

# 中學地理科非選擇題的高層次 認知能力測驗之設計

林聖欽\* 教授

劉臻 碩士

國立臺灣師範大學地理學系

## 摘要

高層次認知能力的操作型定義，是指個體在學習或處理事情時，表現出轉化、分析、判斷、決定與解決問題的一連串複雜的心理歷程。由於論文題的作答，常促使受試者需要透過應用、分析、評鑑、創造等高層次認知能力，因而非選擇題中的論文題，很適合作為高層次認知能力的測驗題型。

在進行地理的論文題設計前，宜先掌握六項編寫原則。在設計題目時，則可依五個流程進行設計：1.界定欲測量能力的內容細目；2.列出該內容細目的具體行為；3.選出較具有代表性的具體行為；4.將具體行為轉為文字敘述的題目；5.修飾與精簡題目的敘述。最後，在評分規準設定，則是依照每個子題，條列出所欲評量的內容細目之具體行為，並依照受試者所能完成的具體行為程度，區分出未達成、基礎、精熟的三個等第。

**關鍵詞：**12年國民基本教育、多元評量、非選擇題、高層次認知能力、地理教育

\*本篇論文通訊作者：林聖欽，通訊方式：t24010@ntnu.edu.tw。



# The Design of Higher-Order Thinking Skills Test of the Free-Response Questions in High School's Geography Education

Sheng-Chin Lin\* Professor  
Chen Liu Graduate Student

Department of Geography, National Taiwan Normal University

## Abstract

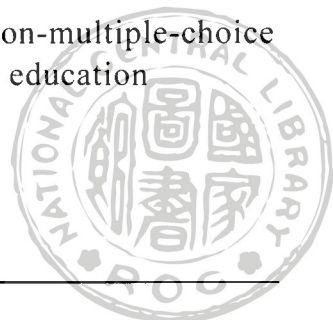
The higher-order thinking skills are someone's a series of complex psychological processes of transformation, analysis, judgment, decision and problem-solving when he learn or deal with things. Because someone needs to answer essay questions through the application, analysis, evaluation, creation of high-level thinking abilities, the essay questions of the non-multiple-choice questions are suitable as a high-level thinking ability tests.

Before designing the geography questions, it is better to master the six principles about the composition of questions. When designing the questions, it according to the five processes: 1. Define the content of abilities to be measured. 2. List the specific behaviors of the above contents. 3. Elect the more representative of the specific behaviors. 4. Transfer the specific behaviors into the narrative questions. 5. Modify and simplify the questions. Finally, we need to set the scoring rubric. It lists the specific behaviors which want to assess by each question, and in accordance with someone complete the degree of specific behaviors, to distinguish score into "below basic", "basic" and "mastery" of the three rankings.

**keywords:** 12-Year Basic Education, multiple assessments, non-multiple-choice questions, higher-order thinking skills, geography education

---

\*Corresponding author: Sheng-Chin Lin, E-mail: t24010@ntnu.edu.tw



## 壹、序論

民國103年11月，教育部公布了十二年國民基本教育課程綱要總綱，作為107學年度國民教育實施的準則。其中在實施要點中的「學習評量與應用」，提到在新的教育制度下，「為了解學生的學習過程與成效，應使用多元的學習評量方式，並依據學習評量的結果，提供不同需求的學習輔導。」(教育部，2014)，可知在十二年國教中，評量的多元化是一大推動重點，而且其他如：補救教學、差異化教學等相關設計，基本上都與多元評量的理念密切相關。

依照十二年國民基本教育課程綱要總綱，多元評量至少指的可以是：

- 一、實施評量時間的不同：包括形成性評量、總結性評量、診斷性評量、安置性評量或學生轉銜評估等。
- 二、學生屬性的歧異：即評量內容應考量學生身心發展、個別差異、文化差異及核心素養內涵。
- 三、教學目標的區別：即評量內容要兼顧認知、技能、情意等不同層面的學習表現。
- 四、教育類型的差別：即因應特殊類型教育學生之個別需求，學校與教師應提供適當之評量調整措施。
- 五、評量方式的差異：如紙筆測驗、實作評量、檔案評量等多元形式。
- 六、評量報告形式的異同：包括量化數據與質性描述。

本文基於上述十二年國民基本教育課程綱要總綱中的多元評量精神，將針對社會學習領域的地理學門，說明非選擇題中，屬於高層次認知能力測驗評量方式的設計原則。由於長期以來，地理學門的評量多是以選擇題為主，即使是非選擇題，也多屬封閉性的填充題、簡答題等測驗題型，所能評量的認知歷程向度，較偏向記憶、了解的低層次。<sup>1</sup>因此，本文提出非選擇題中屬於應用、分析、評鑑、創造等高層次認知能力測驗的評量設計，不僅可以反映前述十二年課綱強調的評量方式多元性，實際上，它也可以呈現教育目標的多元性。

<sup>1</sup> 依照教育部國民及學前教育署公布之國民中學學生學習成就評量標準（試行版）社會學習領域——地理科，其將表現等級分為A至E五個等級，其中A等級（優秀）為分析認知歷程向度，B等級（良好）為理解認知歷程向度，C等級（基礎）為記憶認知歷程向度，D等級表不足，E等級表落後。基於此表，本文的操作型定義，係將布魯姆認知歷程向度中的記憶、理解視為低層次；將應用、分析、評鑑、創造視為高層次。參見：[http://www.sbasa.ntnu.edu.tw/SBASA/Subject/SubjectGeography\\_1.aspx](http://www.sbasa.ntnu.edu.tw/SBASA/Subject/SubjectGeography_1.aspx)（2016/4/20 瀏覽）。此外，目前一般的命題競賽，亦大致將題目的認知歷程向度分為記憶、理解、高層次三種分類。

以下本文將以文獻整理、行動研究等方法，針對下列三個層面，來說明非選擇題中屬於高層次認知能力測驗的評量設計原則。

- 一、高層次認知能力測驗與布魯姆的認知教學目標分類。
- 二、高層次認知能力測驗的題型與編寫原則。
- 三、高層次認知能力測驗的評分規準設定原則。

## 貳、高層次認知能力測驗與布魯姆的認知教學目標分類

實際上，臺灣地理教育評量並非沒有採用非選擇題的評量方式，歷年來的大學入學地理指定科目考試，每年皆有非選擇題的評量設計，而且配分幾乎占整份試卷的1/4左右（24分／100分），只是如同前述，這些非選擇題目的形式，多屬封閉式的填充題、簡答題<sup>2</sup>。若依照布魯姆（B. S. Bloom）認知教學目標的分類，這些填充題、簡答題的題型通常多偏向記憶、了解等低層級教學目標的評量；簡言之，若要設計高層次認知能力測驗的非選擇題題目，評量的認知歷程向度必須提升到應用、分析、評鑑、創造的層級。

這種將布魯姆認知教學目標分成低層次與高層次的目標分類，其實也是近10年來積極發展翻轉教育（Flipped Learning）的設計重點。2007年，美國科羅拉多州林地公園高中（Woodland Park High School）兩位化學老師 Jonathan Bergmann 和 Aaron Sams 開始嘗試將大量的線上教學影片應用至一般課程之中，其先把課程先分成線上課程與實體課程兩個流程，並將布魯姆的教學目標分成低層次與高層次兩類，其中線上課程主要達成低層次認知教學目標，而高層次認知教學目標則透過「以學生為中心」的實體課堂教學達成（Hamdan et al., 2013）。

### 一、布魯姆認知教學目標分類的修訂

自1956年布魯姆等人提出教學目標分類（A taxonomy for educational objectives）後，該分類廣受國內外教育界所青睞，用以設計各科的教材與教學方法。然而，歷經多年使用，且隨著相關領域的研究成果累積，對於1956年提出的分類系統已出現需要檢討的聲音（葉連祺、林淑萍，2003）。

2001年 Anderson、Krathwohl、Airasian、Cruikshank、Mayer、Pintrich、Raths 與 Wittrock

<sup>2</sup> 例如，104年指考非選擇題的參考答案，其中8分的配分，是屬於多組答案中，答對其中一組答案即可的題目，如：以文中的國際移工為對象，舉出在哪個場所出現了什麼現象（或活動），最足以說明臺灣擁有「跨國社會空間」的事實；另外的16分配分，則屬只有1組答案的題目，如：在殖民統治時期，Atacama地區屬於哪個國家的殖民地。但不管多組還是一組答案的題目，皆屬封閉式的填充題、簡答題題目形式。

等人主編的《學習教學與評量的分類：Bloom 教育目標分類的修訂》一書，修訂 1956 布魯姆教學目標分類。基於有意義學習（meaningful learning）中的主動（active）、認知（cognitive）和建構歷程（constructive process）等概念，著重於學習者知（know，指 knowledge）與如何思考（how they think，指 cognitive process）的探討。因此，修訂版將其中的認知教學目標，分為「知識向度」（Knowledge Dimension）和「認知歷程向度」（Cognitive Process Dimension），前者在協助教師區分教什麼（what to teach），後者則在促進學生保留（retention）和遷移（transfer）所習得的知識（葉連祺、林淑萍，2003）。Anderson 等人將修訂版與 1956 年版本進行比較，提出三個層面的差異，分別是：重點強調、使用術語、目標分類結構，並將每個層面的差異分出四個項目，共計十二項的改變。大致來說，新版與舊版內容的更動情形可用圖1來說明<sup>3</sup>。

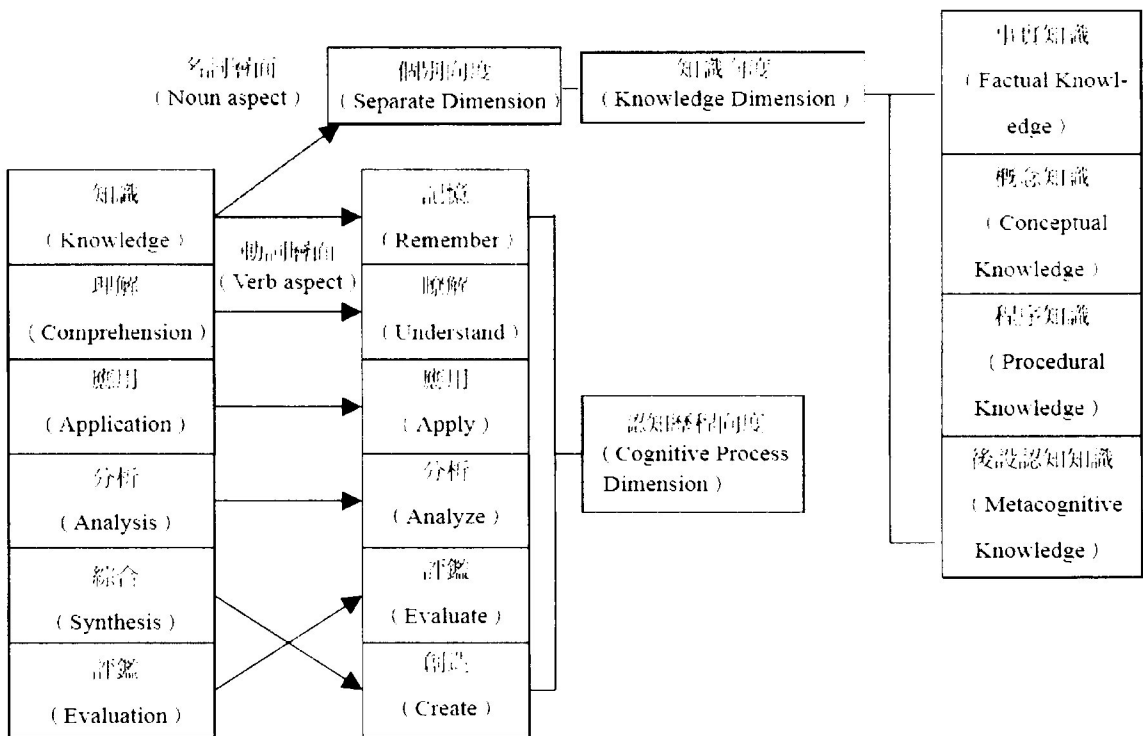


圖 1 布魯姆認知領域教學目標原架構與新架構的比較

資料來源：李坤崇（2004）：〈修訂 Bloom 認知分類及命題實例〉，《教育研究月刊》，122，頁 99。

<sup>3</sup> 其中布魯姆認知歷程向度的新舊版差異，可參見李坤崇（2004）：〈修訂 Bloom 認知分類及命題實例〉，《教育研究月刊》，122，一文。

## 二、新版布魯姆認知教學目標的知識向度與認知歷程向度分類

國內學者葉連祺、林淑萍(2003)及李坤崇(2004, 2008, 2009)將2001年Anderson等人主編的修訂版布魯姆教育目標分類進行整理,並提出新版知識向度與認知歷程向度中主類別與次類別的應用步驟及示例,其整理及摘要文字如下:

依照近年來認知心理學的研究,在修訂版中的知識向度分為事實知識(factual knowledge)、概念知識(conceptual knowledge)、程序知識(procedural knowledge)、後設認知知識(meta-cognitive knowledge)四項。

- (一) 事實知識：指學生在經過學習之後應該了解,以及解決問題時所應該具備的知識,常被視為獨立且特定的元素。包括：1.術語知識；2.特定細節及元素知識。
- (二) 概念知識：指在較大、較複雜的結構中,其基本元素間,歸納共同功能屬性,並將之與以分類及區別關係成為一類事物全體的知識。包括：1.分類和類別的知識；2.原理和原則的知識；3.理論、模式、結構的知識。
- (三) 程序知識：指知道如何進行完成某事的知識,常是指一系列或一連串的步驟流程,或是判斷何時要使用適當程序的規準。包括：1.特定學科技能和演算的知識；2.特定學科技術和方法的知識；3.決定何時使用適當程序的規準知識。
- (四) 後設認知知識：指對於認知的認知,對於自己認知能力與歷程的知識。包括：1.策略的知識；2.認知任務的知識,包括特有脈絡和狀態的知識；3.自我認知的知識。

至於在認知歷程向度方向,主要分為六項,由認知層次的低到高依序為「記憶」、「了解」、「應用」、「分析」、「評鑑」、「創造」。除了「記憶」是與學生的學習保留相關,其餘五項皆與學生的學習遷移較相關。此六項分類主要是基於建構主義者(constructivist learning)即有意義學習(meaningful learning)的看法,其目的在於使教師能夠配合學生的認知結構進行教學,以促進學生有意義的學習。

- (一) 記憶：是指學生從長期記憶中提取相關的知識,其知識可以是事實知識、概念知識、程序知識、後設認知知識,或是此四項知識的組合。「記憶」是學生進行學習的基礎,有了知識的記憶學生才能解決更複雜的問題。其評量方式可以是讓學生認得或記得某事物。記憶包括：1.1 再認；1.2 回憶。
- (二) 了解：指學生從教師的教學資訊中建構意義。換句話說,學生將所建立的新知識與故有舊經驗形成連結,藉以獲得新知識的理解。了解包括：2.1 詮釋；2.2 舉例；2.3 分類；2.4 摘要；2.5 推論；2.6 比較；2.7 解釋。
- (三) 應用：指透過一定的程序(步驟)來執行作業或解決問題,常與程序知識緊密結合。學生先得知執行的程序,再以此解決問題。應用包括：3.1 執行；3.2 實行。

- (四) 分析：將整體材料分解成局部，並指出局部間與整體結構的關聯性。主要是讓學生區別訊息的關聯性與重要性，並透過訊息的組織，覺察出隱含的訊息目的。分析包括：4.1 辨別；4.2 組織；4.3 歸因。
- (五) 評鑑：依據規準 (criteria) 與標準 (standards) 做判斷 (judgement)。規準是指品質、效果、效率與一致性；標準是指質的優劣，以及量的多寡。有明顯的規準及標準才能稱之為評鑑。評鑑包括：5.1 檢查；5.2 批判。
- (六) 創造：指將各元素重新組裝，形成一新的且具功能的整體。創造歷程須涉及協調過去學習經驗，並非毫無限制的自由創作。創作包括：6.1 產生；6.2 規劃；6.3 製作。

### 三、高層次認知能力測驗的操作型定義

透過 2001 年 Anderson 等人修訂的新版布魯姆認知教學目標的認知歷程向度分類，其將應用、分析、評鑑、創造等方面都視為高層次認知能力的範疇 (吳清山、林天祐，2010)。因而本文對於高層次認知能力的操作型定義，是指個體在學習或處理事情時，表現出轉化、分析、判斷、決定與解決問題的一連串複雜的心理歷程，屬於複雜層次思考，與記憶、了解等基礎層次認知能力高上許多。

若依照新版布魯姆認知教學目標的認知歷程向度分類，未來在設計高層次認知能力的測驗時，期待學生能夠達成的認知目標，其細節可以包含表 1 的內容。

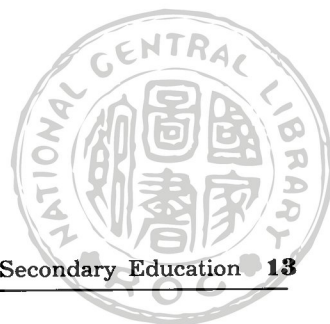
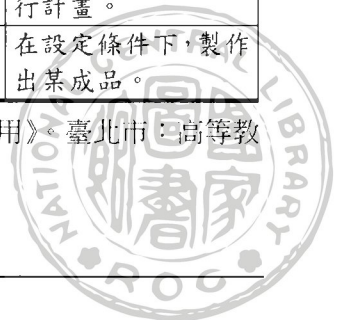


表 1 新版布魯姆認知教學目標的高層次認知歷程向度

主類別／次類別	相關詞	定義	示例
3.應用 (apply)		面對某情境執行或使用一個程序。	
3.1 執行 (executing)	進行 (carrying out)	應用一個程序於熟悉的任務。	(計算題) 整數加、減、乘、除。
3.2 實行 (implementing)	運用 (using)	應用一個程序於陌生的任務。	(應用題) 運用習得經驗解答一題陌生情境的應用題。
4.分析 (analyze)		將整體分解為部分，並決定整體與部分，以及部分與部分彼此間的關係。	
4.1 辨別 (differentiating)	區別 (discriminating) 分別 (distinguishing) 聚焦 (focusing) 挑選 (selecting)	自整體中，根據關聯性與重要性，區分局部。	從閱讀文章中區辨與解題有關或無關的訊息。
4.2 組織 (organizing)	尋找 (finding) 連結 (coherence) 概述 (outlining) 剖析 (parsing) 結構化 (structuring)	確定要素在結構中的適合性與功能。	依據全班體重，繪製次數分配質方圖。
4.3 歸因 (attributing)	解構 (deconstructing)	確認現有材料中隱含的觀點、偏見、價值觀或意圖。	故事隱含的意義。
5.評鑑 (evaluating)		根據規準與標準下判斷。	
5.1 檢查 (checking)	協調 (coordinating) 檢視 (detecting) 監視 (monitoring) 施測 (testing)	檢視某程序或產品中的不一致性或錯誤，確定某程序或產品的內部一致性，察覺正在運行的程序之效能。	檢查實驗過程的問題；檢查實驗結果與某定律間的一致性。
5.2 批判 (critiquing)	判斷 (judging)	檢視產品和外部規準的不一致性，確認產品是否有外部一致性，察覺解決問題的方式適切性。	從提升學習績效來判斷學校取消寒暑假的適切性。
6.創造 (create)		集合要素以組成一個具協調性或功能性的整體，重組要素為一個新的模型或結構。	
6.1 產生 (generating)	提出假設 (hypothesizing)	根據許多規準，建立假設。	提出繪製主題地圖的標準作業流程。
6.2 規劃 (planning)	設計 (designing)	建立一個操作程序已完成某些工作。	提出解決問題的可行計畫。
6.3 製作 (producing)	建立 (constructing)	發明新產品。	在設定條件下，製作出某成品。

資料來源：李坤崇 (2009)：《認知情意技能教育目標分類及其在評量的應用》。臺北市：高等教育，頁 28-31。



## 參、高層次認知能力測驗的題型與編寫原則

在進行高層次認知能力測驗的題型設計時，大致有兩個問題需要考量，一是評量要以應用、分析、評鑑、創造等認知歷程向度作為目標，二是評量的題幹設定要有助於評分標準的擬訂。

### 一、高層次認知能力測驗的題型

評量認知能力的測驗題型大致可以分為兩種，選擇反應題型與建構反應題型（圖2）。選擇反應題型的特徵為受試者須從特定選項中判斷出正確答案，常見的題型有是非題與選擇題，其測驗的認知歷程方面可以達到記憶、了解，有時甚至可以達到應用、分析，但通常無法測驗評鑑與創造等更高層次的能力（陳柏熹，2011）。

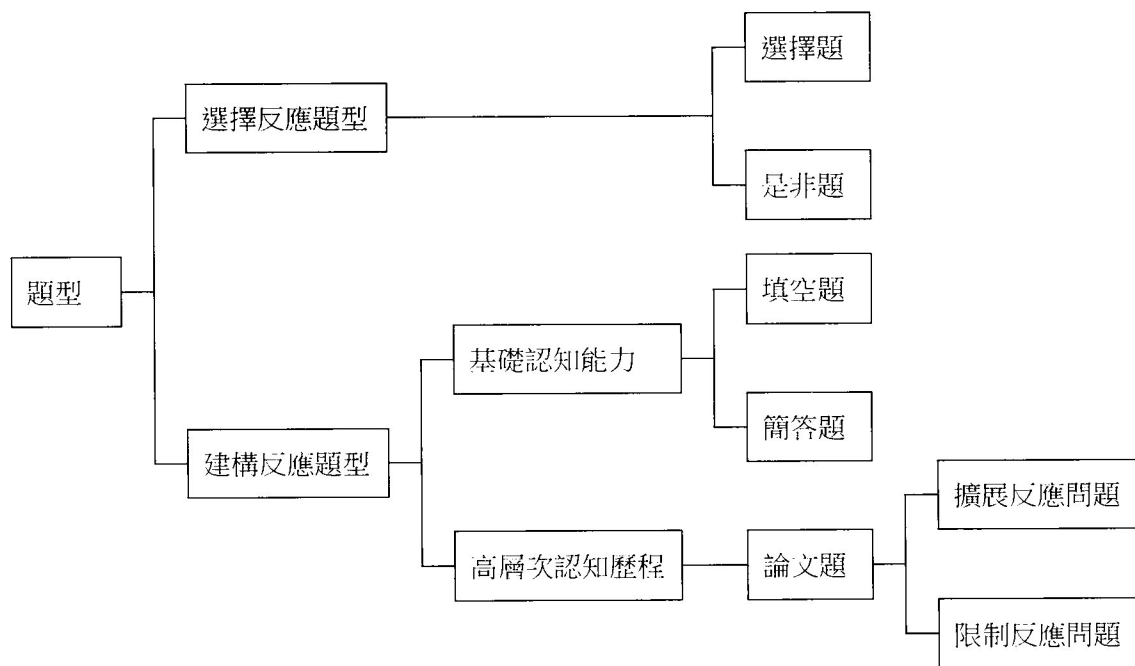


圖2 題型分類下的論文題位置

建構反應題型又可分為兩類，一類是測量基礎認知能力的建構反應題型，像是填空題、簡答題。填空題因為受試者須回答出特定的答案，因此使用記憶及了解的能力即可；簡答題則可以測驗到應用的能力。另一類，則是測驗高層次認知歷程的建構反應題，題型主要為論文題，由於論文題的作答將要求受試者結合多項概念來完成某項計畫或解決問題，又或是給予受試者解決問題的條件以及未曾見過的問題情境，讓受試者根據某一標準來判斷事物，要求其應用所學解決問題。因此，若要測驗受試者應用、分析、評鑑、創造等高層次認知能力，使用論文題的題型是較佳的選擇（陳柏熹，2011）。

實際上，論文題並非是俗稱的長文題，而且這種說法也簡化了論文題題幹內容的表現方式，論文題之所以稱為論文，是指受試者的回答須以論述性的文字表現，而非題幹是以長文的方式來呈現。至於論文題題幹內容的表現方式，除了閱讀式之外，舉凡圖表式、情境式、地圖式、實驗式、類推式等，都有助於評量受試者高層次的認知能力（郭生玉，2004）。

而論文題又可分為限制反應問題與擴展反應問題兩種次類型。前者給予學生部份作答的自由，常使用介紹性題材或具體指導語來限制學生反應在部分特定形式或範圍，常運用列出、界定、說明理由等術語，有時會限制學生作答篇幅；而後者又稱為申論題，僅稍做適當限制讓學生表現要測量的教學內涵，給予學生相當大的反應自由，表達其組織、綜合、統整、評鑑、創意與表達能力（李坤崇，2006）。

## 二、高層次認知能力試題的編寫原則

1985年，Edys Quellmalz 曾提出在設計高層次認知能力試題時的六點基本建議：

- （一）設計的問題必須可以回歸反映到學術科目的領域，針對學生學習過的科目領域進行測驗，學生需應用推理能力回答問題。例如進行不同程度的解釋，因果關係的解釋，評估優點與缺點等。
- （二）命題答案的設計必須平衡知識的廣度與深度，不僅重視學生可以回答出多項的答案，還必須注意命題時要有前後整合性、連貫性。讓學生回答出的每個答案之間有所關聯。
- （三）設計的問題允許多元的解釋與解決方法，學生運用所學知識推論出合理的答案，由學生採用的資訊是否正確，推論是否合乎邏輯，進行評量。
- （四）開放式答題要求學生解釋其推理過程，改變以往選擇題、填充題、簡答題的形式。要求學生以短文的形式闡述自己的推理過程。
- （五）建立的問題必須要訓練學生進行概括化的答題以及概念上的遷移，課題的設計可以讓學生產生學習遷移，舉一反三，將所學之概念應用至現實生活的情境之中。
- （六）發展可以評量學生後設認知的課題，舉例來說，可以使用情境狀況，當一個問題解決到一半的時候，請學生回答如果換作是他，將如何繼續解決該問題，或是讓學生寫出一個計畫，如何解決問題。（Quellmalz, 1985）

實際上，臺灣地理學門的教學不能說從未注意論文題的題型，只是在大型考試領導教學的過程中，論文題因非評量的重點，才導致教育現場對其關注非常有限。例如：國內教育測驗學者郭生玉，先在1985年提到了論文題的編寫原則，當然其所提出的是各學科的論文題通則。



- (一) 只有在客觀試題無法滿意測量學習的結果時才使用論文題。
- (二) 明確界定所預測量的行為。
- (三) 題目的敘述需使學生知道要回答的方向。
- (四) 標明每一題的回答時間，並給予充分的作答時間。
- (五) 避免使用選題作答方式。
- (六) 使用多題短答的方式比少題長答的方式為佳。(郭生玉，1985)

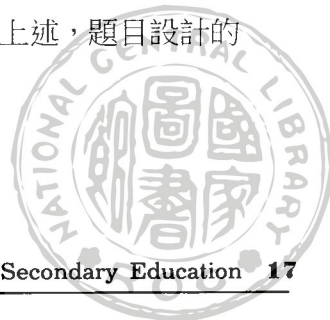
不過就其上述來看，郭生玉所提的論文題編寫原則，比較傾向是限制反應問題的論文題類型，且適用於形成性評量與總結性評量。也許是這個緣故，地理學門，很快地在1990年，國內的施添福、王秋原、陳文尚、陳國川等地理學者，即在郭生玉提出的編寫原則基礎上，於《大學入學考試地理科試題評鑑和命題原則》中，專門針對限制反應問題的論文題，提出以下的編寫原則：

- (一) 試題應只用於測量學生複雜的學習結果。
- (二) 題幹敘述，必須明確界定所欲測量的行為。必要時，可將題目再細分為若干小題。
- (三) 題幹的敘述，要使學生知道所要回答的方向。
- (四) 一個試題若區分為若干小題時，必須明確指出各小題的佔分。
- (五) 試題的編製，盡量使用多題短答的方式，即題目較多，但各題的答案要短而明確，以利評分的客觀性，並增加試題取樣的廣度。
- (六) 避免使用選題作答的方式。即避免讓學生在若干題目中選幾題作答，以免影響測量結果的效度。(施添福等，1990)

施添福等所歸納出來的編寫原則，後來在2012年臺灣國民中學社會學習領域輔導團進行地理學門非選擇題的研發時繼續被採用(賈生玲等，2012)，在此即不再特別說明。

### 三、高層次認知能力試題的設計流程

在對高層次認知能力試題的題型、編寫原則有基礎的認識後，接著可依照陳柏熹(2011)提出的五項試題設計流程，進行論文題的設計，以下即以實際題目為案例來介紹；惟考量地理教育目前所發展的高層次認知能力試題編寫原則，以限制反應問題的論文題為主，加上現階段教學現場對論文題尚不熟悉，因此以下暫時以限制反應問題的論文題為例，未來有機會再針對擴展反應問題的論文題進行討論。依據上述，題目設計的最後成果，詳見第四項的內容。



### (一) 界定欲測量能力的內容細目

假設現在要設計與「人口成長」概念有關的題目，則須先釐清此概念其實是由多個小概念所組成，包括出生率、死亡率、移出率、移入率、自然增加率、社會增加率、人口成長率、人口轉型模式、推拉理論（圖 3）。至於命題情境，則擬以中華民國友邦巴拉圭與布吉納法索兩國的人口成長為主要內容（表 2）。

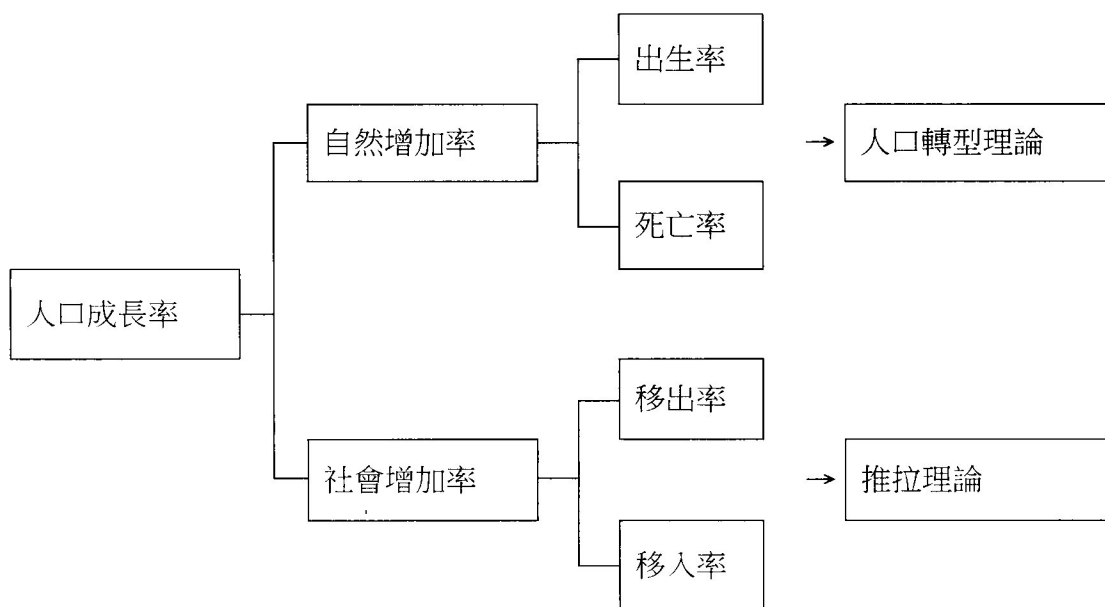


圖 3 人口成長率、人口轉型理論、推拉理論的概念構圖

表 2 2000—2012 年巴拉圭、布吉納法索的人口資料

年度		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
巴拉圭	出生率(‰)	31.27	30.88	30.50	30.14	29.78	29.43	29.10	28.77	28.47	28.17	17.73	17.48	17.22
	死亡率(‰)	4.81	4.75	4.69	4.64	4.58	4.53	4.49	4.54	4.49	4.46	4.55	4.57	4.59
	平均壽命	70.08	70.35	70.62	70.87	71.11	71.33	71.51	71.67	71.81	71.93	72.03	72.11	72.19
	人口數(百萬)	5.35	5.46	5.57	5.68	5.79	5.90	6.01	6.13	6.24	6.35	6.45	6.57	6.68
	人口成長率(‰)	20.8	20.4	20.1	19.7	19.3	19.0	18.6	18.2	17.9	17.7	17.5	17.4	17.2
布吉納法索	出生率(‰)	45.26	44.79	44.78	44.34	44.46	44.17	45.62	45.28	44.68	44.33	43.98	43.59	43.20
	死亡率(‰)	17.04	17.05	17.07	18.76	18.79	18.86	15.06	15.31	13.59	13.30	13.02	12.74	12.47
	平均壽命	50.48	50.86	51.27	51.72	52.18	52.66	53.14	53.62	54.09	54.56	55.01	55.54	55.86
	人口數(百萬)	11.61	11.95	12.30	12.66	13.03	13.42	13.82	14.24	14.66	15.10	15.54	15.99	16.46
	人口成長率(‰)	28.5	28.7	28.9	29.1	29.2	29.3	29.4	29.4	29.4	29.3	29.1	28.9	28.6

資料來源：美國 CIA，WorldFactBook。



若進一步將該題所欲評量的認知能力設定為「分析」，亦即指將材料分解成部分，並指出部分與部分之間，以及部分與整體結構之間的關聯。因此，試題設計的方向，可朝向讓受試者透過閱讀人口資料，讀出巴拉圭與布吉納法索兩國的出生率與死亡率變化趨勢，並分析兩國分別處於人口轉型歷程中的哪一階段。

## (二) 列出該內容細目的具體行為

受試者在作答該題時應出現下列行為，以展示其分析的能力：

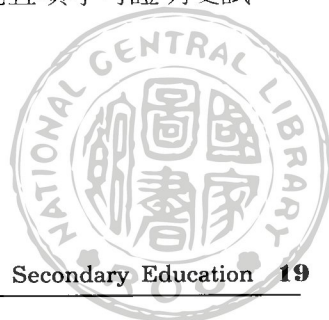
1. 能正確描述出生率、死亡率的變化趨勢。
2. 能依據自己對於數據的描述判斷巴國處於「人口轉型模式」的哪一個階段。
3. 能依據自己對於數據的描述判斷布國處於「人口轉型模式」的哪一個階段。
4. 能判斷巴國為晚期擴張。
5. 能判斷布國為早期擴張。
6. 能確實引用數據分析出生率、死亡率的變化趨勢。
7. 能使用 30‰標準，定義高出生率、高死亡率。
8. 能使用 10‰標準，定義低出生率、低死亡率。

受試者首先需從人口資料中正確讀出巴、布兩國出生率與死亡率的趨勢。在 2000—2012 年當中，巴拉圭的出生率呈現逐年下降的趨勢，從 2000 年的 31.27‰下降至 2012 年的 17.22‰，死亡率的波動則趨於穩定且數值偏低，皆低於 10‰；布吉納法索的出生率波動趨於穩定且數值偏高，皆高於 30‰，在死亡率方面，除了 2000—2005 年間有上升的現象，從 2000 年的 17.04‰升至 2005 年的 18.86‰，但整體來說還是屬於下降的趨勢。因此，正確的描述應為巴拉圭出生率下降且死亡率低；布吉納法索死亡率下降且出生率高。

接著，依據人口轉型的概念，若一地區出生率與死亡率均高則屬於高穩定階段；出生率高但死亡率下降則屬於早期擴張階段；死亡率低且出生率下降則屬於晚期擴張階段；出生率與死亡率均低則屬於低穩定階段。受試者須根據自身對於巴、布兩國出生率與死亡率變化趨勢的敘述，分析此兩國處於人口轉型模式的哪一階段。

## (三) 選出較具有代表性的具體行為

以前述具體行為中的 1~5 項作為具代表性的具體行為，正確的閱讀人口資料，並分析該資料在人口轉型模式中所代表的意義。受試者必須同時達到此五項才可證明受試者具備基礎的「分析」能力。



#### (四) 將具體行為轉為文字敘述的題目

將具有代表性的具體行為轉換為試題文字：根據表中所顯示的 2000—2012 年的人口資料，判斷巴拉圭與布吉納法索兩國分別處於「人口轉型模式」的哪一個階段。請分別描述巴、布兩國的出生率與死亡率，並寫出判斷兩國「人口轉型模式」的依據標準。

#### (五) 修飾與精簡題目的敘述

將上述試題文字修飾、精簡後，確認提議敘述清楚，並加入配分引導受試者應答。

### 四、高層次認知能力試題的案例

透過了前述高層次認知能力試題的題型、編寫原則、試題設計流程，最後設計出來的實際題目如下所示：

巴拉圭與布吉納法索皆是中華民國重要的邦交國，下表為巴拉圭與布吉納法索的人口資料。請問：

三覽		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
巴拉圭	出生率(‰)	31.27	30.88	30.50	30.14	29.78	29.43	29.10	28.77	28.47	28.17	17.73	17.48	17.22
	死亡率(‰)	4.81	4.75	4.69	4.64	4.58	4.53	4.49	4.54	4.49	4.46	4.55	4.57	4.59
	平均壽命	70.08	70.35	70.62	70.87	71.11	71.33	71.51	71.67	71.81	71.93	72.03	72.11	72.19
	人口數(百萬)	5.35	5.46	5.57	5.68	5.79	5.90	6.01	6.13	6.24	6.35	6.45	6.57	6.68
	人口成長率(‰)	20.8	20.4	20.1	19.7	19.3	19.0	18.6	18.2	17.9	17.7	17.5	17.4	17.2
布吉納法索	出生率(‰)	45.26	44.79	44.78	44.34	44.46	44.17	45.62	45.28	44.68	44.33	43.98	43.59	43.20
	死亡率(‰)	17.04	17.05	17.07	18.76	18.79	18.86	15.06	15.31	13.59	13.30	13.02	12.74	12.47
	平均壽命	50.48	50.86	51.27	51.72	52.18	52.66	53.14	53.62	54.09	54.56	55.01	55.54	55.86
	人口數(百萬)	11.61	11.95	12.30	12.66	13.03	13.42	13.82	14.24	14.66	15.10	15.54	15.99	16.46
	人口成長率(‰)	28.5	28.7	28.9	29.1	29.2	29.3	29.4	29.4	29.4	29.3	29.1	28.9	28.6

1. 根據表中所顯示的出生率與死亡率資料，判斷巴、布兩國分別處於「人口轉型模式」的哪一個階段？請分別描述巴、布兩國的出生率與死亡率，並寫出判斷兩國「人口轉型模式」的數字標準。
2. 當一國的經濟生產力不足時，將會產生推力，促使人民移出前往他國工作謀生。相反的，當國家經濟發展較興盛時，便會產生拉力，使他國人民移入該國工作。請依據 2012 年巴拉圭與布吉納法索的人口資料，計算出巴、布兩國在 2012 年哪一國的推力較強。需寫出計算過程，並說明判斷依據。



(一) 在題型部分，此試題屬於限制反應問題的論文題。主要原因在於該題主要以表格作為命題主題幹，而兩個子題的回答，需要透過「應用」、「分析」的認知能力才能進行作答；而答题的方向，則被限制在「人口轉型理論」、「推拉理論」、「人口增加率」的主要概念內進行回應，又這些概念的理解，得依賴相關的低階層知識(圖3)，因而學生在利用這些概念進行解題時，很自然就會表現出轉化、分析、判斷、決定與解決問題的一連串複雜的心理歷程，故題目的設計符合高層次認知歷程向度的設計目標。

此類題目若是以選擇題方式來命題，往往第1子題就會將人口轉型理論中的高穩定時期、前期擴張時期、晚期擴張時期、低穩定時期作為選項，等於提供學生答题更多的線索；而第2子題也勢必無法要求寫出計算過程，進而無法判斷學生答题是猜對，還是真正理解推拉理論的概念。

在該題組第1子題部分的設計，表中數值的變動趨勢，是依實際國家內部情勢而有所變動，受試者需依據整體時間尺度，「辨別」數據資料中有代表性的數值，判斷出生率與死亡率的整體變化趨勢，並說明此趨勢在人口種型模式當中的意義，進一步分析巴、布兩國分別處於人口轉型模式中的哪一階段。因而，此題在布魯姆認知歷程向度上屬於高層次認知能力「分析」中的4.1「辨別」：自整體中，根據關聯性與重要性，區分出局部。若受試者僅寫出巴、布兩國的人口轉型階段，沒有寫出判斷依據或是無法正確判斷兩國的人口轉型歷程階段，則表示受試者未達到認知歷程向度中高層次認知能力「分析」的程度。

而該題組第2子題部分的設計，則是要讓學生透過人口增加率的公式計算巴、布兩國的社會增加率，進而應用推拉理論的概念說明這些數據資料的意義。因而，此題在布魯姆認知歷程向度上屬於高層次認知能力「應用」中的3.2 實行：應用一個程序於陌生的任務。若受試者僅能計算巴、布兩國的社會增加率，但無法將其與他拉理論的概念作連結，則表示受試者未達到認知歷程向度中高層次認知能力「應用」的程度。

(二) 在命題原則部分，若依照施添福、王秋原、陳文尚、陳國川等地理學者，專門針對限制反應問題的論文題提出的命題原則，至少該題符合以下幾項要求：該題的確是複雜的學習結果，該題組至少細分了兩個子題，而兩個子題又各自問了兩個小問題；題幹敘述明確提供了人口轉型理論、推拉理論、人口增加率的答题方向，盡量使用多題短答的方式，未提供學生以選題作答的方式進行評量。

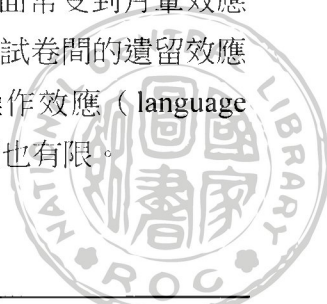
從表3中，可以見到學生回答第1子題的方式，先不論其答题情況，但明顯地，三位學生的答题方向限縮在人口轉型理論，而且也是以短答的方式進行作答，而非長文式的作答情況。

表3 三位同學針對案例題的答題情況

學生	作答情況	等第
甲	<p>巴拉圭—晚期擴張，出生率從20‰<sup>前</sup>降到10‰，而死亡率一直都很低(4‰)，為晚期擴張時期。</p> <p>布吉納法索—早期擴張，死亡率由24‰<sup>前</sup>降到15.24‰，而出生率一直很高(30‰)，為早期擴張時期。</p>	精熟
乙	<p>巴拉圭 晚期擴張階段</p> <p>布吉納法索 早期擴張階段</p> <p>巴拉圭：出生率大幅下降 死亡率下降幅度不大且死亡率低</p> <p>布吉納法索：出生率小幅度下降且出生率高 死亡率大幅下降</p> <p>判斷標準為出生率和死亡率數值高低和變化幅度</p>	基礎
丙	<p>巴：晚期擴張 布：高穩定</p> <p>∴ 出生率漸少 死亡率低 且平均壽命漸高</p> <p>∴ 出生率、死亡率都高 且平均壽命較短</p>	未達成

### 肆、高層次認知能力測驗的評分規準設定原則

雖然論文題的優點，如論文題的特徵一般，可測量到其他題型難以測量的複雜學習結果；但論文題的缺點與限制在於測驗信度不佳，在評分者信度方面常受到月暈效應 (halo effect)、項目間的遺留效應 (item-to-item carryover effect)、試卷間的遺留效應 (test-to-test carryover effect)、次序效應 (order effect) 和文字操作效應 (language mechanics effect) 的影響，縱使是閱卷經驗豐富的教師，評分者信度也有限。



此外，評分者信度高不代表測驗本身的信度高，由於論文題測驗的寫作時間較長，因此論文題測驗的題數較少，當測驗的題數少隨機因素的影響就會增加，導致測驗的信度降低。為改善論文題測驗的信度，李茂興（2002）提出了三個論文題評分的改進方法：一、先依測驗的目標決定答案的重點，二、就每個答案的每個要點進行配分，三、先批改所有試卷在同一道題目上的作答，然後再批改另一道題目。其中第一、二項，其實指的就是評分規準的設定。

## 一、高層次認知能力試題的評分規準設定

不過就論文題而言，測驗主要應該是評量受試者的能力，根據受試者高層次認知能力的達成度，作為學習成效的判斷依據，因而論文題非常適合以標準參照測驗（*criterion-referenced testing*, CRT）的角色去評量受試者的學習成效，因此，高層次認知能力測驗的論文題，其評分方式採等級計分法較佳，可依照受試者對於題目的反應優劣分成幾個等級。

而這個等級計分法，基本上可以分為 A、B、C 三等第：A 等第為精熟，代表受試者熟練使用原理、法則或規準，精確、完整地回答試題；B 等第為基礎，代表受試者有使用原理、法則或規準回答試題；C 等第為未達標準，代表受試者未使用，或未正確使用原理、法則或規準回答試題。簡單來說，C 是受試者未達應用、分析、評鑑、創造的高層次認知能力，B 是受試者可以初步達到高層次認知能力，A 是受試者已經相當符合高層次認知能力。

一般而言，每個子題設計出來後，設計者需再條列出期待受試者的作答步驟，也就是該子題所欲評量的內容細目之具體行為；接著判斷作答步驟所代表的認知能力層次高低；最後依照受試者所能完成的步驟程度，區分出 C 等第（未達標準，未達到高層次）、B 等第（基礎，初步達到高層次）、A 等第（精熟，已相當符合高層次）。

## 二、高層次認知能力試題的評分規準設定案例

以下以前一節巴拉圭與布吉納法索的人口地理題目為例，說明該題的評分規準設定流程。

首先，針對該高層次認知能力試驗，列出期待受試者的作答步驟，亦即該子題所欲評量的內容細目之具體行為：

- （一）能正確描述出生率、死亡率的變化趨勢。
- （二）能依據自己對於數據的描述判斷巴國處於「人口轉型模式」的哪一個階段。
- （三）能依據自己對於數據的描述判斷布國處於「人口轉型模式」的哪一個階段。
- （四）能判斷巴國為晚期擴張。

- (五) 能判斷布國為早期擴張。
- (六) 能確實引用數據分析出生率、死亡率的變化趨勢。
- (七) 能使用 30‰標準定義高出生率、高死亡率。
- (八) 能使用 10‰標準定義低出生率、低死亡率。

接著，判斷作答步驟所代表的認知能力層次高低。

- (一) 就上述的 8 個步驟中，若受試者能夠完成前 5 個，即能初步達到高層次認知能力。但因為無法排除學生是否以猜測方式得到答案，因此僅將其視為初步達到高層次認知能力。
- (二) 受試者除完成前 5 個作答步驟外，又能完成（六）～（八）步驟中的其中一項，即可判斷受試者並非以猜測方式得到答案，並且其的確已經符合高層次認知能力的「分析」中的 4.1 區辨：自整體中，根據關聯性與重要性，區分出局部。
- (三) 就上述的 8 個步驟中，若受試者僅能夠完成前 5 個的其中幾項，但卻能完成（六）～（八）步驟的其中一個時，因其無法完成前 5 個步驟，故無法排除學生可能以猜測方式完成（六）～（八）的步驟，因此仍將其無法達到高層次認知能力。

最後，設計者即能夠依照受試者所能完成的步驟程度高低，區分出 C 等第、B 等第、A 等第。

C 等第（未達到標準），即受試者未能達到高層次認知能力中的分析，主要是未能同時達到（一）～（五）項的作答步驟。包括以下幾種情況：

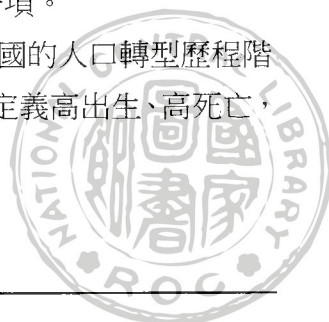
- 未作答。
- 僅描述或錯誤描述出生率與死亡率的變化關係，且未寫出或未正確寫出巴國與布國的人口轉型歷程階段。
- 未描述或錯誤描述出生率與死亡率的變化關係，但正確寫出巴國與布國的人口轉型歷程階段。

B 等第（基礎），即受試者能達到基本的分析層次，主要是能同時達到（一）～（五）項的作答步驟。

- 正確描述出生率與死亡率的變化關係，且正確寫出巴國與布國的人口轉型歷程階段。

A 等第（精熟），即受試者能達到精熟的分析層次，主要是能同時達到（一）～（五）項的作答步驟外，而且還能達成（六）～（八）作答步驟中的任一項。

- 正確描述出生率與死亡率的變化關係，且正確寫出巴國與布國的人口轉型歷程階段，以及能引用數據描述出生率與死亡率，與使用 30‰標準定義高出生、高死亡，10‰標準定義低出生、低死亡。



依上述的評分規準設定，若受試者在該題未作答，或者在答案中未正確描述出生率與死亡率的變化關係，且未同時正確寫出巴國與布國的人口轉型歷程階段，則受試者在該題的得分為 C（未達到）等第。若受試者能正確描述出生率與死亡率的變化關係，且正確寫出巴國與布國的人口轉型歷程階段，則可以達到 B（基礎）等第。最後，若受試者還可以引用數據描述出生率與死亡率，以及使用 30% 標準定義高出生、高死亡，10% 標準定義低出生、低死亡，則受試者在該題的得分將達到 A（精熟）等第。

學生實際答題的內容可見表 3，其中甲生即屬 A 等第（完成（一）～（五）的答題要求，而且也符合（七）（八）的答題要求），乙生屬 B 等第（僅完成（一）～（五）的答題要求），丙生屬 C 等第（無法達成（三）、（五）的答題要求）。但要特別強調的是，C 等第的學生只是未能達到高層次認知能力，但並不代表他無法達到記憶、了解的低層次認知能力。

### 三、其他評分規準設定的相關問題

評分規準的設定，還有技術性的問題必須思考，這是未來尚需繼續處理的課題。

（一）上述案例的評分規準，其實只是針對單一子題來進行評分規準的設定，而非是整個題組。因此，若要設計整個題組的評分規準，則題組內部的子題，彼此間需要形成問題鷹架，且各子題的排列順序，也需由低層次的認知能力排到高層次，才能評量出受試者是否達到高層次的認知能力。

（二）上述案例的評分規準，雖然是按照標準參照測驗精神來設定，但若就一整份試卷角度來看，則每個題目之間的 A、B、C 等第的評分規準，就需要在同個基礎上進行設定，才不會造成題題不等值的問題出現。因此，雖然目前將應用、分析、評鑑、創造都視為高層次的認知能力，但在考量整份試卷各題之間評分規準的統一，應用、分析也許需要再與評鑑、創造區分成不同等級的認知能力，以避免上述題題不等值的問題出現。尤其是若該份試卷是作為常模參照測驗（norm-referenced test），採取相對性標準比較，以判斷受試者成績在整個團體中所處位置，此時各題之間評分規準的統一更顯重要。

## 伍、結論

因應十二年國民基本教育課程綱要總綱的多元評量精神，並期待未來學生的學習能達成高層次的認知能力，本文針對了社會領域地理學門在設計非選擇題的論文題時，需要注意的編寫原則與評分規準的設定原則。

其中本文對於高層次認知能力的操作型定義，是指個體在學習或處理事物時，表現出轉化、分析、判斷、決定與解決問題的一連串複雜的心理歷程，屬於複雜層次思考，與傳統的記憶、了解等基礎層次認知能力略有不同。

為了能評量學生的高層次認知能力，本文認為論文題為適當的題型。主因論文題的作答常要求受試者結合多項概念來完成某項計畫或解決問題，又或是給予受試者解決問題的條件以及未曾見過的問題情境，讓受試者根據某一標準來判斷事物，並要求其應用所學解決問題，因而促使受試者需要透過應用、分析、評鑑、創造等高層次認知能力，才能針對題目進行作答。

地理學門論文題在設計前，需先把握一些編寫原則，以限制反應問題的論文題為例，注意事項有六項：一、試題應只用於測量學生複雜的學習結果。二、題幹敘述，必須明確界定所欲測量的行為，必要時，可將題目再細分為若干小題。三、題幹的敘述，要使學生知道所要回答的方向。四、一個試題若區分為若干小題時，必須明確指出各小題的佔分。五、試題的編製，盡量使用多題短答的方式，以利評分的客觀性，並增加試題取樣的廣度。六、避免使用選題作答的方式，以免影響測量結果的效度。

地理學門論文題在編訂時，則可依下列五個流程進行設計：一、界定欲測量能力的內容細目。二、列出該內容細目的具體行為。三、選出較具有代表性的具體行為。四、將具體行為轉為文字敘述的題目。五、修飾與精簡題目的敘述。

最後，地理學門論文題的評分規準設定：一、首先，係依照每個子題，條列出期待受試者的作答步驟，也就是該子題所欲評量的內容細目之具體行為。二、接著，判斷作答步驟所代表的認知能力層次高低。三、最後，依照受試者所能完成的步驟程度，區分出C等第（未達標準，未達到高層次）、B等第（基礎，初步達到高層次）、A等第（精熟，已相當符合高層次）。

由於上述只是針對單一子題來進行評分規準的設定，因而未來有關評分規準的設定，還有幾個技術性的問題必須思考：一、整個題組的評分規準的設定，這牽涉到題組內部的子題，彼此間需要形成問題鷹架。二、整份試卷各題之間評分規準的統一，以降低題題不等值的問題出現，這牽涉到是否需要進一步將應用、分析與評鑑、創造區分成不同等級的認知能力。實際上，這也是未來在大學入學考試推動高層次認知能力測驗題時，需要再去克服的問題。



## 參考文獻

- 李坤崇 (2004)。修訂 Bloom 認知分類及命題實例。《教育研究月刊》，122，98-127。
- 李坤崇 (2008)。修訂 Bloom 認知歷程向度命題實例。《教育研究月刊》，173，105-120。
- 李坤崇 (2009)。認知情意技能教育目標分類及其在評量的應用。臺北市：高等教育。
- 李茂興譯 (2002)。《教育測驗與評量》。臺北市：學富文化。
- 吳清山、林天祐 (2010)。高層次思考。《教育研究月刊》，200，123-124。
- 施添福、王秋原、陳文尚、陳國川 (1990)。《大學入學考試地理科試題評鑑和命題原則》。臺北市：中華民國大學入學考試中心。
- 陳柏熹 (2011)。《心理與教育測驗：測驗編製理論與實務》。臺北市：精策教育。
- 郭生玉 (1985)。《心理與教育測驗》。臺北市：精華書局。
- 郭生玉 (2004)。《教育測驗與評量》。臺北市：精華書局。
- 教育部 (2014)。十二年國民基本教育課程綱要總綱。臺北市：教育部。
- 賈生玲、李鳳華、蘇芳如、蒲思穎 (2012)。國中社會學習領域 (地理) 非選擇題紙筆評量試題研發之初探。《中等教育》，63 (2)，157-176。
- 葉連祺、林淑萍 (2003)。布魯姆 (B.S. Bloom) 認知領域教育目標分類修訂版之探討。《教育研究月刊》，105，94-106。
- Hamdan, N., McKnight, P., McKnight, K., & Arfstrom, K. M. (2013). *A review of flipped learning*. Retrieved from <http://www.flippedlearning.org/review>
- Quellmalz, EdysS (1985). *Needed : Better methods of testing higher-order thinking skills*. DOCUMENT RESUME, 338-353.

