

國內學生學業成就評價研究綜述

鐘媚

華南師範大學課程與教學系

廣州 510631

電郵：zhmei80@163.com

收稿日期：二零零八年三月二日(於六月三十日再修定)

內容

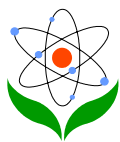
- [摘要](#)
 - [《教育部初中畢業學業考試評價理科科目組》](#)
 - [《新課程背景下高中學生學業評價》課題組](#)
 - [《促進教師發展和學生成長評價研究》課題組](#)
 - [結語](#)
 - [參考文獻](#)
-

摘要

弱化評價的管理性功能，突出評價的教育性功能是新一輪國家課程改革中一個亟待解決的問題。雖然從理論層面上學者們對學生學業成就評價研究進行了大量的探討，但從實踐層面上對這一方面的研究還非常薄弱，直接關注科學學科學業評價研究的更加少。本文選取了三個比較有代表性的學業成就評價研究課題組：（1）《教育部初中畢業學業考試評價理科科目組》，主要關注初中理科試題的命制工作；（2）《新課程背景下高中學生學業評價》課題組，主要關注過程性評價和開放性試題的命制工作；（3）《促進教師發展和學生成長評價研究》課題組，主要關注發展性評價和檔案袋評估的研究工作，對他們的具體工作進行簡要綜述。

《教育部初中畢業學業考試評價理科科目組》

《教育部初中畢業學業考試評價理科科目組》於 2003 年開始對初中理科會考試題的命制和開發進行了廣泛的實踐探索。雖然該課題組在總體理論框架上沒有明確的導向，但在對科學概念知識、科學探究能力和情感態度價值觀方面的考查上提出了很有建設性意見的命題原則和要求（全國初中畢業升學考試綜合理科學科評價課題組，2004；教育部



初中畢業學業考試評價理科科目組，2005)。另外，在命題的總體形式上，無論是對科學概念知識、科學探究能力還是情感態度價值觀的考查，該課題組的關注點都集中在對情境性試題的研究工作上，即考察學生在具體情境中運用所學知識與技能，分析和解決問題的能力，這是很值得理科考試命題工作者借鑒的地方。

(一) 科學概念知識方面

在科學概念知識方面的考察上，該課題組認為命題的原則是：要密切聯繫學生的學習和生活實際，考查學生對科學知識的理解以及應用科學知識解釋現象、分析和解決問題的能力。比如對溶液概念的考察，傳統的做法是直接考定義，現在的做法是考概念的運用。該課題組認為，對於命題基礎知識的命題要避免兩種傾向：一是只測試一堆高度抽象、高度概括的知識，完全脫離生活實際；另一種則相反，測驗只是簡單地重複一下學生已有的生活經驗，沒有超越經驗上升為科學（資料來源：教育部初中畢業學業考試評價理科科目組，2005，頁 30-34）。

例題 1： 以下是溶液的一些說法： ① 無色 ② 混合物 ③ 化合物 ④ 均一 ⑤ 穩定 ⑥ 各部分性質相同，其中正確的是（ ）

A. ① ③ ④ ⑤ B. ② ④ ⑤ ⑥ C. ① ② ④ ⑤ D. ① ③ ⑤ ⑥

答案：B

例題 2（上海理化，2003 年）：將白醋、食用油、食鹽、酒精分別放入水中，不能形成溶液的是（ ）

A. 白醋 B. 食用油 C. 食鹽 D. 酒精

答案：B

例題 3：下列物質中，沒有顏色、沒有氣味、沒有味道的液體是（ ）

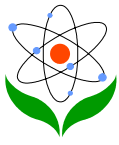
A. 啤酒 B. 食鹽水 C. 純淨水 D. 食醋

答案：B

分析：和例題 1 一樣，例題 2 同樣考察溶液的概念，但卻不像例題 1 那樣對溶液概念的定義直接設問，而是通過設置一個實際的情境，讓學生回答將生活中常見常用的白醋、食鹽、食用油、酒精分別放入水中能否形成溶液。例題 3 則停留在常識水準，不需要教學和考試也可以回答。

(二) 科學探究能力方面

該課題組認為，對科學探究的考查應該從實際問題出發，重點關注學生的探究能力。應該避免將科學探究分解為生搬硬套的知識和按部就班的程式進行考查。在實際教學中，學生應有較完整、自然地體驗科學探究過程的機會，但在大規模考試的情景中，要考查一個完整的探究過程往往是困難的，選擇探究過程的部分內容，對學生探究能力的某一方面進行考查更現實可行。也就是說，為了保證考查的有效性和可操作性，一道試題不必貪多求全，可以只集中考察探究能力的一個或兩個方面。如考查學生提出猜想的能力、設計和實施實驗的能力、表達和交流的能力等。



1. 考查針對科學性問題提出猜想的能力

例題 1：在煮餃子的過程中，可觀察到如下現象：餃子剛放入時，沉在水底，煮了一段時間後，就浮起來。餃子為什麼會浮起來？請提出你的猜想並簡要說明。（資料來源：全國初中畢業升學考試綜合理科學科評價課題組，2004，頁 13）

我想原因可能是：_____

我的根據是：_____

還有可能的原因我認為是：_____

我的根據是：_____

.....

可能的答案是：

- (1) 沸水上下翻滾帶起餃子。
- (2) 水沸騰後餃子表面吸附了一些氣泡，托起餃子。
- (3) 煮熟後的餃子密度小，因此會浮起來。
- (4) 餃子由於溫度升高而體積變大，所受浮力也增大，且大於重力，因此浮起來。
- (5) 餃子內的空氣受熱膨脹，使餃子體積變大，所受浮力也增大，並使浮力大於重力（儘管煮過的餃子因吸水會變重，但浮力增加顯然更大），因此浮起來。
- (6) 隨著餃子內的溫度升高，水蒸氣增加，導致餃子體積變大，所受浮力增大，並使浮力大於重力（儘管煮過的餃子因吸水會變重，但浮力增加顯然更大），因此浮起來。

評分標準：滿分為 6 分，答 (1) 或 (2) 得 1 分；答 (3) 得 3 分，其中若只答餃子煮熟了就會浮起來，而未提到密度變小的，得 1 分；答 (4)、(5) 和 (6) 中的任意一個得 5 分，同時考慮餃子中的空氣和水蒸氣受熱膨脹的得 6 分，其中若沒有考慮煮過的餃子吸水變重而認為重力不變則扣 1 分；能綜合考慮餃子本身和餃子內氣體受熱膨脹並判斷氣體受熱膨脹可能是主要因素的得 6 分；提出其他合理猜想的酌情給分。

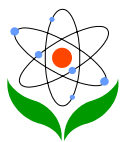
分析：一般來說，學生對煮餃子的過程並不陌生。對於餃子煮後會浮起來的現象，學生可根據經驗和所學的知識，提出許多有一定根據的猜想。評分主要依據的是學生通過猜想所表現出的正確運用科學知識、方法和推理的能力。

2. 考查設計和實施科學實驗的能力

例題 1：（新疆物理，2002 年）幾個同學爬杆比賽時，要比較誰的功率大，分別需要哪些測量儀器和測出哪些物理量？（資料來源：羅星凱等，2004）

答案及評分標準：本題有多種答案，只要合理均得分。幾種可能的答案如下：

- (1) 用體重計測體重，用碼錶測時間，用長杆和卷尺測高度。
- (2) 用體重計測體重，用碼錶測時間。（要求比賽的同學都爬到同一高度）



(3) 用體重計測體重，用長杆和卷尺測高度。(要求比賽的同學在幾條杆上同時開始，同時停止)

分析：要設計一個實驗來測量人爬杆時的功率，首先就要建立一個科學模型，高度抽象、概括地描述人爬杆時的物理狀態。有了正確的模型，爬杆人的體重、爬杆的高度和時間是三個必須考慮的因素。常規的方法是像答案(1)那樣分別去測量三個量，而用控制變數的思想(保持同樣的高度或用同樣的時間)，就可以在時間和高度這兩個量中少測一個(作為試題，這樣的設計應該得分最高)。因此，這樣的設計性實驗試題，雖然並不複雜，但卻考查了學生建立科學模型和正確應用控制變數的方法解決問題的能力。

3. 考查解釋和表達科學探究結果的能力

例題 1：(南通物理，2004 年)小明同學在做完“觀察水的沸騰”實驗後，又進一步研究了沸水自然冷卻過程中溫度隨時間的變化情況，將實驗資料記錄在下表中(資料來源：教育部初中畢業學業考試評價理科科目組，2005，頁 81-82)：

時間 t/min	0	5	10	15	25	35	45	55	65
溫度 t/°C	100	71	55	45	35	28	24	22	22

請在座標圖中通過描點畫出上述溫度隨時間變化的曲線。

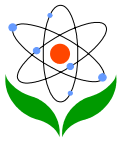
如果有一天你要喝一杯奶茶，可以有兩種方法供你選擇：① 先將滾燙的熱茶冷卻 5min，然後加一匙冷牛奶；② 先將一匙冷牛奶加進滾燙的熱茶中，然後冷卻 5min。你認為方法_____的冷卻效果較好。

分析：一杯熱水放在室溫下，溫度不斷下降，是一個司空見慣的現象。通過探究，從這個司空見慣的現象中發現規律，並把這個規律巧妙地應用到日常生活中，可以讓學生感受到科學探究的獨特魅力。

(三) 情感態度價值觀方面

該課題組認為，情感態度與價值觀是新課程三維目標體系中的重要一維，其重要性和知識與技能、過程與方法至少是等同的。但在科學學科的教育活動中，不能把這方面的東西當成一些條文背誦，而在考試中要求簡單地復述這些條文恰恰會導致這樣的教學。該課題組提倡通過在試題中滲透情感態度與價值觀教育的方式實現考察意圖，比如滲透到具體生動的科學事例與活動中，設計一些開放性的試題等等。該課題組認為，是否完全可以通過考試的方式來檢測這方面的內容，是值得商榷和進一步研究的問題。

例題 1：在今年的氣象日，《海口晚報》刊登了一則氣象資料：“由於空氣污染的影響，地球上陽光越來越少。與二十世紀 50 年代相比，如今到達地表的太陽光平均減少了 10% 左右……”。請你依據光的傳播知識分析，影響地表日照的原因可能是什麼(資料來源：羅星凱等，2005)。



答案：空氣中的污染物（塵埃）將太陽光向各個方向反射。

分析：學生做完這道題後，收穫可能不僅僅是這道試題的正確答案。很多學生都知道大量排放煙塵會造成污染，但沒有想到連光照也會影響。無形中也許學生就意識到防止大氣污染，保護環境是多麼的重要。

《新課程背景下高中學生學業評價》課題組

《新課程背景下高中學生學業評價》課題組於 2003 年開始對學生學業成績評價進行理論研究和實踐探索。該項目研究分為兩個階段，第一階段是 2003-2005 年，主要就高中學業評價的基本理念和框架、模組終結性測驗的定位和方案、過程性評價的理念和方法、開放性試題的命制等進行了探討；第二階段是 2005-2008 主要就高中必修課程學業評價進行探討，主要考察必修模組的水準測試，在命題與考試方面開展實踐研究。這裏主要對該課題組在過程性評價的理念和實施、開放性試題的命制和開發兩個方面進行探討。

（一）過程性評價的理論和實踐研究

1. 過程性評價的理念

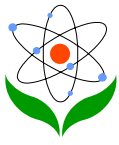
該課題組認為，過程性評價不是對微觀意義上的學習過程的評價，也不是只注重過程而不注重結果的評價，而是對課程實施意義上的學習動機、過程和效果的三位一體的評價（高凌飆，2004）。為此，相對於終結性的或形成性的測驗方式，過程性評價更關注動態的學習表現，評價的內容主要包括以下三個方面（高凌飆，2004；鐘媚，2005）。

（1）學習方式

新課程要求改變被動的、接受式的學習，提倡自主、合作和探究的學習方式，使學生的學習發生實質性的變化。自主學習、合作學習和探究性學習是對學習方式的形式化描述，那麼，學習方式的本質是什麼？

這裏指的學習方式來源於英語“approaches to learning”，具體含義包括：當學生面對學習任務時，他們對學習任務的價值取向和興趣，以及他們對學習本身的認識，導致他們對學習效果產生特定的期望，這種期望驅使學生採用特定的策略和方法完成學習的任務，從而產生與期望相匹配的結果。可見，學生學習前的預備狀態（presage）、學習過程中思考和活動的方式（process）、學習的結果（product），是一個密切相關的系統。簡言之，學習動機 + 學習策略 = 學習方式，學習方式是決定學習品質的重要一環，對學生的學習結果產生重大的影響。根據所涉及的動機和策略的不同，比格斯(Biggs 1987)把學習的方式分為表層式、深層式和成就式等。新課程所宣導的是深層式的學習，即學習者的學習由內在動機或好奇心產生，為滿足興趣、探究意義而自發、主動學習。

（2）反映學生智力發展的過程性成果



終結性測驗可以很好地考察經過一個學習階段後學生對知識技能的理解和掌握，而對反映學生智力發展的過程性成果則無法很好地反映。這就成為過程性評價關注的內容之一，具體來說，包括：①學習策略和方法。在學習過程中，以過程性評價作為交流的平臺，通過學生之間的相互觀察和提醒，可以促進他們不斷自我反思，從而提高學生的學習策略和方法水準。②與學科學習緊密相關而難以通過終結性評價衡量的智力因素，如解決現實科學問題的能力、創新能力等，這是考試測驗不能完全考出來的。③通過學科學習而達成的高層次學習目標，如正確的科學觀等。

(3) 與學習密切相關的非智力因素

學生的學習過程受多種因素的影響，其中非智力因素對學生的學習有十分重要的作用。過程性評價通過考察與學習密切相關的非智力因素，比如對學習任務的意向，學習的習慣，以及透過學習過程反映出來的情感和態度等，給學生以資訊回饋，促進學生進行反思和總結，使其對自身建立更加全面和深入地認識，從而提高學習的效果和品質。

2.過程性評價的實施

如何進行過程性評價？評價的實質在於價值判斷。從操作層面上看，評價的過程包括設計評價方案和工具，利用評價工具收集資料資料，對資料資料進行分析得出結論，對評價結論進行解釋。因此，過程性評價的實施應該包括以下四個工作環節（鐘媚，2005）：

(1) 明確評價的內涵和標準

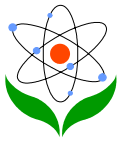
從評價的使用目的及結果來看，過程性評價屬於課堂學習評價，一般由學校自主進行，評價的主體是教師和學生。在評價的過程中，教師雖然也作為評價者參與評價過程，但更多的是作為評價的監控者和組織者，所以評價的主體主要是學生，評價的責任也就相應地落到了學生的身上。因此，評價工作的第一步是使學生理解評價的內涵和功能，明確評價的內容和標準。過去，由於學生長期處於被評價者的地位，對於什麼是評價，評價的目的和作用是什麼，以及如何正確地評價自己和他人等的認識是比較薄弱的，導致在評價過程中不能正確地把握和對待評價的標準。如何使學生提高對過程性評價的認識，使評價的水準由完全的下意識或潛意識水準以及情感水準逐步上升到理智水準，應該成為評價過程中的一項重要內容。當然，教師作為評價的組織者和主體之一更應該積極地提高對評價的認識，在教學的過程中逐步引導學生學會評價。

(2) 設計評價方案和工具

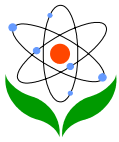
從評價的內涵中可以看出，過程性評價是基於人們對學習品質的認識提出來的，是一種較具體方法更為上位的評價方式，因此，不能將過程性評價與某種特定的評價方法甚至評價工具等同起來，應該根據學校和學生的實際情況，針對具體問題，選擇合適的評價工具。該課題組開發了評價量表作為過程性評價的工具（見表 2-1）。

表 2-1：高中語文過程性學習評價用表（資料來源：鐘媚，2005）

模組_____高____級____班 學生_____學號_____200__年__月



評價項目	評價內容	評價等級			
		自評		組評	
1. 情感、態度、價值觀	①熱愛祖國的語言文字，積極提高運用欣賞能力				
	②維護祖國的語言文字，抵制不良語言現象				
	③關注生活中的語言現象，並能分析判斷、取捨				
	④按質按量完成學習任務				
	⑤學習中表現得自信、勤奮、堅毅，追求卓越				
2. 過程與方法	①獨立思考，善於發現問題，有自己獨到的見解				
	②合作意識強，勇於表達，專注傾聽，樂於交流				
	③注重體驗美，創造美，陶冶性情，涵養心靈				
	④善於計畫、比較、反思、總結、調整				
	⑤勤於整理、辨析、積累，以各種方式充實自己				
3. 學習成果	①閱讀廣泛，讀高文化品味著作，讀完整著作				
	②表達交流語言規範、清晰、誠懇、有感染力				
	③積累豐富、完整、準確，靈活運用				
	④書寫工整認真，整潔美觀				
	⑤作品豐富、完整，有創意；筆記有條理，清晰				
特長及成果展示列舉	背誦的篇章				
	閱讀的書籍				
	發表及其它				
個人小結					



小組 評議					
自評等級	本人簽名	組評等級	組長簽名	教師總 評	教師簽名

表格填寫說明：

1.過程性學習評價的方式為等級制，分四個等級：A等（優秀）、B等（良好）、C等（合格）、D等（仍需努力）。

2.各“評價欄目”的等級確認辦法：先給5個具體分項評定等級；再將等級轉換為分值：A等為3，B等為2，C等為1，D等為0分；最後把5個分數相加，得12—15分的為A等，8—11分為B等，4—7分為C等，0—3分為D等。

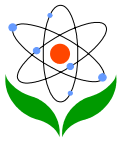
上述高中語文過程性學習評價用表的優點在於：評價所指向的內容比較全面，涉及情感態度與價值觀、過程與方法、學習成果等領域；評價的主體多元，有學生自評、小組評、教師評；評價指標體系的分解比較詳細、科學、可行；各指標的權重比較清晰；量化評價的等級的計算比較清晰；評價含有成果展示等評定的證明材料，使得評價的結果比較可信。

（3）解釋和利用反映學習品質的結果

在學年或一個學習階段結束之後，應如何報告、解釋、利用過程性評價的結果？有的學校建議把“過程性評價的結果賦予權重，折算成一個分數，然後與終結性評價的結果相加，作為最終學習結果”。那麼，過程性評價與終結性評價要不要整合？過程性評價屬於個體內差異評價，評價的目的並不是對學生的學習下一個終結性的結論，而是促進學生的學習與發展；評價的內容主要不是學生最後達到的水準，而是學生的進步情況。教師和學生應該通過對收集資料的分析，讓學生瞭解自己的進步和不足，在此基礎上提出建議，使學生明確將來繼續努力的方向。所以，簡單地把過程性評價的情況折算為一個分數加到考試成績裏面去是沒有意義的。過程性評價的結果與終結性評價的結果應該分別給予呈現報告。

（4）反思和改進評價方案

整個評價方案實施結束以後，還需要對於評價方案進行再評價，也稱“元評價”，即“在評價的過程中，為檢討評價方案，實施過程與結果，藉以總結成功的經驗和糾正評價工作之不足，而對正在進行或已完成的評價進行價值判斷。”元評價的關鍵是確定元評價的標準問題，即一份好的方案的標準是什麼。我們可以以下述問題作為準則考察評價的方案：



①有效。這個方案是為瞭解決什麼問題，是否最合理地解決了問題，在多大程度上達到了我們的目標？

②可行。方案的要求是否超出了學校現有的條件和能力？評價的工具是否繁瑣，是否給教師和學生帶來很大的負擔？

③可信。評價的標準是否恰切？評價工具收集的資訊是否準確？

通過對這三方面的考察，學校和教師可以找出評價方案的不足，改進評價的方式和工具，更好地完成下一步的評價工作。過程性評價是一個動態多元的，而不是一個僵化單一的評價方式，因此，即使在實施的過程中，學校和教師仍然可以根據變化的情況及時修正評價方案。

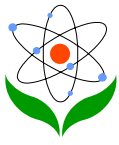
(二) 高中必修課學業水準測試

新課程強調學生的模組學分由各高中學校自主授予，這一措施大大增加了高中學校的辦學自主權和靈活性，有利於各高中學校根據自己的實際進行教學，全面落實素質教育方針，形成自己的風格和特色。然而，由於學校條件和辦學理念上的差異，各高中學校的教學品質也有一定的差別。需要通過適當的檢測，對各高中學校的教學品質進行監控，保證每一間高中學校的教學品質以及每一位高中畢業學生的學業都達到國家規定的基本要求。為此，教育部華南師範大學基礎教育課程研究中心，從 2006 年開始，系統組織了高中語文、數學、英語、物理、化學、生物、政治、歷史、地理九學科的必修課學業水準測試。為了避免過去考試命題過於倚重封閉性試題的弊端，該課題組對開放性試題的命制方法進行了有益的探索。

1. 開放性試題的命制理念

過去的學業測評主要採用封閉性試題，該課題組認為，適當採用開放性試題在很大程度上彌補了封閉性試題的種種不足，特別在考查學生思維的靈活性和廣泛性，考查學生的實踐能力和創新意識，以及情感、態度、價值觀等方面有著封閉性試題所無法取代的優點。開放性試題更符合新課程發展性評價的理念，在新課程的模組學業評價中應提倡適量使用開放性試題（高凌飆，吳維寧，2004）。

該課題組認為，開放性在編制上需要注意兩方面的問題：①試題是否有明確的檢測目標和檢測效度；②試題是否違背學科的基本規律。為了避免開放性試題評分主觀隨意性的弊端，該課題組集中研究了 SOLO 分類法（Structure of the Observed Learning Outcome）在命題中的運用。SOLO 分類評價法的基本理念源于皮亞傑的認知發展階段論。皮亞傑的認知發展階段論指出，兒童在成長的過程中認知的發展是有階段性的，不同的階段之間的認知水準有質的區別。比格斯(Biggs 1982)和他的同事通過研究發現，人的認知不僅在總體上具有階段性的特點，在對具體知識的認知過程中，也具有階段性的特徵。人在學習新知識過程中表現出來的思維階段是可以觀察到的，因此稱為“可觀察的學習成果結構”。也就是說，學生在具體知識的學習過程中，都要經歷一個從量變到質變的過程，每發生一次躍變，學生在對於這一種知識的認知就進入更高一級的階段，可以根據學生在回答問題時的表現來判斷他所處的思維發展階段，進而給予合理的評分。SOLO 分類法將學生學習的結果由低到高分為五個不同的層次，即前結構（prestructural）、單點結構（unistructural）、多點結構（multistructural）、關聯結構（relational）、拓展抽象結構（extended abstract）等。這五種結構的基本含義如下。



- (1) 前結構：沒有形成對問題的理解，回答問題邏輯混亂，或同義反復；
- (2) 單點結構：回答問題時，只能聯繫單一事件，找到一個線索就立即跳到結論上去；
- (3) 多點結構：回答問題時，能聯繫多個孤立事件，但未形成相關問題的知識網路；
- (4) 關聯結構：回答問題時，能夠聯想多個事件，並能將多個事件聯繫起來；
- (5) 拓展抽象結構：回答問題時，能夠進行抽象概括，結論具有開放性，使得問題本身的意義得到拓展。

SOLO 的五種層次代表了學生對於某項具體知識的掌握水準，從學生對某個問題的回答中，教師可以對照上述標準就學生對該項知識內容的掌握情況做出判斷。因此，這種評價方式可以幫助教師進行教學診斷，同時，也可以向學生提供有用的反饋，所以 SOLO 分類法可以用於形成性的學生學業評價；另一方面，如果將上述五個層次賦予不同的等級分數，那麼學生對問題回答的品質就可以被量化，量化的分數可以作為終結性評價的依據。以下是依據 SOLO 的基本思想進行等級評分的歷史開放性試題和物理開放性試題，從中我們可以瞭解基於 SOLO 的評分標準編制的一般方法。

2. 開放性試題的命制實例

(1) 歷史開放性試題

題目：請結合英國資產階級革命的相關史實，談談你對英國議會制度的理解（資料來源：黃牧航，2004）。

本題為一道問答題，考查學生思維能力的種類包括再現歷史知識、概括歷史知識、分析歷史事物的本質、闡述歷史發展的規律等等。而根據 SOLO 分類法，可以把學生的回答分為以下五種情況：

①前結構的回答。這種回答只出現了一個簡單的判斷，我們不知道學生得出這種判斷的根據何在；或者只提供了一個史實，我們不知道這個史實說明瞭什麼問題。如：

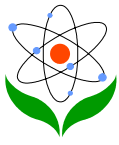
答（1）：英國的議會制度對英國歷史產生了巨大的作用。

答（2）：1649 年，議會把查理一世送上了斷頭臺。

②單點結構的回答。這種回答只依據了一個方面的史實，說明學生還沒有充分理解歷史發展的複雜性。如：

答：英國資產階級革命期間，議會組成了議會軍，打敗了王黨軍隊，使立憲立憲制能夠在英國確立。

③多點結構的回答。這種回答能夠依據兩個以上的史實，但是未能建立起這些史實之間的聯繫。如：



答：1640年，查理一世為籌措軍費而被迫重新召開議會，而議會卻以此為契機要求限制王權。當兩者的矛盾不可調和時就只能兵戎相見了。1645年，議會軍最終擊敗了王軍，並於1649年把國王送上了斷頭臺。

④關聯結構的回答。這種回答不但能夠提供多個史實，而且能夠在這些史實之間建立起某種聯繫。如：

答：英國資產階級革命其實是國王與議會之間的矛盾鬥爭引起的。1640年，查理一世為籌措軍費而被迫重新召開議會，而議會卻以此為契機要求限制王權。當兩者的矛盾不可調和時就只能兵戎相見了。1645年，議會軍最終擊敗了王軍，並於1649年把國王送上了斷頭臺。但是，議會與國王之間的矛盾並未根本解決。1660年，因為議會內部矛盾不斷，查理二世借機復辟。斯圖亞特王朝復辟後，國王與議會的矛盾更加尖銳，最後，議會不得不拋棄查理二世，把荷蘭的執政威廉三世迎來英國當國王，使立憲君主制最終在英國確立，國王與議會的矛盾也最終得以解決。

⑤抽象拓展結構的回答。這種回答不但能夠提供多個史實，而且能夠把對問題的認識上升到抽象的理論認識層次，能夠從理論上分析為什麼會出現這種現象，以及這種現象的發展規律。如：

答：英國的議會起源於中世紀，議會與國王之間的矛盾是一直存在的。在中世紀，兩者的矛盾主要體現在王權與貴族特權的紛爭。到了近代，隨著資產階級勢力的發展，兩者的矛盾則體現在新興的資產階級對君主專制統治的不滿，要求分享政治的權利。……（省略具體的史實）“光榮革命”後，資產階級立憲君主制最終在英國確立起來，表面上這種制度的特點是國王與議會並存，而實質上是議會的權力日益超過國王，因此，資產階級終於能夠通過議會來實現對國家政權的控制。

(2) 物理開放性試題

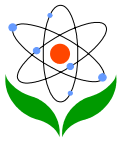
教師任務：將一氣墊導軌傾斜放置，用尺規標明導軌上各點的位置，讓滑塊從導軌的高端滑下。不斷重複這一過程（資料來源：高凌飆，2005）。

學生任務：觀察上述過程並進行你認為必要的測量（可以利用自己的手錶和導軌上的尺規），記錄下結果，利用觀察和測量的結果，對上述過程進行描述。應清楚地寫出必要的文字說明、推導論證、公式計算、資料結論、圖線表格等等。

評分細則：

本題無參考解答，教師可參照下麵標準給學生評分：

- 0分：沒有回答，或給出的回答言不及義、不對題。
- 1分：只有定性的描述，基本上能說清滑塊的運動過程。
- 2分：有定性的描述和定量的測量。其中定量的測量只是測出幾個時間及距離而沒有做進一步的分析整理，也沒有結論；或測得的資料與實際的差距很大。
- 3分：有定性的描述和定量的測量。其中定量的測量包括對滑塊運動到不同位置時，滑塊與原點的距離及運動所經過的時間，並能用這一資料計算出滑塊運動的平均速度或即時速度這兩個量中的一個。測得的資料基本上合理。
- 4分：情況同上，但能將滑塊運動的平均速度和即時速度兩個量都求出來。



- 5 分：在前面的基礎上，整個報告的敘述，包括目的、原理、過程、資料記錄和整理、最後的結論等，條理清楚，結論明確。
- 7 分：在前面的基礎上，進一步求出滑塊運動的加速度。
- 9 分：在前面的基礎上，能利用所掌握的多種數學工具，包括圖表、圖線等等，將實驗觀察的結果非常清楚地表達出來。
- 10 分：在前面的基礎上，能從實驗事實出發對滑塊的運動以至普遍情況下的勻加速運動，或其他任何問題進行討論，且言之成理。

其他形式的答案，可參照上述等級給分。

《促進教師發展和學生成長評價研究》課題組

《促進教師發展和學生成長評價研究》課題組於 2002 年開始對發展性評價的理念和實施進行了廣泛的探討。課題組認為，發展性評價區別於選拔性評價和水準性評價，指的是在發展的整個過程中進行的，旨在促進被評價者不斷發展的評價（董奇，趙德成，2003）。發展性評價的目的是為了更好地促進學生的成長，促進教師教育教學水準的提高，促進學校的發展。發展性評價主張評價主體、評價內容的多元化和評價方法的多樣化，提倡形成性評價和終結性評價的有機結合。在評價方式和工具的開發方面，該課題組著重探討了質的評價方式—檔案袋評估的實施。

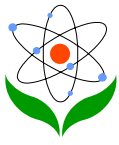
（一）檔案袋評估的定義

檔案袋評估（portfolio assessment），又稱作“成長記錄袋評估”。該課題組認為，檔案袋評估就是指“學生作品的有意收集，以反映學生在特定領域的努力、進步或成就。它必須包括內容選擇過程中的學生參與、選擇的指南、評分的標準，以及學生自我反思的證據”（趙德成，2003a）。檔案袋評估之所以受到重視是因其具有呈現學生學習進步情況和學習過程的潛能，以及重視學生對學習的自我反省能力，是一種能夠適應學生個別差異的評估方式。綜合而言，檔案袋評估以促進學生發展為根本目的，強調個別差異的學習和成果的呈現，而且著重教學與評估交互進行的過程。檔案袋評估具有以下特點：（1）檔案袋評估的基本成分是學生作品；（2）學生作品的收集是有目的、有計劃的，而不是隨機的；（3）檔案袋評估要給學生提供發表意見和對作品進行反思的機會；（4）檔案袋評估不是在整個教學結束後才進行，其評估與教學是同步交互進行的。

（二）檔案袋評估的類型

1. 展示型檔案袋

展示型檔案袋是學生工作最佳成果的搜集，是教室中最常實施的檔案袋評估類型之一。展示型檔案袋的主要使用者是學生自己，其主要目的是促進學生對自己作品的反思。展示型檔案袋中搜集的事物包羅萬象，包括手工作品、作業、照片、獎狀、紀念品、論文、短評、錄影等。對於每一項放入展示型檔案袋中的作品，學生應該討論將其放入檔案袋的原因，下列問題可供他們參考：（1）為什麼我會選擇這個作品放入檔案；（2）這個作品是如何顯示我的學習和進步；（3）我是怎樣做使得這個作品成為“最優秀作品”等等。



2. 文件型檔案袋

這類型的檔案是學生學習歷程的觀察及記錄，主要是描述學生一段時間內的進步以及教師的期望。文件型檔案袋的內容可以包括草稿、頭腦風暴活動的結果，也可以放入學習歷程的任何作業、觀察記錄、評估清單、討論記錄等。要利用文件型檔案袋有效評估學生的進步情況，必須依據每一項學習內容選擇數個作品作為其進步的證明，因此，選擇的作品要包括三部分：（1）原始的草稿；（2）修改過的草稿；（3）最後的草稿。與展示型檔案袋相比，文件型檔案的限制在於耗費時間長，資料太多，不易評估。針對耗時的問題，教師可以在平時有目的地記錄。至於資料太多的問題，教師應評估檔案可以評估學生的進步的資料，無需對所有資料加以評估。

3. 過程型檔案袋

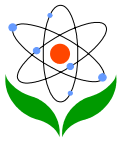
過程型檔案袋記錄了學生在一學年當中對不同課題內容的學習過程，主要包括學習歷程與自我反思兩個方面：第一方面包括作品從萌芽到最終定型的過程，第二方面包括學生對作品的反思過程。自我評價與反思是學生參與檔案袋評估的最高水準。要想使學生的自我評價與自我反思起到預期的作用，教師不能只是簡單地提出要求，要給學生以必要的指導，讓學生知道評價什麼、反思什麼，以及怎麼評價和反思。給學生設計一個用於自我評價或自我反思的表格，是一種值得提倡的做法。表格的欄目設置應該鼓勵學生反思他們為什麼選擇這些內容，以及他們所選擇內容的突出特徵是什麼。

4. 評價型檔案袋

與其他類型的檔案不同，評價型檔案是在預設的要求下搜集資料，並在預設的標準下評估學生，其目的是收集事實以系統評價學生的學習，並將結果報告給其他人。評價型檔案袋試圖通過簡化評價任務、統一評價要求以及指導檔案袋裏面應該放什麼材料來保證評分過程的信度，評價型檔案可以應用於區域性的大規模評價活動。評價型檔案的優勢在於有比較可靠的評分標準，缺點在於：缺乏真實性，資料搜集方向的選擇權在教師而非學生，其主要目的是評估學生在特定工作上的表現，減低了教學目的中所要求的真實性。

（三）檔案袋評估優勢與局限性

該課題組認為，一方面檔案袋評估具有許多傳統的評估方式不具有的優勢，但同時也存在一定的局限性（徐芬，趙德成，2001）。檔案袋記錄了學生成長的“故事”，是評價其最終發展水準、努力、反省和進步的理想方式。相對於紙筆測驗和其他快照式評價而言，它能向教師、家長和學生本人提供豐富的內容，不僅能夠反映學生知道些什麼和能做些什麼，而且能夠反映學生的問題解決能力和思維能力、運用策略和程式性技能的能力，以及建構知識的能力，為評價者提供很多關於學生的學習和發展的重要資訊。由於檔案袋評估並不是局限於教學之後進行，因此可以有效地促進教學與評價的有機結合，幫助教師及時、準確地獲得有關學生的學習資訊，為學生提供適合其特點的教學。另一方面，源於檔案袋評估本身的特性以及我國的教育背景，也帶來很多問題，有待於我們



進一步研究：(1) 檔案袋評估加重教師和學生負擔的問題；(2) 檔案袋評估的信度和效度問題。

結語

學業成就評價是教育評價、課程評價、教學評價、學生評價的重要組成部分，是在系統收集學生課業學習情況的客觀事實材料基礎上的價值判斷過程（吳維寧，2007）。學業成就評價的方法，既包括以紙筆測驗為主要方式的量的方法，也包括以觀察、談話、收集各種反映學生學習過程與結果的作品在內的質的方法。從國內三個主要課題組對學生學業成就評價的研究成果來看，在量的方法和質的方法方面都取得了一定的進展，但仍然存在很多問題。

首先，在量的的評價研究方面，雖然現有的研究根據課程標準的教學要求，提出了與之相匹配的命題原則，但是如何細化和具體化這些原則使之更具有系統性，建立規範的評價理論依據和基本框架，仍然需要進一步的思考。例如，如何界定科學概念理解的層次水準，科學探究能力發展的層次水準。在紙筆測驗中，如何衡量情感態度價值觀的考察限度等等。

其次，在開放性試題的命制方面，現有的研究已經意識到評分標準的效度和信度問題，嘗試引用教育評價領域的分類理論克服開放性試題的主觀隨意性弊端，但是如何結合各個學科的特性，選擇與之相適切的評價理論和方法，還需要研究者在實踐中不斷摸索。例如，SOLO 分類評價理論適合哪些學科的試題命制？如何借助於新一代的測量理論和技術促進開放性試題的發展？

最後，在質的評價研究方面，實現評價與教學的有效結合，開發具有可行性和操作性的評價工具，避免質的評價方法異化成管理和束縛學生的手段，是未來研究的著力點。值得思考的問題是，現代教育評價理念提倡將各種非正式評價納入評價的範圍，如課堂提問，這樣的做法是否會造成評價概念的泛化？如何結合我國的教育背景開展質的評價？

參考文獻

Biggs, J. B., & Collis, K. F (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press.

Biggs, J. B. (1987). *Student Approaches to Learning and Studying*. Hawthorn, Vic.: Australian Council for Educational Research.

董奇，趙德成（2003）。發展性教育評價的理論與實踐。中國教育學刊，(8):18-21。

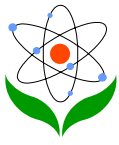
高凌飆（2004）。關於過程性評價的思考。課程·教材·教法，(10):15-19。

高凌飆（2005）。探索和思考：“新課程背景下高中學生學業評價”專案階段總結。
<http://xyji.cersp.com/XMYJ/XGZL/PJLL/200509/19.html>

高凌飆,吳維寧(2004)。開放性試題如何評分？——介紹兩種質性評分方法。學科教育,(8):1-6。

高凌飆，吳維甯，黃牧航（2006）。開放性試題的編制與評分。人民教育，（1）：36-38。

黃牧航（2004）。SOLO 分類評價理論與高中歷史試題的命制。歷史教學，（12）：58-63。



教育部初中畢業學業考試評價理科科目組.中考命題指導（理科）。江蘇教育出版社。

教育部華南師範大學基礎教育課程研究中心（2004）。高中學生學業評價研究資料彙編。

羅星凱，劉小兵，曾平飛，趙光平，薛躍規，黃都（2004）。新課程：“雙基”和“探究”如何考——關於科學學科中考走向的研究與思考。人民教育,(9):33-38。

羅星凱，曾平飛，趙光平，劉小兵，薛躍規，吳德漢，唐力（2005）。理科中考：從 2004 看 2005。

全國初中畢業升學考試綜合理科學科評價課題組（2004）。2004 年中考命題指導叢書（綜合理科）。江蘇教育出版社。

吳維寧(2007)。理科教師學業評價觀研究。華南師範大學博士學位論文（未出版）。

徐芬，趙德成（2001）。檔案袋評價在中小學教育中的應用。教育研究與實驗，（4）：50-54。

趙德成（2003a）。成長記錄袋的創建與使用。天津師範大學學報，（4）：38-42。

趙德成（2003b）。成長記錄袋在大規模、高利害評價中的應用。教育理論與實踐,(8):26-29。

趙德成，徐芬（2002）。成長記錄袋應用的反思與改進。語文建設,(9):43-44。

鐘媚（2005）。過程性評價：概念、範圍與實施。當代教育科學，（14）：44-47。