

歷史測驗題型的兩難：如何檢驗出學生的歷史思維情形

蕭憶梅 助理研究員

◎ 國家教育研究院課程及教學研究中心

摘要

隨著國內中學歷史課程綱要強調「歷史思維（historical thinking）」的重要性，國中基本學力測驗與高中升大學的學科能力測驗與指定科目考試均宣稱測驗目標配合課程綱要中所強調的能力。然而，卻很少有人去分析這些宣稱檢測歷史思維能力的題目是否真的名符其實。目前，國內的歷史考試真的能夠檢測學生的歷史思維嗎？

研究主要探討學生在面對選擇題與非選擇題不同測驗题型時的表現有何異同。研究結果發現在測驗較基礎的歷史思維時，學生在選擇題與非選擇題的表現沒有明顯差異，但在測驗較進階的歷史思維時的表現卻相差懸殊。本文結論建議依據測驗目標來決定歷史測驗题型會是一個較折衷的作法，過份重視公平性而只採用選擇題，或只重視效度而全改用非選擇題命題的兩種極端都有其缺失。

關鍵詞：歷史測驗、選擇題、非選擇題、歷史思維

The Dilemma of Designing History Examinations: How to Assess Students' Historical Thinking

Yi-Mei Hsiao Assistant Research Fellow

◎ Research Center for Curriculum and Instruction, National Academy for Educational Research

Abstract

In recent years, the national history curriculum emphasized the importance of historical thinking. All of the national examinations claim that they test the ability in the national curriculum. However, can these history exams assess historical thinking?

The aim of this study was to examine how students react to different types of history questions such as multiple-choice questions and essay questions. The results showed that there was no distinguished difference between students' responses to multiple-choice questions and essay questions when they were assessed about basic historical thinking. On the other hand, there was an eminent difference when students were assessed about higher-level historical thinking. The findings suggested that the best way to make decisions about the type of exam questions is based on its assessment targets.

Keywords: history exams, multiple-choice questions, essay questions, historical thinking

壹、前言

當我們談到歷史知識時，實際上它包含了兩個不同層次的概念：第一種是實質概念（substantive concepts），即歷史知識的實質內容，如：革命、國家、佛教等。另一種是第二層次概念（the second order concepts），即理解歷史內容時會運用到的概念，例如：資料證據、因果解釋、時序變遷等（Lee, 2005）。近年來，國內的歷史課程綱要深受國外教育趨勢影響，強調「歷史思維（historical thinking）」的重要性，關注的焦點不再只是前者的實質概念，還擴及到後者的第二層次概念。例如：九年一貫課程社會領域「人與時間」的能力指標中就有兩點清楚地論及第二層次概念：（1）比較人們因時代、處境、角色的不同，所做的歷史解釋的多元性；（2）瞭解並描述歷史演變的多重因果關係（教育部，2010，p.25）。高中95暫綱與101年9月將實施的正式課綱亦明確指出時序、歷史理解、歷史解釋與史料四項核心能力（教育部，2005；2011）。

在課程綱要調整的情形下，國中基本學力測驗、高中升大學的學科能力測驗與指定科目考試均宣稱測驗目標配合課程綱要中所強調的能力，國中基測的每個示例題目都可對應一條課程綱要的能力指標，大考中心也會在試題分析中說明每個題目的測驗目標，如史料、歷史解釋等（蔡蔚群，2004）。然而，卻很少有人去分析這些宣稱檢測歷史思

維能力的題目是否真的名符其實。目前，國內的歷史考試真的能夠檢測學生的歷史思維嗎？

由於本身學科性質的特性，歷史考試無法只停留在測驗客觀知識的層面。因歷史論述是建構而來，本身就包含了主觀解釋性的元素。若只測驗客觀不變的歷史事實（如人物、時間或地點）易陷入偏重記憶的傳統窠臼，但若測驗事件發生的原因或影響，也常因不同解釋的緣故而使答案有爭議。因此，歷史測驗與評量的問題其實相當複雜，要在設計出良好的試題與確保評分客觀之間抓到平衡點並不容易。對於選擇題這種「標準化測驗（standardized test）」能否檢測學生的歷史思維一直是個爭議，目前有很多相異的論點（Dickinson, 1978；1984、Rothstein, 2004、蔡蔚群，2004）。在這樣的背景下，筆者決定嘗試用選擇題與非選擇題的兩種題型方式來設計同樣的歷史問題讓同一群學生作答，分析這些學生在不同形式的歷史題目下有何異同的表現，藉此探討不同歷史測驗題型的優缺點及其可能性。

貳、檢測歷史思維的問題設計

在歷屆國中基測、大學學測與指考的考古題，或坊間的參考書與測驗卷等，都可看到琳瑯滿目與史料相關的題目，因此筆者決定選取大家都很熟悉的史料題型來探討檢測



「史料證據」概念的相關議題。在決定要檢測「史料證據」概念後，接著就要面對選取知識內容的問題。當時因為筆者在高中任教時剛教到「世族與門閥政治」這個章節，因此決定就以「魏晉的選才制度」為知識內容來結合「史料證據」概念進行施測。

筆者在決定測驗知識內容與概念後，接著擬定三個由淺入深的測驗目標：

(一)學生能否從史料中選取有意義的（可用）資訊？

(二)學生能否判斷史料可否作為支持某項推論的證據？

(三)學生能否處理矛盾衝突的史料？

筆者依據這幾項測驗目標，選取了與「魏晉的選才制度」相關的四份史料（見附錄），並設計了相對應的三個問題。由於此項小型測試的目的在分析比較學生在不同歷史測驗題型的表現情形，因此針對這三個問題分別設計了選擇題與非選擇題兩種形式的題目（見表1、2、3）。這四份史料均為文言文，在原先設計試卷時曾考慮是否要改成白話文，但因顧慮到改成白話文可能會影響原文意思，因此決定先以原文施測，若學生反映閱讀文言文有困難時再進行修改。實際測試時學生表示這些史料的文言文並不算困難，所以就沒有再作修改。¹

表1 測驗目標（一）的兩種題目型式

測驗目標1：學生能否從史料中選取有意義的（可用）資訊？	
選擇題	根據資料一的內容，曹操對於選才的標準為下列何者？ (A) 只重品德，不論家世 (B) 只重品德，不論才能 (C) 只重才能，不論品德 (D) 只重家世，不論才能
非選擇題	請根據資料一的內容，說明曹操對於選才的標準為何？

表2 測驗目標（二）的兩種題目型式

測驗目標2：學生能否判斷史料可否作為支持某項推論的證據？	
選擇題	陳致平在《中華通史》一書中，論述曹魏行九品官人法的原因是因為漢末天下大亂，士人四處流亡，所以無法依照原來的察舉制度來選舉人才，而改行中正官來品選人物。請問在資料一到資料四中那份資料最足以佐證他的這項論述？ (A) 資料一 (B) 資料二 (C) 資料三 (D) 資料四
非選擇題	陳致平在《中華通史》一書中，論述曹魏行九品官人法的原因是因為東漢末天下大亂，士人四處流亡，所以無法依照原來的察舉制度來選舉人才，而改行中正官來品選人物。請問在資料一到資料四中有那份資料可以做為支持他的這項論述？請闡述你的理由。

1 雖然本次施測資料中以文言文呈現，但仍建議在用文言文的史料命題時，改成白話文是較為適當的，以避免學生因讀不懂文言文而導致出其真正的歷史思維無法被檢測出來。

表3 測驗目標（三）的兩種題目型式

測驗目標3：學生能否分析與比較矛盾衝突的史料？	
選擇題	資料二中提到九品官人法的選才標準是在於人才優劣，而非世族高低，但是為何資料三和資料四中卻說九品官人法中的上品多是公侯勢族？ （A）資料二談的是當初行九品官人法的原意，資料三和四談的是實行後衍生出來的情況 （B）資料三和四均出自於晉書，晉書的作者對曹魏有偏見因而刻意毀謗 （C）資料三和四的論點一致，所以應該是資料二的作者不明真相 （D）資料三和四的論述為真，因為晉書編寫的時間早於宋書，所以比較知道事情當時發生的原因
非選擇題	資料二中提到九品官人法的選才標準是在於人才優劣，而非世族高卑，但是為何資料三和資料四中卻認為九品官人法中的上品多是公侯勢族？

施測的對象為就讀於台北市一所高中一年級的23位學生。學生當時正在學習中國史，剛學完「世族與門閥政治」這個章節，對於九品官人法制度的相關知識有一定程度的認識，但他們未曾在課堂上接觸過這些史料與題目。施測題目分別以A卷（非選擇題型式）與B卷（選擇題型式）兩份試卷呈現。為避免學生先看到選擇題的選項而影響其在非選擇題的作答，學生拿到的第一份試卷為A卷的非選擇題，需在25分鐘內作答完後繳交試卷。收回全部的非選擇題試卷後，

同一批學生再接受B卷選擇題的測試，他們被要求要在10分鐘內完成。實際情形是學生不到5分鐘就全部作答完這三題選擇題。

參、測驗結果的分析與檢討

由於當時設計這份問卷時過於倉促，在收回試卷時發現第二道試題的設計不佳，未能符合測驗目標的涵意，原意是要測驗學生是否能判斷史料作為支持某項推論的證據，但因為選取的史料二內容與題幹中論述本身太過接近，學生並不需要思考「推論」與「史料」之間的證據關係就可輕易作答，而影響此道題目的效度，因此本文中對施測結果的討論僅限於第一題與第三題。原則上，學生在選擇題形式的試卷表現優於非選擇題型式的試卷（包括第二題²）。但仔細分析學生在各題目的答題情形，卻出現一些令人意想不到的結果。

在第一題（曹操對於選才的標準為何？）的施測結果上，同一群學生在選擇題試卷的答對情形為74%，非選擇題試卷則為87%。這一個題目算是很簡單的題目，只要能夠概略理解資料一中的文字內容就可正確作答，所以大部分的學生都答對。但為何這題的非選擇題作答情形竟高出選擇題？依常理推斷，要書寫出內容應該比選取一個選項來得困難，但為何結果卻是相反？將學生在兩份不同形式的答題情況相比較後發現，有4位

2 學生在第二題的選擇題試卷答對情形為78%，非選擇題試卷則為74%。



學生在非選擇題型式的這題能寫出「重視才能，無論品德」這個正確答案，但在答選擇題時卻選到其它錯誤答案。這幾位學生會寫出正確答案但卻選錯答案的原因應是他們在答選擇題時不小心「看錯選項內容」，可見在評估選擇題試卷的結果時時，也要考慮到有些學生可能因粗心而有較差的表現。

相對之下，試卷中的第三題（資料二中提到九品官人法的選才標準是在於人才優劣，而非世族高卑，但是為何資料三和資料四中卻認為九品官人法中的上品多是公侯勢族？）是較困難的題目，主要在測試學生是否理解九品官人法的制定背景與後續發展，及他們否處理矛盾衝突史料的能力。出乎意料的是，這班學生在選擇題的答對情形居然為100%，比第一題的簡單題目表現還要好。但他們在第三題選擇題與非選擇題型式的表現卻相差懸殊，非選擇題答得較完整的學生只有13%。學生在第三題非選擇題的表現不佳是可理解的，因這題處理的問題較複雜，學生需要交叉比對不同史料，並提出具體的解釋說明，而大部分的學生並沒有被鼓勵去思考這樣的問題，或有機會用文字的方式去論證某個想法。

但其中值得注意的是，用選擇題來測試學生時，他們全部都答對這道較困難的題目，這可能會讓大家有種錯覺，學生在「史料證據」概念的高層次思維情形有很好的表現。但仔細分析學生在這道非選擇題的作答情形，就可看出學生其實並沒有完全理解九品官人法的相關內容，對「史料」與不同

「推論」之間的關係也未能具體掌握。這些學生思考這個問題時展現出幾種特性：

- 一、僅以常理推論，未能運用史料
- 二、能擷取史料中的訊息，但未能對不同資料進行分析比較
- 三、能針對題意，對不同資料內容進行分析比較

以下針對這幾種不同層次思維過程舉例說明：

（一）僅以常理推論，未運用史料

有部分學生在回答過程中，會用一些通俗想法帶有偏見的觀點來解釋九品官人法中上品多為公侯勢族的現象，以一位林同學的回答為例：

公侯勢族有財力，權大勢大，若要求中正官提拔自己的子孫，中正官豈敢不從？況更可藉此打好關係，何樂而不為？而尋常家世的人，中正官踐踏都來不及了，何況為了他們擠掉公侯勢族的名額？

像林同學有類似想法的學生都只用某種「成見」去解釋上品多為世族的原因，如：「這是中國人的黑暗面，不能公正」、「小蝦米如何能對上大鯨魚呢？」這類似是而非的推論。但他們卻沒有運用史料中所提供的訊息來說明這個現象，也沒有回答這幾份史料為何對九品官人法中上品的成員背景有不同的描述。

（二）能擷取史料中的訊息，但未能對不同資料進行分析比較

另一群學生雖懂得應用史料中提供的訊息，但只停在擷取與摘要資料訊息的階段，

還未能進一步去對資料之間做出分析比較，如黃同學的回答：

所謂「選才在於人才優劣」只是個幌子，資料三提及「台閣選舉，徒塞耳目」，資料四提及公侯勢族賄賂自通的黑暗。

或只針對某份資料詳加解釋，如闡述資料三或資料四中九品官人法中的上品多是公侯勢族的原因，如賄賂、人情和世族勢力，但仍然沒有針對題意回答這幾份史料為何對九品官人法中上品的成員背景有不同的描述，如吳同學的回應：

就像資料四所說：愛憎難防，情偽難明，人物難也，而「愛憎決於心」這就是重點了，一個人的貪心會使他偏私，個人的情懷也會使他偏頗自己人，而自己為中正，所認識的人，大多是公侯勢族，所以偏頗他們也不會使人覺得奇怪。

(三)能針對題意，對不同資料內容進行分析比較

少數幾位學生卻能夠針對題意，說明為何資料中對上品的選才標準有不同的說法，如周同學：

因為「資料二」是指立法之初，大家的心態：用多方位考量選出人才，但久而久之，就如演變成「資料三」和「資料四」敘述的：各位主考官由於利益或人情，而選「地位」不選「人才」。

像周同學一樣能做出清楚回答的學生雖然不多，但卻非常難得，因為他們展現出在「史料證據」這項核心能力中高層次的思維過程，能做到處理不同史料中的矛盾敘述。

從學生在非選擇題上的書寫情形就可看出歷史思維的複雜性，這些不同層次的想法是無法透過傳統的選擇題顯現出來的。從選擇題的結果來看，全體的學生都選對了答案，彷彿他們都達到了測驗目標。但若從非選擇題的答題狀況就可發現，其實很多學生尙未能運用史料進行推論，並慣性地讓一些「成見」主宰他們的思考過程，如：「選才在於人才優劣只是個幌子」、「這是中國人的黑暗面，不能公正」。

由此可見，我們必須重新思索「測驗」的目的與方式。一份試卷是否真能檢測出原訂的測驗目標？當學生在某項測驗中表現很差時，它是否就代表學生的學習情形不佳？亦或其實是測驗本身的設計與方式不良？相反地，當學生在某項測驗中表現很傑出時，它是否就表示學生都達到預定的學習成果？

肆、天秤的兩端：效度與公平性

從有電腦讀卡的技術引進後，國內的歷史考試多半都以選擇題為主，目前的國中基本學力測驗、高中升大學的學科能力測驗與指定科目考試也多是選擇題的題型。選擇題是一種標準化測驗，好處是批改快速簡單，沒有閱卷者的個人主觀因素影響成績的結果。但其缺失就在於選擇題有其本身的侷限性，通常越以記憶為導向的題目，越容易以選擇題的方式出題（蔡蔚群，2004）。從前面的小型研究結果可發現，雖然選擇題並



非完全與歷史思維無關，但因受限於題型本身四選一的型式，它所能檢測的歷史思維層次是較初階的。無論是採取選擇題或非選擇題的方式，學生在第一題的表現上並沒有過大的懸殊，因此題的測驗目標是較基礎的，只在檢測學生是否能從史料中選取有意義的（可用）資訊。

然而，學生在第三題的表現卻由於不同測驗題型的方式而有極大差異，學生在回答選擇題的試卷時有相當傑出的表現，但在以非選擇題作答時就顯露出學生並非真正有達到測驗目標。因此，用選擇題的方式去測驗學生分析比較不同史料的能力是較不適當的，因為本身這題的測驗目標屬於高層次思維，牽涉到的論證過程是複雜的，將其簡化成短短一兩個句子的「正確選項」不但不夠完整，也容易造成爭議。讓學生用選取的方式挑出一個正確答案的方式所運用的思維只在「辨認」的層次，而無法展現出「分析」與「論述」的能力。此外，命題者還得花很多精力去設計有「誘答力」但又不合理的選項供學生選擇，更不用談如今的命題技術還得講求各選項的字數必須一致（筆者這一題在命題技術上就不及格了！）。

由於學生只能從列出的選項中選出一個（或多個）較正確的答案，因此我們只能知道學生選對或選錯，但卻無法看出學生作答背後的思維過程。學生能選出正確選項並不保證他們真的懂，甚至對同樣都答對這道題目的學生而言，他們的知識理解程度也不見得是一樣的（蕭憶梅，2009）。比較學生

在第三題選擇題與非選擇題的答題情況就可清楚地看出這一點，學生在這個題目用選擇題型式測驗時都答對了，但用非選擇題型式來測驗時就顯示出他們對歷史的不同理解層次，有的學生僅以某些先入為主的偏見作判斷，未能運用史料進行推論；有的學生則只擷取史料中的訊息，但未能對不同資料進行分析比較，但少數學生的確能針對題意，對不同資料內容進行分析比較。由此可見，選擇題在測驗學生較高層次的歷史思維時的確有其限制，不容易達到測驗效度。相對下，非選擇題較適合用來檢測學生的歷史思維，可清楚看到學生不同層次的思維過程。然而，這代表我們就應該全面在歷史考試中採取非選擇題的方式嗎？

也許這個答案並沒有那麼簡單，因為非選擇題第一個要面對的問題就是評分問題：評量標準該如何制定？閱卷的公平性該如何確保？以試卷的第三題為例（資料二中提到九品官人法的選才標準是在於人才優劣，而非世族高卑，但是為何資料三和資料四中卻認為九品官人法中的上品多是公侯勢族？），學生面對這種開放式的題目時出現五花八門的回答。在沒有所謂「標準答案」的前提下，該如何處理評分的問題？評量標準的制定就是其中第一步，須先釐清這個測驗目標由淺入深的思維層次，這些層次可在進行試卷測試時依據學生的回答情形調整如：層次一、僅以常理推論，未能運用史料；層次二、能擷取史料中的訊息，但未能對不同資料進行分析比較；層次三、能針對

題意，對不同資料內容進行分析比較。每個層次可再依據整份試卷的比例配分做調整，例如若這個題目為8分，層次一可給0~2分，層次二可給3~5分，層次三可給6~8分。每個層次可再做更進一步的區分與定義，以提供閱卷者明確的評分依據。

當評量標準建立後，就需要訓練有經驗的教師成為適任的閱卷員。以香港與英國為例，負責考試的機構會招聘有數年教學經驗的現職中學教師為閱卷員，讓閱卷員依據擬定的評分標試批改考卷，以確認閱卷員對評分標準有一致的理解。每份試卷也會有兩位閱卷員批改，以避免個人主觀意見影響試卷分數（蕭憶梅，2003、香港考試評核局，2007）。當然，無論如何小心謹慎地研擬非選擇題的評分方式，它的公平性是絕對比不上選擇題這種標準化測驗。但是，非選擇題在檢測學生高層次歷史思維時的效度卻遠遠超過選擇題。如果我們一味只講求百分之百的公平性，我們可能會失去真正檢測出學生不同層次歷史思維的可能性。

測驗的「效度」與「公平性」其實同樣重要，只重視公平性而使用選擇題或只重視效度改採非選擇題的兩種極端都不是最好的方法，也許依據測驗目標來決定歷史測驗題型會是比較好的作法。例如，試卷中的第一題較為基礎，主要在檢測學生從史料中提取資訊的能力。此題所牽涉的歷史思維較初階，用選擇題的方式其實就可以測量出來。在這樣的情況下，就沒有必要用簡答題之類的非選擇題方式出題，避免增加閱卷的成

本。但試卷中的第三題較為困難，它牽涉到的歷史思維較複雜，僅用選擇題的方式無法檢測出學生真正的能力（在本文第二部份已舉例分析），這時就可考慮用非選擇題的方式出題。這樣的抉擇並不容易，但唯有真正去思考要用什麼樣的題型才能檢測出測驗目標中的概念或能力時，歷史測驗才能發揮其正面的效益。

伍、結語

隨著歷史教育漸漸重視歷史思維，如今的歷史考試面對巨大的挑戰。雖然現今的歷史測驗題目不斷活化，試圖用史料、情境或圖表來測驗學生的歷史思維，但這些方式都不能解決選擇題題型本身外在型式侷限性所造成的問題。選擇題本身簡單的形式（題幹不能過長、史料不能過多、思考範圍已限制在選項內）很難承載需要運用高層次歷史思維的測驗目標，為了符合選擇題的型式，命題者必需將許多複雜的概念濃縮在一行或兩行的選項內。這些限制常常讓一道命題立意良好的題目變得支離破碎，學生應對的方式通常是用「找線索」或「刪去法」來答題，最後題目本身與其宣稱的測驗目標早已出現很大的距離。從同一道題目以選擇題與非選擇題兩種不同形式來檢測學生歷史思維時，就可證明此點。試卷中的第三題在用選擇題方式出題時，就無法真正檢測出學生高層次的歷史思維，但用非選擇題時就有比較大的空間讓學生表達其思維過程。



許多歷史教育工作者其實早已明白選擇題的缺點，但仍持續努力思考選擇題不同形式的命題方式，其背後的主因就是因為選擇題的優點難以取代，它畢竟是標準化測驗，不但方便閱卷又可確保評分的公平性。而非選擇題雖然擺脫選擇題外在型的限制，但它卻有與生俱來的巨大包袱，就是評分的複雜性。因此即使高中升大學的大學指考中有非選擇題的形式，但它其實是簡答題，仍然具有標準答案，我們還是無法藉由測驗檢視出學生歷史思維的過程。因此，到底該用哪種題型來檢測學生的歷史思維？這是一個兩難。但如文中所述，依據測驗目標來決定歷史測驗題型會是比較折衷的作法，選擇題可用來測試較基礎的歷史思維，如指出史料中的資訊或史料間的異同處，但非選擇題可用來測試較進階的歷史思維，如判斷為何有不同的論述或解釋某個現象的原因。

無論是以哪種測驗題型來命題，最重要的還是要回到測驗目標來思考。並非題目中有出現史料，就代表是在測試學生的「史料證據」概念，有時只是換湯不換藥，改用史料的方式測試學生對某些歷史事實的認識而已。如果命題者關注的僅止於學生對歷史知識的熟練度，改變測驗題型的方式也就沒有太多助益，只是徒增困擾。

測驗本身只是一個工具，主要的目的在檢驗出學生學習的情況，它與教學本身的關係是密不可分的。只有教學目標與測驗目標能相互配合時，測驗才有其意義。若歷史測驗沒有關注到教學，只陷在題目的難易度、

鑑別度等這些統計數字裡，它就失去其促進學生學習的重要價值，反而成為監禁學生學習的牢籠。

參考文獻

- 香港考試評核局（2007）。香港高級程度會考——評閱答卷及分數處理。取自http://www.hkeaa.edu.hk/tc/hkale/Marking_and_Processing_of_Marks/
- 教育部（2005）。「普通高級中學課程暫行綱要」——歷史課程綱要。取自http://www.edu.tw/files/site_content/b0037/8.doc
- 教育部（2010）。國民中小學九年一貫課程綱要社會學習領域。教育部。
- 教育部（2011）。「普通高級中學課程綱要」——歷史課程綱要。取自http://www.edu.tw/files/site_content/B0036/公告用課程綱要_1000527c.doc
- 蔡蔚群（2004）。考試能否檢驗教學的成果。清華歷史教學，15，頁97-118。
- 蕭憶梅（2003）。英國與臺灣歷史科考試之比較。清華歷史教學，14，頁69-81。
- 蕭憶梅（2009）。理解學生歷史思維的重要性。歷史教育，14，頁87-102。
- Dickinson, A. (1978). The New Examinations: Servants or Lipservers of Historical Understanding? In Dickinson, A. and Lee, P. (Eds) *History Teaching and Historical Understanding*. London: Heinemann Educational Book, pp.133-155.