroups within country
- IRT based methods are used due to the level of complexity of the data
- A multilevel IRT model with predictors based on the background data is used to analyze the data
 - Often called conditioning or latent regression model
- Reporting uses imputations (plausible values) to generate distributions of proficiencies, given reporting (major background) variables.



23

Listening Learning Leading

Latent regression IRT

The model LR-IRT represents is a latent regression model or a multi-level, better: 2-level, IRT model:

- The first level is the measurement model, a multidimensional, simple-structure IRT model
- The second level is a linear regression of the (multidimensional) latent trait θ on covariates, referred to as background variables in large scale assessments



24

Listening Learning Leading

LR-IRT Measurement model, 1

The LR-IRT measurement model is

- A composition of standard one - dimensional IRT models for multiple scales.
- Each item measures only one scale
- That means simple structure
- Which implies a factored likelihood



25

Listening Learning Leading

LR-IRT Measurement model, 2

- For dichotomous multiple choice items, often the 3PL is used
- For dichotomous constructed response items, the 2PL should be used
- For polytomous items, the generalized partial credit model (GPCM) is used
- One can use the Rasch model as well (Conquest, PISA & TIMSS 1995)



26

Listening Learning Leading

LR-IRT Measurement model, 3

- IRT estimation is carried out before the latent regression is estimated
- -> item parameters are considered as fixed and known constants in LR-IRT
- Adams, Wilson and Wu show that impact is minimal for the Rasch model case
- Patz & Junker talk about this as adopting a *divide and conquer* strategy



27

Listening Learning Leading

LR-IRT Estimation

We use the (EM) algorithm, a standard tool used a lot for IRT and other latent variable model estimation problems

- E-step: Estimate posterior means and posterior (co-)variances
- M-step: Maximum likelihood estimates of the latent regression parameters



28

Listening Learning Leading

LR-IRT Estimation, EM-cycle

The E-Step produces preliminary estimates of the posterior means and variances given previous regression parameters

The M-step produces preliminary estimates of the regression parameters given the previous E-steps estimates of posterior moments

This cycle is repeated until parameters converge or a predetermined maximum of cycles is reached



29

Listening Learning Leading

Drawing Plausible Values

- NAEP and other survey assessments report subgroup estimates, not individual scores
- Covariates are used to incorporate reporting variables and other auxiliary information into a comprehensive model where they help avoiding biased group estimates
- NAEP uses multiple imputations derived from the LR-IRT model estimates to generate subgroup means, variances and other statistics



30

Listening Learning Leading

Plausible Values (PVs)

- Plausible values (PV's) are multiple imputations drawn from examinee posterior distributions
- Plausible values together with sampling weights are also used to estimate variances of the reported statistics
- This is done outside of LR-IRT, LR-IRT just generates the PVs using the regression estimates relating proficiency to background data



31

Listening. Learning. Leading.

Summary

- Survey assessments collect student proficiency data using sparse responses from multiple booklets
- In addition, multiple other sources of information are available on student background
- Latent regression models integrate IRT based proficiency estimation with population models using background data
- Reporting is based on group-level statistics aggregated using major background variables like gender, ethnicity, school type, parents education...



32

Listening. Learning. Leading.

Latent Regressions in DESI / MGROUP & Drawing Plausible Values

IERI Fall Academy 2008
Matthias von Davier
ETS
Research & Development

ETS Confidential & Proprietary

Overview

- Latent regression in a nutshell
 - i.e., DESI / MGROUP in a nutshell
- Measurement model, Population model
- Estimation
- Drawing plausible values
- Outlook

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP in a nutshell

The model estimated by DESI is a latent regression model or a multi-level, better: 2-level, IRT model:

- The first level is the measurement model, a multidimensional, simple-structure IRT model
- The second level is a linear regression of the (multidimensional) latent trait θ on covariates, referred to as background variables in NAEP

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

Let $x_1, \dots, x_I$ denote item responses, θ the latent trait, and let the $z_1, \dots, z_K$ denote the covariates:

- The IRT model donates:
 $P(x_1 \dots x_I | \theta) = \prod_{i=1}^I P(x_i | \theta)$
- The latent linear regression donates:
 $f(\theta | z_1, \dots, z_K) = \phi(\theta | \mu(z_1 \dots z_K), \Sigma)$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

The marginal probability of observed responses $x_1, \dots, x_I$ given the $z_1, \dots, z_K$ covariates is then

$$P(x_1 \dots x_I | z_1 \dots z_K) = \int_{\theta} \left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_1 \dots z_K), \Sigma) d\theta$$

The $z_1, \dots, z_K$ determine the prior mean $\mu(z_1 \dots z_K)$ of the ability distribution, Σ denotes the common (co-)variance (matrix), $\phi(\cdot)$ denotes the (multivariate) normal density.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI / MGROUP model

Bayes' theorem is applied to find the posterior distribution of θ given observed responses $x_1, \dots, x_I$ and the $z_1, \dots, z_K$ covariates

$$P(\theta | x_1 \dots x_I, z_1 \dots z_K) = \frac{\left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_1 \dots z_K), \Sigma)}{P(x_1 \dots x_I | z_1 \dots z_K)}$$

The posterior distribution of θ is central to many subsequent calculations, it is used to calculate posterior means and variances (and finally PV's)

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

DESI Measurement model, 1

The DESI measurement model is

- A composition of standard one - dimensional IRT models for multiple scales.
- Each item measures only one scale
- That means simple structure
- Which implies a factored likelihood

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

7

DESI Measurement model, 2

- For dichotomous multiple choice items, often the 3PL is used
- For dichotomous constructed response items, the 2PL should be used
- For polytomous items, the generalized partial credit model (GPCM) is used
- One can use the Rasch model as well (Conquest, PISA, and TIMSS in 1995)

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

8

DESI Measurement model, 3

- IRT estimation is carried out before the DESI latent regression is estimated
- -> item parameters are considered as fixed and known constants in DESI, not as estimated quantities
- Adams, Wilson and Wu show that impact is minimal for the Rasch model case
- Patz & Junker talk about adopting a divide and conquer strategy

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

9

This is roughly DESI Phase 1

- Read in item parameters from previous IRT estimation, use as fixed constants
- Read in observed responses and covariates for each examinee
- Calculate likelihood $L(x_{1..}x_I|\theta) \sim P(x_{1..}x_I|\theta)$ for each examinee, store for next phase
- Calculate covariance matrix of covariates, $KOV(z_{1..}z_K)$ and store for next phase

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

10

DESI Population model, 1

- The population model describes the relationship between covariates (background variables) and the ability distribution.
- In DESI it is assumed that a latent regression can be used to describe this relation

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

11

DESI Population model, 2

The regression is latent since the dependent variable is not directly observed

Recall that we only have:

$$P(\theta | x_{1..}x_I, z_{1..}z_K) = \frac{\left[\prod_{i=1}^I P(x_i | \theta) \right] \phi(\theta | \mu(z_{1..}z_K), \Sigma)}{P(x_{1..}x_I | z_{1..}z_K)}$$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

12

DESI Population model, 3

Instead of an observed dependent variable, we use the posterior mean and posterior variance in the regression estimation

$$E(\theta^q | x_1 \dots x_I, z_1 \dots z_K) = \int_{\theta} \theta^q P(\theta | x_1 \dots x_I, z_1 \dots z_K) d\theta$$

$$V(\theta | x_1 \dots x_I, z_1 \dots z_K) = E(\theta^2 | x_1 \dots z_K) - (E(\theta | x_1 \dots z_K))^2$$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

13

DESI Population model, 4

For the regression in question, we have the usual system that relates a linear combination of the covariates for each examinee n to the posterior mean of this examinee

$$E(\theta | x_{n1} \dots x_{nI}, z_{n1} \dots z_{nK}) = \sum_{k=1}^K \beta_k z_{nk} + \varepsilon_n$$

The β_k are regression coefficients that have to be estimated, which is roughly what Phase II of DESI does.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

14

DESI Estimation

DESI uses an Estimation – Maxi-mization (EM) algorithm, a standard tool used a lot for IRT and other latent variable model estimation problems

- E-step: Estimate posterior means and posterior (co-)variances
- M-step: Maximum likelihood estimates of the latent regression parameters

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

15

DESI Estimation, E-step

E-step estimates posterior means and posterior (co-)variances:

- Calculation of posterior moments requires potentially multidimensional (numerical) integration
- There are many ways to do that, approximate solutions, smart integration methods, not so smart brute force methods or Monte- Carlo methods
- More about different E-step versions next time

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

16

DESI Estimation, M-step

The M-step in DESI uses posterior means and covariates to estimate regression parameters. It is assumed that we can use the OLS estimator.

- OLS estimator is $\beta = (Z^T Z)^{-1} Z^T \theta$, with posterior mean vector (matrix) θ
- The Σ estimator is the sum of the regression error variance $V(\varepsilon)$ and the average posterior variance $E(V(\theta | x_1 \dots z_k))$

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

17

DESI Estimation, EM-cycle

The E-Step produces preliminary estimates of the posterior means and variances given previous regression parameters

The M-step produces preliminary estimates of the regression parameters given the previous E-steps estimates of posterior moments

This cycle is repeated until parameters converge or a predetermined maximum of cycles is reached

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

18

EM is roughly DESI Phase II

The EM cycles can be time consuming, but modern hardware allow samples of 200,000 students or more with 500 background variables and 5 math scales:

- Computational cost is not an issue any more, large samples of students are fine
- More of an issue are small student samples with large numbers of background variables, numerical instabilities can result.

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

19

Drawing Plausible Values

- NAEP and other survey assessments report subgroup estimates, not individual scores
- Covariates are used to incorporate reporting variables and other auxiliary information into a comprehensive model where they help avoiding biased group estimates
- NAEP uses multiple imputations derived from the DESI model estimates to generate subgroup means, variances and other statistics

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

20

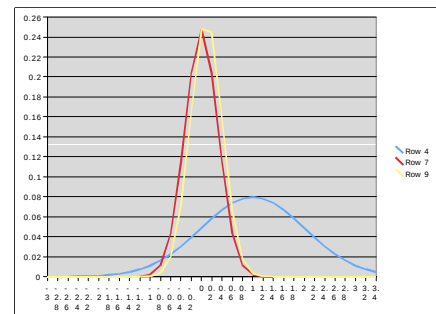
Plausible Values ~ Phase III

- Plausible values (PV's) are multiple imputations drawn from examinee posterior distributions
- Plausible values together with sampling weights are also used to estimate variances of the reported statistics
- This is done outside of DESI, DESI just generates the PV's using the posterior moments we talked about

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

21

Recall $\text{Prior} \times \text{Likelihood} = \text{Posterior}$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

22

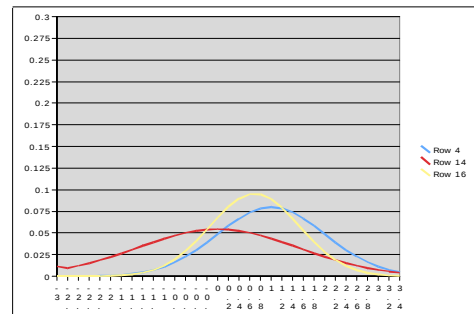
Testing versus Survey Assessments

- Individual testing provides a narrow likelihood based on many items, provides a lot of information
 - prior does not matter much
 - population needed neither needed nor wanted
- Survey assessments use fewer items, the likelihood is broader, represents less information
 - priors matter, covariates of proficiency are useful
 - population model providing priors is essential

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

23

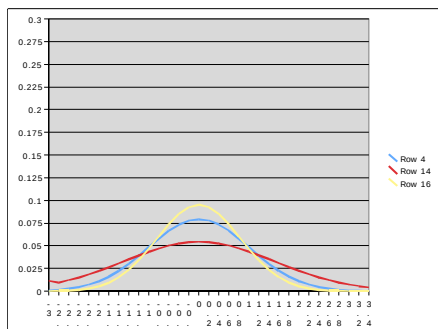
Ed. Surveys: Prior Information, $M=1.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

24

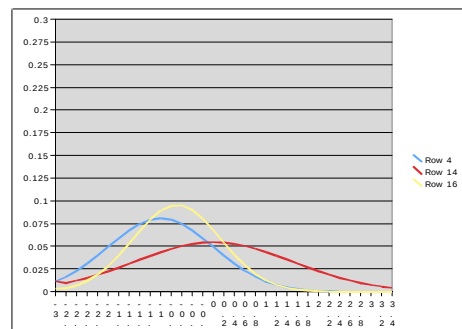
Ed. Surveys: Prior Information, $M=0.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

25

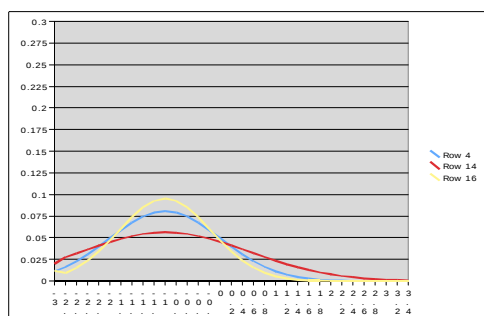
Ed. Surveys: Prior Information, $M=-1.0$



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

26

Prior Information, $M=-1.0$, and Likelihood $M=-1.0$ (below average Joe)



Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

27

!!! Priors are estimated, not fixed !!!

- Many average-Joe cases go into the prior
- Easy(?) to see in multiple-group models
- In latent regressions, the average effect of background variable on posterior is estimated
- Likelihood and prior do not need to agree
 - If not, posterior moves away, given information
 - If in agreement, prior*likelihood works like a charm
- 'borrowing information' from similar cases is the key concept

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

28

Plausible Values ~ Phase III

- *DESI provides estimates of subgroup means based on PVs for user defined grouping variables*
- *PVs are a means to an end, not optimal for anything, but very good for many types of analyses*
- *Many background variable effects are represented in PVs, allows analyses using these without bias*

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

29

Summary

- Latent regression model extends IRT
- Allows the estimation of effects of background variables
- Provides average effects based on sample correlations between θ and BQ
- Used for prior distributions, borrowing information from similar cases

Matthias von Davier: DESI Theory and Practice 1
ETS Confidential & Proprietary

30

出席國際學術會議心得報告

計畫編號	NSC 97-MOE-S-003-003
計畫名稱	國際數學與科學教育成就趨勢調查2007 (4/4) 甲
出國人員姓名 服務機關及職稱	任宗浩 國立台灣師範大學科學教育中心 助理研究員 李哲迪 國立台灣師範大學科學教育中心 助理研究員
會議時間地點	2009 年 5 月 26 日- 5 月 28 日 Singapore
會議名稱	Analysis of Large Scale Survey Data

一、 與會歷程

此工作坊為 ACER (Australian Council for Educational Research) 所主辦。ACER 是澳洲非常重要的教育評鑑機構，同時也是負責統籌執行 PISA (The Programme for International Student Assessment) 的單位。此工作坊之目的在介紹大型調查資料的統計分析方法。工作坊的講者是 ACER 的資深研究員—Wolfram Schulz 博士。工作坊分三天進行（議程如下表）。第一天的主題是取樣誤差。第二天的主題是運用似真值進行資料分析。第三天的主題是階層線性模式分析。工作坊的每個場次都是先介紹概念和理論，然後實做分析，交錯進行。

Tuesday 26 May, 2009

09.30 – 11.00 Introduction and Overview

Presentation: Sampling and weighting

11:00 – 11:30 Coffee break

11:30 – 13:00 Presentation: Estimation of sampling variance

Practice session: Calculating sampling variance

13:00 – 14.00 Lunch break

14.00 – 15.00 Presentation: Using SPSS for the estimation of sampling variance

Practice session: Learning how to use SPSS module

15:00 – 15:15 Short coffee break

15:15 – 16.00 Practice session: Reproducing tables from report in international studies

Wednesday 27 May, 2009

09.30 – 10.30	Presentation: Scaling methodologies used in international studies Practice session: Estimation of combined measurement and sampling error
11:00 – 11:30	Coffee break
11:30 – 13:00	Practice session: Analysis of plausible values
13:00 – 14.00	Lunch break
14.00 – 15.00	Presentation: Regression analysis
15:00 – 15:15	Short coffee break
15:15 – 16.00	Practice session: Estimating regression models with large scale survey data

Thursday 28 May, 2009

09.30 – 11.00	Concepts: Multi-level modelling Introduction to software packages for multi-level modelling
11:00 – 11:30	Coffee break
11:30 – 13:00	Practice session: Estimating multi-level models with large scale survey data
13:00 – 14.00	Lunch break
14:00 – 15.00	Practice session: Comparing multi-level and single-level regression results
15:00 – 15:15	Short coffee break
15.15 – 16.00	Conclusion and open discussion

二、 與會心得

此工作坊最大的特色是 ACER 所開發的統計分析軟體—Replicates。此軟體是專為大型調查資料分析而設計的。在運用大型調查研究所蒐集的資料進行統計推論時，所有統計量的標準誤都必須考慮測量誤差和抽樣誤差。以 TIMSS 和 PISA 的資料為例，其測量誤差以五個似真值來體現，而抽樣誤差則要透過 Jackknife 的方法來估計。為了取得測量誤差和抽樣誤差的估計，相同的統計估計的計算要重複執行將近 80 次。在實際的操作上，需要撰寫巨集語法來完成重複的計算。Replicates 的功能就是在產生 SPSS 的巨集語法，以計算諸如平均值、百分比、相關係數、迴歸方程的參數等等統計量的標準誤。一般研究者對巨集語法的撰寫並不熟悉，就算熟悉該語法，每次分析都要撰寫巨

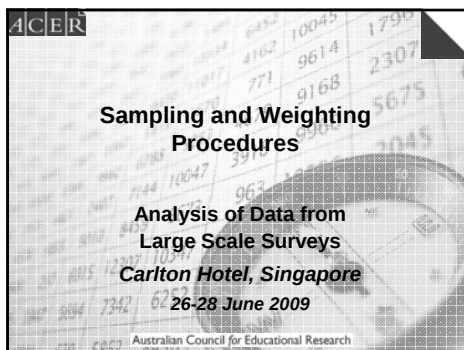
集程式也是很方便。有了此一整合在 SPSS 當中的外掛程式，在分析上就極為便利。

雖然 Replicates 很方便，但是它主要是針對 PISA 的資料而設計的，內定的變數名稱的命名規則是 PISA 資料庫的命名規則。如果要以 Replicates 來分析 TIMSS 的資料，則需要修改它所產生的巨集程式檔。不過，IEA 也發展了一套類似的軟體，稱為 IDB Analyzer。這兩套軟體可在 ACER 和 IEA 的官方網站上直接下載。比較兩套軟體，Replicates 較有彈性，考慮的分析項目較為周全。

綜觀國內研究，研究者進行大型調查資料二次分析時，一般皆無考慮測量誤差和抽樣誤差，統計量的標準誤都被低估了。在此情況下，統計分析的計算結果雖然顯示有顯著差異，但由於真實的標準誤可能很大，因此其實並沒有顯著。學者目前的作法或許是因為手邊沒有方便的統計分析套件所致，此工作坊所介紹的 Replicates 若能引介給國內學者，相信對國內二次分析此研究領域的發展必有極大的助益。

三、 攜回資料名稱及內容

1. 各場次講者提供之簡報檔及習作資料。
2. Replicates (V 6.2.1) 程式軟體一套。



Topics of this presentation

- Sampling in large-scale data surveys
- Computation of sampling weights
- How to apply replication methods

Sample versus Census - 1

- Surveys usually draw a sample because a census:
 - Requires unavailable budgets
 - Time consuming and thus incompatible with deadlines
 - Does not necessarily bring additional and required information

Sample versus Census - 2

- Several ways of drawing a sample
- Two major principles:
 - Avoiding bias in the selection
 - Randomness
 - Sampling frame covers the population adequately
 - Ensuring the varying probabilities of selection are reflected $\Rightarrow$ weights
 - Achieving maximum precision, given a set of constraints (time, funds, and so on)

Selecting a simple random sample

- An SRS gives an equal probability of selection for each n selected from population N which is

$$p_i = \frac{n}{N}$$

- and each selected n has a weight of

$$w_i = \frac{1}{p_i} = \frac{N}{n} = \frac{400}{40} = 10$$

Using weights with SRS

- Since the weights are the same for all selected members of the sample, weighted and unweighted means are identical:

$$\hat{\mu}_{(X)} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \frac{w_i \sum_{i=1}^n x_i}{w_i \sum_{i=1}^n 1} = \frac{\sum_{i=1}^n x_i}{n}$$

Using weights with SRS

- The sum of the weights always rebuilds the size of the population

$$\sum_{i=1}^n w_i = \sum_{i=1}^n \frac{N}{n} = N$$

Student surveys

- Simple random sampling generally too expensive and often impossible.
 - Often no central database for sampling of individual students
 - Use of SRS is impractical for survey administration.
 - No link of students to classrooms or schools available.
- Therefore, two-stage designs common, where schools are selected at first stage and students or intact classrooms at the second.

Two stage sample

- n schools are selected from a population of N schools
 - $\Rightarrow$ what is the probability of a particular school to be drawn?
- Within school i , n_i students or n_i classrooms are drawn from a population of N_i students or N_i classroom
 - $\Rightarrow$ what is the probability of a particular student in school i to be drawn?

An example of a two stage sample

- Let us imagine:
 - a population of 10 schools
 - Within any school, there are 40 students
 - So the educational system counts 400 students
- 4 schools are drawn according to a SRS procedure
- 10 students within the 4 selected schools are selected according to a SRS

Two-stage probabilities

- Four schools selected out of 10

$$p_{1_i} = \frac{n_{sc}}{N_{sc}} = \frac{4}{10} = 0.4$$

- 10 students selected within each school

$$p_{2_{ij}} = \frac{n_i}{N_i} = \frac{10}{40} = 0.25$$

Final weights (SRS two-stage)

- The final selection probability is

$$p_{ij} = p_{1_i} p_{2_{ij}} = \frac{n_{sc} n_i}{N_{sc} N_i}$$

- and the final total weight is the product of school and student weight or the inverse of the selection probability:

$$w_{ij} = \frac{1}{p_{ij}} = \frac{1}{0.1} = 10$$

Weights summary

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	40							
2	40	0.4	2.5	0.25	4	0.1	10	100
3	40							
4	40							
5	40	0.4	2.5	0.25	4	0.1	10	100
6	40							
7	40	0.4	2.5	0.25	4	0.1	10	100
8	40							
9	40							
10	40	0.4	2.5	0.25	4	0.1	10	100
Total			10					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10							
2	15	0.4	2.5	0.66	1.5	0.27	3.75	37.5
3	20							
4	25							
5	30	0.4	2.5	0.33	3	0.13	7.5	75
6	35							
7	40	0.4	2.5	0.25	4	0.1	10	100
8	45							
9	80							
10	100	0.4	2.5	0.1	10	0.04	25	250
Total	400		10					462.5

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10	0.4	2.5	1	1	0.4	4	40
2	15	0.4	2.5	0.66	1.5	0.27	3.75	37.5
3	20	0.4	2.5	0.5	2	0.2	5	50
4	25	0.4	2.5	0.4	2.5	0.16	6.25	62.5
Total			10					190

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
7	40	0.4	2.5	0.250	4	0.10	10.00	100.0
8	45	0.4	2.5	0.222	4.5	0.08	11.25	112.5
9	80	0.4	2.5	0.125	8	0.05	20.00	200.0
10	100	0.4	2.5	0.100	10	0.04	25.00	250.0
Total			10					662.5

From SRS to PPS Sampling

- As school sizes differ across schools, then:
 - student weights will vary across schools.
 - In smaller schools, probability of selection is higher.
 - In larger schools, probability of selection is lower.
 - Very large uncertainty on the student population estimates $\Rightarrow$ population estimates will have a large sampling variance

Sampling with school probabilities proportional to size (PPS)

- Larger schools get a higher probability of selection than smaller schools...
- ...but students in larger schools have a smaller within-school probability of being selected than students in small schools, and...
- the school and within school weights differ between schools, but final student weights do not vary ('self-weighted sample').

School weights with PPS sampling

- Probability of a school to be selected is equal to the ratio of the school size multiplied by the number of schools to be sampled and divided by the total number of students in the population:

$$p_{1_i} = \frac{N_i * n_{sc}}{N}$$

PPS sample of schools

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10							
2	15							
3	20	0.2	5.00	0.500	2.0	0.1	10	100
4	25							
5	30							
6	35							
7	40	0.4	2.50	0.250	4.0	0.1	10	100
8	45							
9	80	0.8	1.25	0.125	8.0	0.1	10	100
10	100	1	1.00	0.100	10.0	0.1	10	100
Total	400		9.75					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
1	10	0.1	10	1	1	0.1	10	100
2	15	0.15	6.7	0.67	1.5	0.1	10	100
3	20	0.2	5	0.5	2	0.1	10	100
4	25	0.25	4	0.4	2.5	0.1	10	100
Total			25.67					400

Different school sizes

School label	School size	School prob.	School weight	Within school prob.	Within school weight	Final student prob.	Final student weight	Sum of final weights
7	40	0.4	2.5	0.25	4	0.1	10	100
8	45	0.45	2.2	0.2	4.5	0.1	10	100
9	80	0.8	1.25	0.125	8	0.1	10	100
10	100	1	1	0.1	10	0.1	10	100
Total			6.97					400

PPS sample of schools

- With a PPS sample of school:
 - Self weighted sample of students
 - Rebuild the student population
- But
 - Large variability of the school population estimate
- ⇒ Use of a systematic procedure

PSS Sampling example (sampling interval 100, start 75.2)

School label	School size	From student number	To student number	Part of the sample
1	10	1	10	No
2	15	11	25	No
3	20	26	45	No
4	25	46	70	No
5	30	71	100	Yes
6	35	101	135	No
7	40	136	175	No
8	45	176	220	Yes
9	80	221	300	Yes
10	100	301	400	Yes

PPS Sampling - summary

- More efficient way of sampling
 - Reduction of schools
 - Reduction of variation in weights
- Weights may still differ because of ...
 - Over- or under-sampling of some strata in the population
 - Lack of accuracy in measures used for selection of sample
 - Adjustments to weights for school and student non-response

Possible weight components

- School base weight
- School weight trimming factor (to avoid extremely large or small school weights)
- Student or Classroom base weight
- School non-response adjustment
- Student non-response adjustment
- Final student or classroom weight trimming factor (to avoid final student weights more than four times the median)

School-level weights

- The **School base weight** (for PPS samples) is the population size over the number of sampled schools multiplied with the school size:

$$BW_{sc}^i = \frac{N_{pop}}{n_{sc} * N_{non}}$$

- The **school non-response adjustment** is the number of sampled schools over the number of participating schools (in PISA: schools weighted with size):

$$Sc_{adj} = \frac{n_{sc}}{n_p}$$

Student-level weights

- The **Student base weight** is the population size over the number of sampled students multiplied with the school size:

$$BW_{st}^i = \frac{N_{st}^i}{n_{st}^i}$$

- The **student-level non-response adjustment** is the number of sampled (and eligible) students within a school over the number of participating students:

$$St_{adj} = \frac{n_{st}^i}{n_p^i}$$

Final sampling weight

- The final student weight (in simple designs) is equal to multiplying school and student base weights and the non-response adjustments:

$$W_i = BW_{sc}^i * Sc_{adj} * BW_{st}^i * St_{adj}$$

- Final student weights are often (but not always) the same within schools!
- Population size can be estimated by summing students weights within and across schools

Summary of Weighting

- One-level SRS sampling usually not appropriate for student surveys.
- PPS sampling more efficient than Two-level SRS.
- Weights do vary across students/schools.
- Generally, analysis of large-scale survey data should use sampling weights!!!

Questions?

Practice Session - weighting

- Generated data with 998 schools and 99,479 students in population
- Variables are
 - SCORE: Achievement score with mean of 500 and SD of 100
 - GENDER of students
 - SES: Socio-economic status index with mean of 0 and SD of 1
 - PROG: General vs. vocational study program (1, 0)

Sampling design

- PPS sample
- Explicit stratification: PROG
- Implicit stratification: High/low SES schools
- 50 schools per explicit stratum
- 30 students randomly selected per school
- Schools sorted by enrolment

Your task

- Please compute the sampling weights for this sample
 - WGT_EXMPL_DATA.SAV (96 schools)
- Student-level weights can be computed at school level
- Only participating schools in data!

Variables in datafile

- POPSIZE: Population size of each explicit stratum
- NSCHOOL: Number of schools in sample
- ENROL: Number of enrolled students at each school
- STUDSMP: Student sample size within schools
- SPAR: Number of participating students within schools

Weights (within explicit strata)

- School base weight:
 - POPSIZE / (50 * ENROL)
- School non-response adjustment
 - 50 / PSCHOOL
- Student base weight
 - ENROL / STUDSMP
- Student non-response adjustment
 - STUDSMP / SPAR
- Final student weights
 - Product of all four weights

```

* VIEW WEIGHTS with SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Tables Window Help

COMPUTE SCL_BW = FREQ100 / (50 * ENROL) .
COMPUTE SCL_AP = 50 / FREQ100 .
COMPUTE STD_BW = ENROL / STDENR .
COMPUTE STD_AP = STDENR / SPAR .
COMPUTE TOTAL_W = SCL_BW * SCL_AP * STD_BW * STD_AP .
EXECUTE .

WEIGHT
BY STD_W .
FREQUENCIES
VARIABLES=PROG
/ORDER= ANALYSIS .

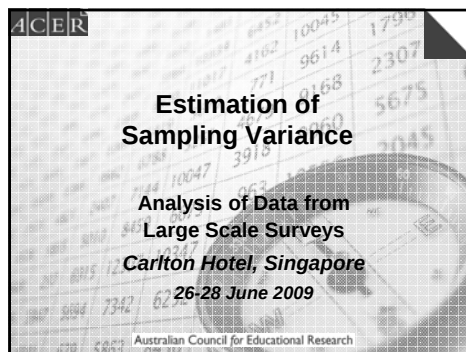
```

Check of results

- Compute SUMWGT
 - Product of Final weight and participating students (SPAR)
- Weight by SUMWGT
 - Frequency of PROG
 - Compare with table:

PROG

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	42022	42.2	42.2	42.2
	1.00	57457	57.8	57.8	100.0
Total		99479	100.0	100.0	



Topics of this presentation

- Error components in measurement
- Implications of cluster sampling for the estimation of sampling variance
- Methods to estimate sampling variance with cluster sample data

Error components associated with measurement

- Sampling error
 - Random (-> estimation of sampling error)
 - Systematic (e.g. non-response bias)
- Measurement error
 - Random (-> reliability, multiple imputation)
 - Systematic (e.g. unfavourable test conditions)

Simple Random Sample (SRS)

- A sample is random if all the individuals in the population have a known probability to be selected
- The sample will be simple if the individuals are all selected independently and every individual has the same chance of being selected

Sampling variance

- In the case of a SRS, the sampling variance is:
 - proportional to the variance
 - inversely proportional to the sample size
- The standard error can be obtained from the variance in the population and the sample size
- However, the variance in the population is (typically) not known

Calculation of Standard Errors

- As we don't know the variance in the population, the SE for a mean of θ as obtained from a sample is calculated as

$$SE_{\bar{\mu}} = \frac{\hat{\sigma}_{\bar{\mu}}}{\sqrt{n}}$$

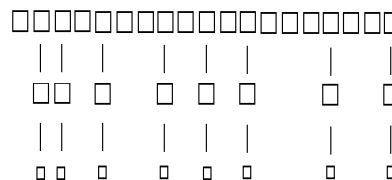
- similarly, the SE for a percentage P is calculated as

$$SE_P = \sqrt{\frac{PQ}{n}}$$

Drawing a sample of students

- As already mentioned, for practical, conceptual and financial reasons, SRS is not appropriate
- Cluster sampling provides a more adequate solution:
 - At the first stage schools are drawn from a sampling frame,
 - then classes and or students are selected randomly within school

Cluster sampling (1)



Cluster sampling (2)

- The variance in a cluster sample consists of *Between-school* and *Within-school* variance
- In an infinite population of clusters and an infinite population within clusters, and if the sample sizes m are equal within l sampled clusters, the variance is:

$$V(\bar{y}) = \frac{\sigma^2_{\text{between}}}{l} + \frac{\sigma^2_{\text{within}}}{l \cdot m}$$

Cluster sampling (3)

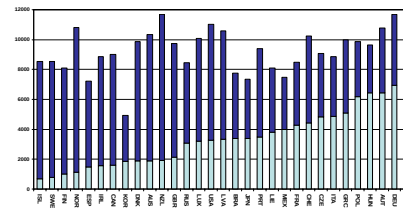
Comparison with a SRS:

- For example, 150 schools with 35 students per schools. Expected STD= 100

$$\sigma^2_{(u)} = \frac{4000}{150} + \frac{6000}{150 * 35} = 26.667 + 1.143 = 27.8$$

$$\sigma^2_{(u)} = \frac{0}{150} + \frac{10000}{5250} = 1.90$$

School and Within school variance (Reading Pisa 2000)



How to estimate sampling variance for complex designs?

- Design so complex that there are no formulae, even for a mean
 - Multilevel analysis (tends to overestimate the sampling variance for stratified samples)
 - Taylor series linearization (requires data on stratification, e.g. STATA)
 - Re-sampling methods
 - Bootstrap methods.
 - Replication methods (e.g. WESVAR, AM but can also be programmed in SPSS or SAS).

Types of replication methods

- Jackknifing
 - JK1 (only recommended with no information on sampling design)
 - JK2 (TIMSS 1995, 1999, 2003, PIRLS, CivEd 1999)
- Balanced Repeated Replication (BRR)
 - BRR
 - Fay's Method (applied in PISA)

Jackknife for unstratified sample

Replicate	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
School 1	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 2	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 3	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11
School 4	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11
School 5	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11
School 6	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11
School 7	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11
School 8	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11
School 9	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11
School 10	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00

Jackknife for stratified sample

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1		2	1	1	1	1	1	1	1	1
1	2		0	1	1	1	1	1	1	1	1
2	3		1	0	1	1	1	1	1	1	1
2	4		1	2	1	1	1	1	1	1	1
3	5		1	1	2	1	1	1	1	1	1
3	6		1	1	0	1	1	1	1	1	1
4	7		1	1	1	0	1	1	1	1	1
4	8		1	1	1	2	1	1	1	1	1
5	9		1	1	1	1	2	1	1	1	1
5	10		1	1	1	1	0	1	1	1	1
6	11		1	1	1	1	1	2	1	1	1
6	12		1	1	1	1	1	0	1	1	1
7	13		1	1	1	1	1	1	0	1	1
7	14		1	1	1	1	1	1	2	1	1
8	15		1	1	1	1	1	1	1	0	1
8	16		1	1	1	1	1	1	1	2	1
9	17		1	1	1	1	1	1	1	1	0
9	18		1	1	1	1	1	1	1	2	1
10	19		1	1	1	1	1	1	1	1	2
10	20		1	1	1	1	1	1	1	1	0

Balanced Repeated Replication (BRR)

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1	2	0	0	2	0	0	0	2	2	0
1	2	0	2	0	2	2	0	0	0	0	2
2	3	2	2	0	0	2	0	0	0	2	2
2	4	0	0	2	2	0	2	2	2	0	0
3	5	2	0	2	0	0	2	0	0	2	2
3	6	0	2	0	2	2	0	2	2	0	0
4	7	2	2	0	2	0	2	0	0	2	2
4	8	0	0	2	0	2	2	0	2	2	0
5	9	2	2	2	0	0	2	0	0	0	2
5	10	0	0	0	2	0	2	2	0	2	2
6	11	2	2	2	2	0	2	0	0	2	0
6	12	0	0	0	0	2	2	0	2	2	2
7	13	2	0	2	2	2	0	2	0	0	2
7	14	0	2	0	0	2	0	2	2	0	2
8	15	2	0	0	2	2	0	2	0	0	2
8	16	0	2	2	0	0	2	0	2	2	0
9	17	2	0	0	0	2	2	0	2	0	2
9	18	0	2	2	2	0	0	0	2	0	2
10	19	2	2	0	0	0	2	2	2	0	0
10	20	0	0	2	2	2	0	0	2	0	2

Fay's variant of BRR

Pseudo- rep (r)	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5	0.5	1.5
1	2	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5	1.5	0.5
2	3	1.5	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5	0.5
2	4	0.5	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5	1.5
3	5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	1.5	1.5
3	6	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	0.5	0.5
4	7	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5	1.5	1.5
4	8	0.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5	0.5	0.5
5	9	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5	1.5
5	10	0.5	0.5	0.5	1.5	1.5	1.5	1.5	1.5	1.5	0.5
6	11	1.5	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5	0.5
6	12	0.5	0.5	0.5	0.5	1.5	1.5	1.5	1.5	1.5	1.5
7	13	1.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	1.5	0.5
7	14	0.5	1.5	0.5	0.5	1.5	0.5	1.5	1.5	1.5	1.5
8	15	1.5	0.5	0.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5
8	16	0.5	1.5	1.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5
9	17	1.5	0.5	0.5	0.5	1.5	1.5	0.5	1.5	0.5	1.5
9	18	0.5	1.5	1.5	1.5	0.5	0.5	1.5	1.5	1.5	0.5
10	19	1.5	1.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5
10	20	0.5	0.5	1.5	1.5	1.5	0.5	0.5	1.5	1.5	1.5

Replication Theory

- Variance estimators generally take the form of

$$Var_{rep}(\theta) = c \sum_{k=1}^G [\theta_k - \bar{\theta}]^2$$
- θ is the parameter of interest
- G is the total number of replicates with k denoting the k -th replicate
- c is a constant that depends on the method
 - For JK1, $c = (G-1)/G$
 - For JK2, $c = 1$
 - for BRR, $c = 1/G$
 - for Fay's BRR, $c = 1/(G(1-k^2))$

Computing sampling variance in PISA

- In Pisa
 - Fay's method
 - 80 replicate weights in data file
 - Perturbation factor k is 0.5
 - $c = 1/(80(1-0.5)^2) = 0.05$

$$\sigma_{(\hat{\theta})}^2 = \frac{1}{G(1-k)^2} \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{80(1-0.5)^2} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{20} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2$$

- To compute sampling errors, one needs to estimate once with the full PISA weight and 80 times with each replicate weight!

Computing sampling variance in CIVED, TIMSS or PIRLS

- In IEA studies
 - Jackknife (JK2) method
 - 75 replicate weights in data file
 - $c = 1$
- To compute sampling errors, one needs to estimate once with the full PISA weight and 75 times with each replicate weight!

How can we achieve this?

- Special software
 - WESVAR (Westat)
 - AM (AIR)
- Macros (requires syntax handling)
 - SPSS
 - SAS
- IEA Database Analyser
- SPSS Analysis module
 - Based on SPSS macros but menu-driven

Questions?

Example for practice - 1

- Compute Standard errors based on example with Jackknife (JK2) replication (generated data)
- Please open SPSS file
 - REPLICATES_EXAMPLE.sav
- Variables in File
 - Mean of achievement scores for each replicate weights (SCORE)
 - Mean of achievement scores with full sample weight (SCORE_FW)

Example for practice - 2

- Compute squared difference between scores with replicate weights and full sample weight
- Compute sum and multiply with correct constant (for JK2 = 1)

ACER

Scaling methodology

Analysis of Data from
Large Scale Surveys
Carlton Hotel, Singapore
26-28 June 2009

Australian Council for Educational Research

Rationale for IRT scaling of data

- Summarising data instead of dealing with many single items
- Raw scores or percent correct sample-dependent
- Makes equating possible and can deal with rotated test forms

The 'Rasch model'

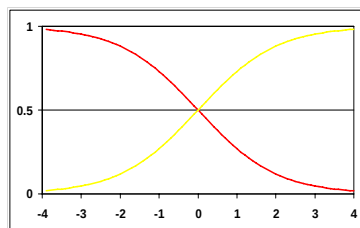
- Models the probability to respond correctly to an item as

$$P_i(X_{ni} = 1) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

- Likewise, the probability of NOT responding correctly is modelled as

$$P(X_{ni} = 0) = \frac{1}{1 + \exp(\theta_n - \delta_i)}$$

IRT curves



Some examples

Student ability	Item Difficulty	Probability of success
-2	-2	0.50
-1	-1	0.50
0	0	0.50
1	1	0.50
2	2	0.50

Some examples

Student ability	Item Difficulty	Probability of success
-2	-3	0.73
-1	-2	0.73
0	-1	0.73
1	0	0.73
2	1	0.73

Student ability	Item Difficulty	Probability of success
-2	-1	0.27
-1	0	0.27
0	1	0.27
1	2	0.27
2	3	0.27

Some examples

Student ability	Item Difficulty	Probability of success
-2	-4	0.88
-1	-3	0.88
0	-2	0.88
1	-1	0.88
2	0	0.88

Student ability	Item Difficulty	Probability of success
-2	0	0.12
-1	1	0.12
0	2	0.12
1	3	0.12
2	4	0.12

Anchoring of tests

- The Rasch model allows to derive comparable scores for two different assessment via the inclusion of "anchor" items.
- Anchoring can be used across time (e.g. repeated surveys) but also across different population assessments (e.g. grades)

Example of anchoring between grades 4 and 8

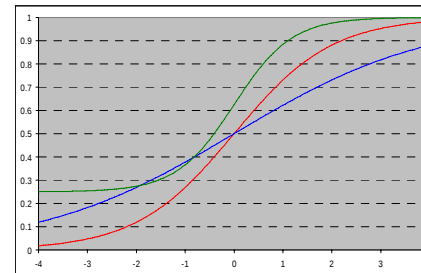


3 Parameter model

- Three-parameter models have additional parameters for item discrimination and guessing:

$$P(X_{ij} = 1 | \theta_i, a_i, b_i, c_i) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta_i - b_i))}$$

Comparisons of the ICC curves

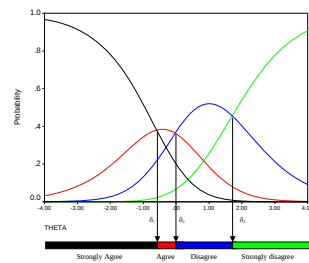


IRT models for polytomous items

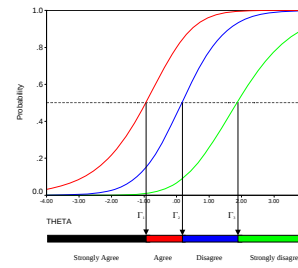
- Extension of Rasch model with additional step parameters τ
- Rating scale model* has equal step parameters across items
- Partial credit model* has different step parameters for each item

$$P_{x_i}(\theta) = \frac{\exp \sum_{j=0}^{x_i} (\theta_n - (\delta_i + \tau_{ij}))}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^h (\theta_n - (\delta_i + \tau_{ij}))} \quad x_i = 0, 1, 2, \dots, m_i$$

Partial Credit Model – probabilities



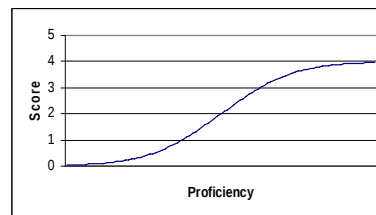
Partial Credit Model – threshold curves



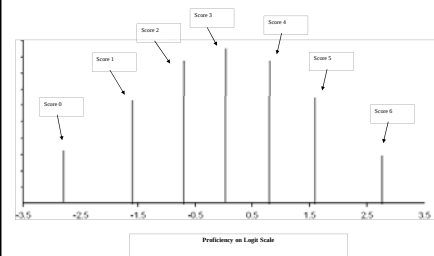
How might we impute a reasonable proficiency value?

- Choose the *proficiency* that makes the score most likely
 - Maximum Likelihood
 - Weighted Likelihood
- Choose the most likely *proficiency* for the score
 - empirical Bayes
- Choose a selection of likely proficiencies for the score
 - Multiple imputations (plausible values)

Maximum Likelihood vs. Raw Score



The Resulting Proficiency Distribution



Characteristics of MLE

- Unbiased at individual level with sufficient information BUT biased towards ends of ability scale.
- Arbitrary treatment of perfects and zeroes required
- Discrete scale & measurement error leads to bias in population parameter estimates

Weighted Likelihood Estimation

- WLE maximizes the weighted likelihood, wL , where w is the weight function.
- Solve for each case

$$\sum_{i=1}^k \left[r_i + \frac{J}{2I} \right] - \sum_{j=1}^k \frac{\exp(\theta_a - \delta_j)}{1 + \exp(\theta_a - \delta_j)} = 0$$

- $\frac{-J}{2I^2}$ being the bias for the ML estimates
- with I being the Information function: $\sum_{i=1}^k \frac{p_{vi}'^2}{p_{vi}(1-p_{vi})}$
- and $J = \sum_{i=1}^k \frac{p_{vi}' p_{vi}''}{p_{vi}(1-p_{vi})}$

Characteristics of WLE

- Less biased than MLE
- Provides estimates for perfect and zero scores
- BUT discrete scale & measurement error leads to bias in population parameter estimates**

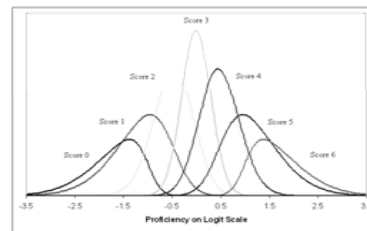
Plausible Values

- What are plausible values?
- Why do we use them?
- How to analyse plausible values?

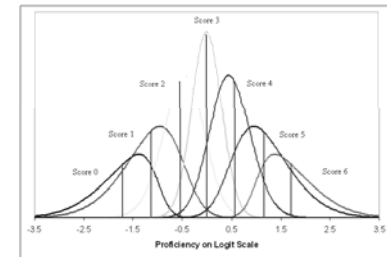
Purpose of educational tests

- Measure particular students**
(minimise measurement error of individual estimates)
- Assess populations**
(minimise error when generalising to the population)

Posterior distributions for test scores on 6 dichotomous items



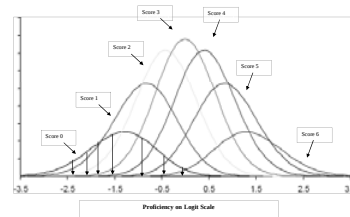
Empirical Bayes - EAPs



Characteristics of EAPs

- Biased at the individual level but unbiased population means (NOT variances)
- Discrete scale, bias & measurement error leads to bias in population parameter estimates
- Requires assumptions about the distribution of proficiency in the population

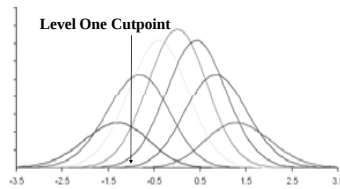
Plausible Values



Characteristics of Plausible Values

- Not *fair* at the student level
- Produces unbiased population parameter estimates
 - if assumptions of scaling are reasonable
- Requires assumptions about the distribution of proficiency

Estimating percentages below benchmark with Plausible Values



The proportion of plausible values less than the cut-point will be a superior estimator to the EAP, MLE or WLE based values

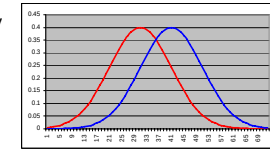
Methodology of PVs

- Mathematically computing posterior distributions around test scores
- Drawing 5 random values for each assessed individual from the posterior distribution for that individual

What is conditioning?

- Assuming normal posterior distribution: $N(\mu, \sigma^2)$
- Model sub-populations: $N(\mu + \alpha X, \sigma^2)$

X=0 for boy
X=1 for girl



$$N(\mu + \alpha X + \beta Y + \gamma Z + \dots, \sigma^2)$$

Multi-dimensional scaling

- 4 scales for major domain => 4 abilities for each person, $\theta_1, \theta_2, \theta_3$ and θ_4
- Covariance between the abilities can be estimated.
- It is assumed that only one ability is drawn upon for answering a particular item.

Conditioning Variables

- Plausible values should only be analysed with data that were included in the conditioning (otherwise, results may be biased).
- Aim: Maximise information included in the conditioning, that is use as many variables as possible.
- To reduce number of conditioning variables, factor scores from principal component analysis can be used.
- Use school means of performance takes between-school variation into account (no inclusion of school questionnaire data needed)

Plausible values

- Model with conditioning variables will improve precision of prediction of ability (population estimates ONLY)
- Multidimensional models improves estimates, through using information about correlation between dimensions.
- Conditioning provides unbiased estimates for modelled parameters.
- For tests with more than one dimension multi-dimensional scaling is recommended.

Results from analysis with simulated data

- Simulation used to test bias in WLE, EAP and PV estimates
- Generating item responses
 - True abilities are known
 - True relationships between abilities and background variables are known
 - Within- and between school variances are known
- Simulated item responses were used to estimate WLEs, EAPs and PVs with ConQuest

Outcomes of simulation studies - 1

- MEAN:
WLE, EAP and PV unbiased
- SD and PERCENTILES:
WLE overestimated
EAP underestimated
PV unbiased

Outcomes of simulation studies - 2

- CORRELATION COEFFICIENT:
WLE underestimated
EAP-C overestimated
PV-C unbiased
- REGRESSION COEFFICIENT:
WLE, EAP-C, PV-C unbiased
- But RESIDUAL VARIANCE:
WLE overestimated
EAP-C underestimated
PV-C unbiased

Calculating of measurement error

- In PISA, TIMSS or PIRLS data files, there are five plausible values for each cognitive domain
- Using five plausible values enable researchers to calculate estimates of the measurement error

How to analyse PVs - 1

- Estimated mean is the AVERAGE of the mean for each PV

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M \hat{\mu}_i$$

- Sampling variance is the AVERAGE of the sampling variance for each PV

$$\hat{\sigma}_{(\hat{\mu})}^2 = \frac{1}{M} \sum_{i=1}^M \hat{\sigma}_{(\hat{\mu}_i)}^2$$

How to analyse PVs - 2

- Measurement variance computed as:

$$\hat{\sigma}_{(PV)}^2 = \frac{1}{M-1} \sum_{i=1}^M (\hat{\mu}_i - \hat{\mu})^2$$

- Total standard error computed from measurement and sampling variance as:

$$\hat{\sigma}_{(\hat{\mu}_{PV})} = \sqrt{\sigma_{(\mu)}^2 + (1 + \frac{1}{M})\sigma_{(PV)}^2}$$

How to analyse PVs - 3

- μ can be replaced by any statistic for instance:
 - SD
 - Percentile
 - Correlation coefficient
 - Regression coefficient
 - R-square
 - etc.

Steps for estimating both sampling and measurement error

- Compute statistic for each PV for fully weighted sample
- Compute statistics for each PV for 80 replicate samples
- Compute sampling error (based on previous steps)
- Compute measurement error
- Combine error variances to calculate standard error

Questions?

Practice 1

- Compute combined standard error for Science achievement in Australia (PISA 2006)
 - [Example_PVs.xls](#)

Steps

- Compute overall results as average of five plausible values
- Compute overall sampling variance from average squared SEs for each PV
- Compute measurement variance
 - Sum of squared deviations from average times $1/(M-1) = 0.25$
- Compute combined standard error
 - Square root of overall sampling variance + measurement variance times 1.2

Practice 2

- Reproducing parts of PISA 2003 report, table 2.5c, page 351 (Maths performance)
 - Means (PV)
 - Standard deviation (PV)
 - Mean by male/female (PV)
 - Difference male/female (PV)
 - Percentiles (PV)

Practice 3

- Reproducing TIMSS 2003 report, page 16, exhibit 1.1 (Maths performance)
 - Means (PV)
 - Mean by male/female (PV)
 - Difference by male/female (PV)

Practice 4

- Reproducing PISA 2003 table 2.5a, page 349 (Maths proficiency levels)
 - percentages for each proficiency level (PV)

Practice 5

- Reproducing PISA 2003 report, table 4.2e, page 387
 - Recode FAMSTRUC into single-parent vs. others
 - Group percentages
 - Maths performance by groups
 - Difference by groups

PISA reporting of achievement by quartiles of questionnaire indices

- In PISA reports achievement means by quartiles of questionnaire indices
- To ensure equal distribution random error is added to questionnaire index scores
- No exact replication possible
- SPSS module provides tool for producing results

Practice 6

- Reproducing parts of PISA 2003 international report, table 3.7, page 368 (variable MATHEFF, MATHS)
 - Means
 - Performance by quartile

Effect Sizes

- Instead of comparing absolute differences researcher may want to report standardised differences:

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

- SPSS replicates enables researchers to estimate standard errors for single variables and plausible values

Practice 7

- Use one PISA 2006 country data set (e.g. Hong Kong data) to compute effect sizes plus standard errors for
 - Gender (make 0, 1 variable)
 - Reading

Relative Risk - 1

- Measure of association between an antecedent factor and an outcome factor based on 4X4 table
- Calculation requires dummy variables (coded 0 and 1).
- For example, the outcome variable might be reaching a performance benchmark (yes=1) and the antecedent a student's gender (female=1)

Relative Risk Ratio - 2

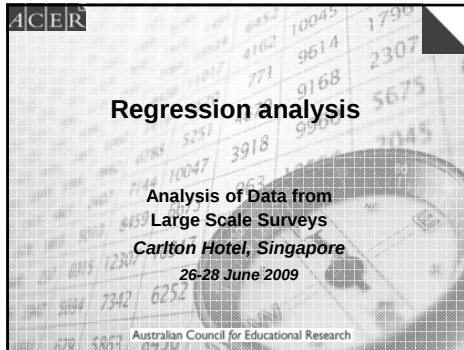
		Outcome measure		Total
		Yes	No	
Antecedent measure	Yes	p_{11}	p_{12}	$p_{1.}$
	No	p_{21}	p_{22}	$p_{2.}$
Total		$p_{.1}$	$p_{.2}$	p

$$RR = \frac{(p_{11} / p_{1.})}{(p_{21} / p_{2.})}$$

- The resulting value how much more likely a respondent would fall into the category given the antecedent

Practice 8

- Use one PISA 2006 country data set (e.g. Hong Kong data) and create dummy variables for each plausible value (Reading) for the proficiency level 3 and above ($>$ value 480.2)
- Create five gender dummies (G1 to G5) with male as 1 and female as 0
- Save dataset and use SPSS Replicates Module



Regression is...

- Explaining the distribution of a dependent variable Y with...
- the distribution of one or more independent variables $X_1, X_2, \dots, X_n$

Types of regression models

- Linear regression
- Non-linear regression
- Logistical regression
- Time Series Regression
- Multi-level Regression (linear and logistical)

Regression can be...

- Bivariate with just one predictor
- Multiple with various predictors
- Path analysis, when there is more than one dependent variables
- Non-recursive, when two variables are both dependent and independent with respect to each other...

The linear OLS (Ordinary Least Squares) regression model

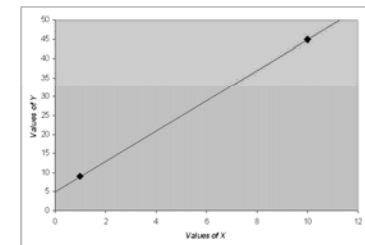
- Regression equation

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- Error term

$$\varepsilon_i \sim N(0, \sigma^2)$$

Graphical display of regression equation



Regression equation

$$\hat{Y} = 5 + 4X$$

- This equation can be graphically represented by a straight line
- A equation with 2 variables (X, Y) can be represented in a two-dimensional space

Linear regression is based on:

- Means: $\bar{X} = \frac{\sum X}{N}$
- Variances: $\text{var}(x) = \frac{\sum X - \bar{X}}{N - 1}$
- Covariances: $\text{cov}(x, y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1}$

Regression Parameters

- Coefficient: $b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$
- Intercept: $a = \bar{Y} - b\bar{X}$
- Predicted Y: $\hat{Y} = a + bX$
- Residual: $e = \hat{Y} - Y$

Partitioning of Variance

Y consists of $Y = \bar{Y} + (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$

Var(Y) is $\sum y^2 = ss_{reg} + ss_{res}$

Explained variance: $r_{xy}^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}$

Unexplained variance: $ss_{res} = (1 - r_{xy}^2) s_y^2$

Tests of significance (simple random samples)

- F-Test for regression $F = \frac{ss_{reg}/k}{ss_{res}/N - k - 1}$
- t-test for parameters $t = \frac{b}{s_b}$
- SE for parameter $s_b = \sqrt{\frac{ss_{res}/N - k - 1}{\sum x^2}}$

Standardised Coefficient

- Regression coefficients can be standardised as: $\beta^* = b \frac{s_x}{s_y}$
- In bivariate regression this is equal to $r_{xy} = \frac{\text{cov}(X, Y)}{s_x s_y}$
- Or a regression between standardised variables $z_y = \beta^* z_x$

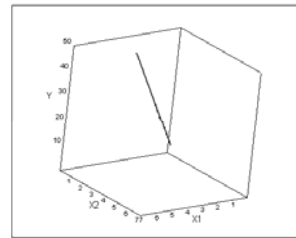
Multiple Regression

- Equation becomes $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + e_i$
- Partial coefficients with standardised scores can be simply calculated as

$$\beta_1^* = \frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2}$$

- and $\beta_2^* = \frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2}$

Equation with two predictors: $Y = 5 + 2X_1 + 3X_2$



How to judge whether a model is good?

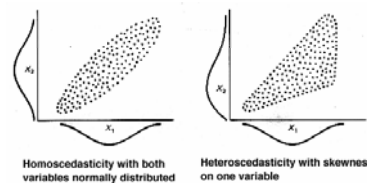
- Explained variance proportion as measures by R Square.
- Size of regression coefficients.
- Significance tests (F-test for model, t-tests for parameters).
- Inclusion of all relevant variables (Theory!).
- Is method appropriate?

Assumptions of linear regression

- Linearity of relationship
- Continuous nature of (at least) the dependent variable
- Normality of distributions
- Homoscedasticity
- Absence of multicollinearity (independent variables not being highly correlated with each other)

Heteroscedasticity

- Heteroscedasticity is caused by non-normality of one of the variables, an indirect relationship between variables, or to the effect of a data transformation.



Analysis of residuals

- Usually done with standardised residuals $z = \frac{Y - \hat{Y}}{s_{y.x}}$
- Detect outliers (extreme values can distort estimates)
- Check random and even distribution

Multicollinearity - 1

- Means that one or more independent variables are linear combination of other independent variables
- This can distort the estimation of coefficients
- Can lead to large standard errors for the coefficients

Multicollinearity - 2

- No 'scientific' rules
- Check correlation between independent variables
- Use multicollinearity statistics (VIF, Tolerance)
- Deletion of variable may lead to misspecification of models
- Sometimes variables can be combined (for examples Mother's and Father's education)

Use of Dummy Variables

- Dichotomous or Categorical variables can be recoded into dummies with (0,1).
- In a bivariate case this leads to the intercept=mean(0), slope=mean(1).
- Sometimes interaction effects can be modelled (different slopes for different categories).

Use of standardised coefficients

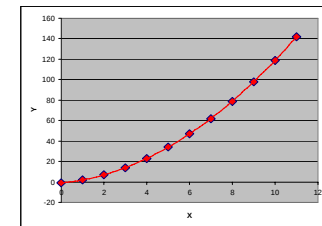
- Often thought to be 'easier' to interpret.
- Standardisation depends on variances of independent variables.
- Unstandardised coefficient are generally preferable (can be translated directly).
- Standardised coefficients cannot always be compared across populations.

Curvilinear relationships

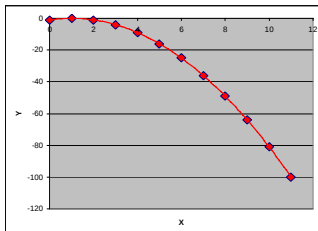
- Often relationships are not linear, but curvilinear (e.g. homework & performance)
- This can be modelled by adding a quadratic term to the equation:

$$Y_i = \alpha + \beta_1 X + \beta_2 X^2 + e_i$$

$$Y = -1 + 2X + 1X^2$$



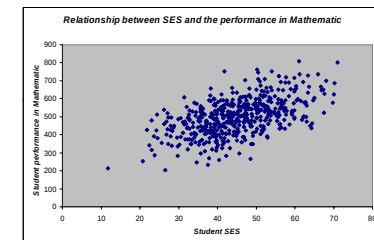
$$Y = -1 + 2X - 1X^2$$



When does linear regression modelling become inappropriate?

- When the dependent variable is dichotomous or polytomous (>> Logistical Regression).
- When data are sequential over time and variables are 'auto correlated' (>> Time Series Analysis).
- When context effects need to be analysed and slopes are different across higher level units (>> Multi-level Analysis).

Example: SES and performance

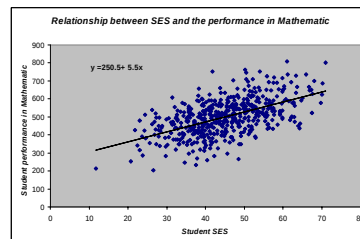


SES and performance

$$Y_i = \alpha + \beta X_i$$

- $Y_i = 250.5 + 5.5(\text{SES})$

SES and Maths performance



Observed, predicted and residual scores

Students	SES	Observed score	Predicted Score	Residual
1	49	463	520	-57
2	53	384	542	-158
3	51	579	531	+48
4	42	404	481.5	-77.5
5	42	282	481.5	-199.5

Questions?

Practice 1

- Estimate regression model for GENDATA_sample.sav
 - Predictors: Gender, SES, PROG
- First using normal syntax
- Then with SPSS module to compute correct standard errors

Practice 2

- Reproducing PISA 2003 Table 4.3a, page 392
 - Linear regression model (PV) Maths on ESCS
 - Quadratic regression model (PV) Maths on ESCS
 - Mean, standard deviation and skew of ESCS
 - Percentage of students with ESCS < -1.

Practice 3

- Reproducing Figure 4.9, page 176
 - Graph of gradient based on quadratic regression
- Use EXCEL spreadsheet [SLOPE.XLS](#)