

一、選擇題(每題 2 分，共 25 題)

- 1、 「實作評量」的設計原則在強調測驗與真實生活情境的結合，例如：實際的科學問題的解決能力的評量、等，因此此類測驗問題的答案，並不一定有唯一的標準答案，所以要檢測此類評量的信度，最好利用以下的何種指標？
 - (1) KR-20
 - (2) Cronbach alpha
 - (3) 概化信度係數
 - (4) 重測信度係數
 - (5) 折半信度係數

- 2、 測驗 A 的常模資料為平均數 56，標準差為 4。測驗 A 另有一套複本測驗為測驗 B；測驗 B 的常模資料為平均數 72，標準差為 6。若甲在測驗 A 的得分為 60，而乙在測驗 B 的得分為 78，丙在測驗 A 的得分為 70，以下何者較正確？
 - (1) 甲比乙的分數高
 - (2) 甲比乙的分數低
 - (3) 乙比丙的分數高
 - (4) 甲與乙的分數相等
 - (5) 乙與丙的分數相等

- 3、 某一研究機構檢驗其所自行研發的基本學力測驗的模擬試題的施測結果與實際舉行的學力測驗的測驗成績的一致程度，這樣的檢驗程序可以被視為是表現該模擬試題的哪一種測驗品質指標？
 - (1) 表面效度
 - (2) 構念效度
 - (3) 內容效度
 - (4) 信度係數
 - (5) 效標關聯效度

- 4、 若研究者想要評鑑整份測驗題本的品質，對於以下的檢定指標的描述，相對而言，何者的描述比較正確？
 - (1) KR-20 係數是檢定題本的穩定度(stability)
 - (2) Alpha 係數是檢定題本的恆等性(equivalence)
 - (3) 重測(test-retest)係數是檢定題本的穩定度(stability)
 - (4) 複本係數是檢定題本的內部一致性(consistency)
 - (5) 折半係數是檢定題本的恆等性(equivalence)

5、以下的各種測驗的相關因子中，相對之下，何者最不可能影響測驗工具的信度係數大小？

- (1) 測驗的題數 (2) 測驗常模的種類 (3) 測驗計分的客觀性
(4) 測驗分數的分散情形 (5) 計算信度係數的方法

6、下表為某次測驗的學生答題狀況，若分別以全部人數的前後 20% 為高、低分組的分組依據，則以下關於各題的鑑別度的描述，何者正確？

	第一題	第二題	第三題	第四題	第五題
學生一	答對	答對	答對	答對	答對
學生二	答對	答對	答對	答對	答錯
學生三	答對	答對	答對	答錯	答錯
學生四	答對	答對	答錯	答錯	答錯
學生五	答對	答錯	答錯	答錯	答錯

- (1) 第一題為 0.1 (2) 第二題為 0.2 (3) 第三題為 0.6
(4) 第四題為 0.4 (5) 第五題為 1.0

7、某研究者為了評定一份測驗工具的品質，若該研究者選定以庫李 (Kuder-Richardson) 係數的大小為指標，則這個評定指標主要是為了要確認該工具的哪一項特質？

- (1) 一致性 (2) 真實性 (3) 可信性 (4) 有效性 (5) 經濟性

8、下列的測驗中，何者是適性化(adaptive)測驗？

甲：基本學力測驗

乙：Raven's Coloured Progressive Matrices (CPM)

丙：Wechsler Intelligence Scale for Children (WISC-III)

- (1) 甲 (2) 乙 (3) 丙 (4) 甲、乙 (5) 甲、乙、丙

9、以下對於 Cronbach alpha 係數用於檢定測驗工具品質時的描述，相對而言，何者的描述比較正確？

- (1) 檢定題本間的一致性 (2) 檢定題本間的恆等性
(3) 檢定評分者的穩定度 (4) 檢定試題間的一致性
(5) 檢定試題間的有效性

10、請依照成就測驗與性向測驗的基本設計原理及邏輯(logic)，仔細比較以下的描述後，選出相對最好的組合。

甲：成就測驗強調學習成果的檢測 乙：性向測驗強調未來潛能的檢測

丙：成就測驗強調教學成果的檢測 丁：兩者的最大差別受測對象的不同

- (1) 甲、乙 (2) 丙、丁 (3) 乙、丙
(4) 甲、乙、丙 (5) 甲、乙、丙、丁

下表是根據某文獻所摘錄的多特質多方法(multitrait-multimethod)分析矩陣，試根據此表回答題 11 至題 15：

	Trait	Method 1			Method 2			Method 3		
		A	B	C	A	B	C	A	B	C
Method 1										
	A	.95								
	B	.28	.86							
	C	.58	.29	.92						
Method 2										
	A	.76	.32	.57	.96					
	B	.25	.90	.40	.39	.87				
	C	.52	.21	.86	.55	.16	.84			
Method 3										
	A	.73	.10	.43	.64	.17	.27	.48		
	B	.15	.68	.17	.26	.67	.19	.25	.41	
	C	.35	.19	.52	.38	.17	.56	.41	.30	.58

N= 213

11、下列哪個數值表示信度 (reliability)？

- (1) .66 (2) .67 (3) .86 (4) .90

12、下列哪個數值表示區別性效度(discriminant validity)？

- (1) .55 (2) .56 (3) .67 (4) .96

13、下列哪個數值表示輻合性效度(convergent validity)？

- (1) .41 (2) .57 (3) .84 (4) .90

14、下列對於多特質多方法的敘述，何者最為正確？(1) 特質(trait)的數目不得少於三個 (2) 多特質多方法的信度係數是屬於內部一致性信度 (3) 它是一種測驗效度的分析方法 (4) 區別性效度數值應該比輻合性效度數值較高

15、假設某研究者實證研究獲得上述的多特質多方法分析矩陣，則有關此研究者的結論，何者最為正確？(1) 由於每個方法要測的特質(trait)數目太多，可能不具任何解釋推論 (2) 受試樣本數不足，其可信度很難接受 (3) 此測驗工具的複本信度 (alternate-forms reliability) 相當高 (4) 此測驗工具的構念效度(construct validity)應該可接受

- 16、下列何者不是影響試題偏差(differential item functioning)之原因？
（1）試題內容僅測量到一個重要概念 （2）試題內容受文化背景因素的影響
（3）試題內容受語言因素的影響 （4）試題內容具有意識型態的偏向
- 17、已知某一標準化測驗其總變異量為 160，誤差變異量為 10，獨特因素變異量為 30，則該次測驗效度的估計為多少？
（1） 0.94 （2） 0.81 （3） 0.75 （4） 0.60
- 18、關於「常模參照測驗」與「標準參照測驗」的敘述，下列何者正確？
（1）標準參照測驗所包含的學習結果較廣
（2）標準參照測驗常以標準分數解釋測驗結果
（3）常模參照測驗強調試題的鑑別力
（4）常模參照測驗目標的陳述較為具體詳細
- 19、根據下列描述，何者不是建置題庫的優點？
（1）改進測驗的品質，並節省編制測驗所花的時間
（2）連結不同測驗間的等化分數，以建立評量量尺
（3）使得所欲測驗的目標，都有適當題數的試題來測量
（4）使得測驗編製者能隨心所欲地編製符合各種目標的測驗
- 20、下列何者不是進行測驗分數等化時，所需符合的條件？
（1）相等性(equity) （2）對稱性(symmetry)
（3）團體不變性(invariance) （4）局部獨立性(local independence)
- 21、今有甲乙兩測驗，已知甲測驗平均分數為 72 分，標準差為 10，乙測驗平均分數為 77 分，標準差為 9，請問經由線性等化轉換過後，甲測驗的 85 分代表乙測驗的多少分？
（1） 85.5 （2） 87.3 （3） 88.7 （4） 90.0
- 22、有一個成就測驗的測量標準誤是 5 分，某生在該測驗之分數為 80 分，請問在 95%之信心水準下，此生的真正分數區間 X 之信賴為何？
（1） $70.2 \leq X \leq 89.8$ （2） $75 \leq X \leq 85$ （3） $65 \leq X \leq 95$ （4） $68.4 \leq X \leq 91.6$
- 23、下列關於評量的敘述何者有誤？
（1）總結性評量可用來評定教學目標是否達成
（2）診斷性評量關心的是學生的起點行為表現
（3）形成性評量提供修正教學的空間
（4）標準參照評量是以事前決定之標準為評鑑之依據

24、依試題反應理論，試題特徵曲線(item characteristic curve)的左上漸近線為能力值極高時，仍無達達到 100%的答對機率稱為什麼？

- (1) 鑑別度參數 (2) 難度參數 (3) 猜測參數 (4) 天花板參數

25、以下何者不是常模之功用？

- (1) 顯示測驗分數在團體中的相對位置
(2) 使用上比較方便，可避免零或負數
(3) 提供比較不同測驗的分數
(4) 常模是解釋測驗分數的依據

二、閱讀選擇題(每題 2 分，共 20 題)

請先閱讀以下的文章，然後根據文章的內容與相關的測驗與評量概念，回答下面的選擇題。

實作評量的本質

實作評量的本質是要求學生投入專業的探究，以創造在他們生活中有價值的知識，而非只是證明他們在學業方面的成就，評量結果的重點不在測驗分數或繳交作業，而是專業領域知識的表現。因此，實作評量一開始在教育上的應用又稱之為表現本位評量。「實作」指的是執行或經歷一個工作，並完成工作。因此，一般以為實作評量必定是要學生實際操作，其實，實際操作只是學生學習的表現形式之一，也並非動手操作就是實作評量的形式。如果實際操作只是簡單的肢體操作，或知識的複製，沒有知識和技巧自我製造性的應用，就沒有學生個人「表現」的意義。實作的精熟通常可表現在完成一篇散文、一項設計工作或同儕和社會大眾有興趣的方案上。實際操作，以及各種形式的表現方式，如學生筆記、電腦虛擬形式、自由申論題與新形式的選擇題，若能測得學生應用知識來表現精熟的能力，不論是評量知識的組織、解題時的策略或監控表現的能力，都可稱之為實作評量或表現本位評量。

雖然實作評量至今未有清楚的定義，各類文獻中的討論，仍提供一些判斷的基本要素：(1) 實作的表現：學生要應用出他們所習得的知識和技巧，表現出認知複雜度；(2) 真實的情境：實作評量是在實際的情境中或模擬的評量作業中完成一個特定的工作，呈現給學生在真實生活中需要面臨與解決的情境，有時候也許是在視聽媒體終端機前舉行評量；(3) 結構的特質：為了提供適度的挑戰與真實性，評量的題目須有一些模糊不清，學生的解題才能有彈性，方式及結果才能多元。(4) 過程性：問題解決的表現不僅限於解題的結果，更在解題的過程，兩者都是評量的目標和媒介；(5) 彈性的解題時間：要求學生呈現個別思考過程，彈性的時間以符合個別差異就是必要的條件。傑出的表現通常無法在一個十分特定的時間內完成；(6) 社會化互動下的結果：思考往往經過激盪後，更多元和深

入；小組的互動更能讓學生發揮思考的廣度與深度；(7) 多向度的評分系統：要求學生表現多元的思考過程與結果，評分的系統就不能單一化，學生所完成的評量項目或結果，得依特定的標準和過程作觀察及評分。以下就將實作評量的這些特點，分別討論。

實作的表現

「實作」強調的是對能力評量的真實性與代表性，及在有意義或真實的情境中學習與評量的重要性；「表現」強調的是對智力、知識與技巧挑戰的必需性，及掌握學習內容與過程的學習結果的必要性。這類的評量項目強調實作的複雜度和知識表現的深度，而非題目的複雜或答案的簡單與否，很複雜的題目未必就是實作評量。實作的特質主要有：(1) 實行並完成一個任務和過程；(2) 表現知識的能力；(3) 是學科的應用，而非只是學習該學科；(4) 是下一個階層的學習準備度的指標，注重的是遷移的能力。

真實性和直接性

在另類評量方式中，真實性評量和實作評量強調的都是真實性和直接性，因此，兩者被視為是可互換的名詞。真實性包含測驗情境與要求標準的真實性，追求測驗情境的真實性，目的在提供給學生一個更有意義的教育經驗，以促成學習和技能的發展。也可對一個好的表現所需的要項和標準做深入的瞭解。但要求測驗情境的設計完全符合真實生活，其實是不大可能的。

強調真實情境並非有意將學習分為學校內與學校圍牆外，事實上，尋找一些學校中發生的真實問題並不如想像中的困難，有些問題在社會上發生，也會在校園中重演，只是大小版本不同。而重點是在使學生從對問題的主控權與對有意義、有興趣的學習探索過程中，尋求問題的解決之道。

豐富的脈絡化真實性問題，可提升學生的動機與興趣，促使學生努力。脈絡化在本質上，意味著題目具有足夠的內容，以使問題情境更有意義，也幫助學生活用其經驗和知識，成為詮釋問題和解決問題的基礎。然而，富於脈絡真實意義的評量題目，不必然對所有學生都是適合的。某一情境對於甲生具有促進其動機及努力的效果，可以讓他充分發揮表現。但這個脈絡情境對乙學生而言，效果可能相反，反而扭曲了他的表現。因此，評量時的真實性應被看成是一種約束，學生在某些特定情境下去經歷並完成問題。

然而，評量「情境化」或「看起來像真的」並非是有效評量的充分條件，情境化也許代表了實作評量一個重要的品質；而這個特質也許只不過代表了表面效度，別無其他。很多在真實情境中的評量表現，其本身極少價值。任何一種實作評量，真實性的設計都必須強調實際性的複雜度，並須包含所有以後學生應用這些評量的知識的可能相關情形。

要求標準的真實性，主要期望對學生的知識和技能做更豐富的瞭解和掌握，

以描述學生表現時的過程和策略，再調整評量，使其更直接靠近教學目標。假如卷宗的內容反映出課程所強調的內容，就符合了要求標準的真實性。重視高層次思考評量的研究者，認為實作評量是一種直接判斷表現的評量而非間接的能力指標。實作的價值就在其本身所代表的意義，而非其他有價值的表現或其相關表現的間接指標。

評量的真實性大致可以下列的幾點作為判斷的準則：(1) 學生必須運用知識，以有效地表現能力，而這些題目必須類似於成人世界或此領域專家所面對的問題；(2) 必須忠實地描述這個領域或真實情境中的重要要素；(3) 問題需要應用一連串的知識，判斷在何時、何處運用哪些知識較為適合，並且需要組織問題並決定解題步驟的優先順序；(4) 這些表現，具有品質的要求；(5) 指標或標準透明化，要求學生對其答案或選擇舉出理由；(6) 在各式各樣的情境中，強調學生表現和所評量的目標行為之間的一致。

弱結構

為了配合真實性與情境性，評量必須有一些模糊不清及無結構的特質，如此學生才會將判斷力集中於瞭解、澄清及解決問題。除了情境之外，實作評量強調比標準測驗還缺乏結構的問題，在評量中，學生應該能做到在真實問題情境下所該做的事，評量應模擬生活上心智的運作及每天工作的情境。在實際情境中所遇到的重要工作通常是沒有公式可循的，這類的問題和一般結構清楚，強調最後結果的問題不同，這類的問題結構不清且(1) 較複雜，對何時才會有解決之道，並無清楚的準則可依循；(2) 問題的說明中，並不全然包括解題所需的所有資訊，僅提供一些十分模糊的相關資料；(3) 並無單一的或唯一的方式去發現每一個解決步驟的可能性。如 Wiggins (1989) 提出經由無結構的問題的應用，實作評量促使學生的思考集中在提出問題、澄清及解決問題上，評量的是知識與思考的深度而不只是廣度。

過程性

學習是十分複雜的過程，不只包含學生所知道，且包含針對所知道的，學生可以作些什麼。「知道如何應用」顯見是專家和生手間最大的差別，有意義的學習，即在從敘述性的知識到過程性的知識間的轉換，也是從記憶到行動間的轉換，對這類轉換的評量，則需在情境中提供不斷應用的機會，而這樣的過程性知識方能培養出不斷進步的學習能力。在無結構的情境下，問題解決的過程更加的重要，既然沒有明確的正確答案，解決之道是否有意義只能依靠推理與判斷，雖然成果的資訊對呈現學生最後所習得的結果十分重要，而若要改進結果，就必須知道學生在這解題過程中的經驗，導致某種結果的策略與努力的相關資訊。

實作評量與傳統紙筆測驗最大的不同，就是在於它評的不只是學生在評量中產生的結果（答案），也重視產出結果的過程（思考歷程）的評量。實作評量要求的不是唯一正確的或最好的答案，一個問題可能有多種解決方法，而是否能有

系統的敘述問題，思想的組織，證據的整合和原創性等過程性的思考都是實作評量的重點。

事實上，表現的過程和結果都是評量的目標和媒介，有時候結果是目標，而過程是媒介。當我們的重點是過程表現或結果的品質時，則表現或結果本身成了評量的目標；而當我們要對學生的技能、能力或某一構念進行推論時，則過程表現或結果只是評量的媒介。在某些領域，如戲劇和舞蹈，過程和結果基本上是同一回事，因此所需評量的是這種或那種的表現形式。在繪畫或寫作領域，可以有許多不同的過程或不同的開始模式，所以關心重點在於結果。而在機械或化學領域，則表現的過程和結果都需要被仔細加以從頭檢查，因為結果和過程都同樣重要。

社會互動

在真實的情境中，問題往往需要一群人合作方能完成。這一群人也許選擇各自感興趣的問題角度或某部分去著手，而非問題的全部，之後一起討論分享心得。實作評量的題目也強調評量不再限於坐在一張桌椅上或單獨完成。學生可以小組的形式，到市場做調查，在走廊討論，更可以尋求社會人士的支援，也就是所謂的多元練習。小組的互動中，自然會有不同的角色需求，有人領導策劃，有人善於執行，也有人負責監督檢查，就如同社會上不同的角色需求，而這也是實作評量的情境最不同於一般標準測驗的特質。

彈性

實作評量的問題也許是相當的侷限，或十分的廣泛可延伸許多天甚或幾週的主題研究。要評量學生在評量項目中的思考過程，時間就必須十分有彈性。標準測驗通常在一個限定的時間內完成，使得評量者無從看到學生在一段時間中的進步。一個領域的傑出成就通常是無法在一個十分特定的時間如 50 分鐘的課堂，或 2 個小時的考試時間內完成的。實作評量的評量時間設計應能讓學生以持久、長時間深入地探討某個概念。當學生的成就可以經由持續的觀察來瞭解其優缺點時，以實作評量來評量學習成就，其結果就越真實了。

多向度

既然對學生反應的評量重在過程與結果，就需要一種多層性的評分系統來分析出複雜的表現的成分，很明顯的，判斷一個人在某個領域的專業程度，單一的指標，如百分等級所能傳達的訊息不但十分有限，且易出錯。真實情境下的表現是無法以一個分數來傳達其品質的，如一份研究報告，可從資料的完整性，用了哪些權威的資源，和公共政策議題的相關性，及報告本身的組織和邏輯性來判斷。家具設計師的工作可從設計的持久性、有效性，使用材料的美觀和品質，及整體完成的品質來評估。學生的口語表達技巧不只從所使用的字彙的範疇來看，且需從對聽者的觀點是否敏感、解釋的清晰和提問適恰等問題來評量。

既然實作評量強調的不只是簡單的紀錄錯誤與統計分數而已，評分的標準相對的也是相當複雜，實作評量的表現結果，以表現的真實標準為參考，複雜表現的多向度要在判斷中一一被分析出來，評分的系統通常必須多層次而非單一的一個分數。因此，實作評量的評分系統至少要具備三個基本標準：知識的產生、專業的探究，及評量之外的價值。有一些這類的評量是根據學生的書面報告來評分，其他則須評量者針對學生工作時所表現的過程予以評估，這兩種方式，評分者都必須有專業訓練，也須有可用的共同標準，多向度的判斷才不致流於拼湊。我們將在後面討論實作評量的評分表設計。

實作評量的類別

從實作評量的定義及特質，一般都以實作限制的程度或實作問題的結構度，將實作評量的題型分類。實作限制的程度呈現在問題層次的設計上，大都以問題的結構度、解決問題必要條件的提供與否、方式的多元性，以及解決之結果的已知程度等為決定層次的依據。實作層次的類別，可視為一條連續性線上的許多點，呈現的結構、已知條件的多寡、可接受的方法或詮釋的多寡，以及問題、方法及結果的彈性來決定在線上的位置（也就是層次、類別）。

線的一端是最有結構、定義最清楚的層次，這個層次的問題敘述清楚，解題條件充分，問題解決的方法很明顯，出題者與解題者都知道；解題的結果出題者知道，解題者須經由一些已知的知識或方法中去發現。這是解題者最熟悉的問題情境；解題條件愈清楚，反應也就受限了。因此，這類實作類別或稱為限制反應式的實作評量。題目敘述清楚標明形式上的限制，優點在於其有更高的結構性，且需要的執行時間較短，所以在一定時間內可以完成更多的題目，以涵蓋更多想評量的領域範圍。而其缺點在於高結構性，不易評量學生資訊整合的能力和學生的原創性。

線的另一端則是最無結構、定義最模糊的問題層次，這個層次的問題中，問題定義不清（問題本身也許還含有許多不清的問題），解決問題的方法未知，解題條件十分不足，有許多不同的方式，也許需要嘗試新的有創意的方式來解題，因此，出題者和解題者都不知道需要用哪個方式最恰當，解題結果出題者和解題者也無法預知，是解題者最陌生的問題情境。這種開闊式的實作類型要求學生從更多樣的來源蒐集資料，學生使用資料的過程是評量的重點，而產生的結果也有多種不同的形式。自由程度高，使學生能證明其選擇、組織、整合、和評價資料的能力。但其缺點就是喪失評量的效率，喪失內容領域的廣度，另外，也增加了評分的困難。

在這兩端之間則是其他層次的問題。為了研究設計及分析上的限制，一般都將問題分為三或五個層次。結構愈清楚，條件愈充分也表示限制愈多，所需的高層次思考愈少；反之，結構愈模糊不清，問題情境愈新奇，條件愈少，難度愈高，也就愈發需要高層次的計畫與策略監控等的思考活動。

請先閱讀以上的文章，然後根據文章的內容與相關的測驗與評量概念，回答下面的選擇題。

26、請問本文接下去，最可能討論哪一個主題？

- (1) 舉例說明實作評量的類別
- (2) 說明實作評量的結構與表現
- (3) 討論實作評量與標準測驗的差異
- (4) 討論真實評量的類別

27、請問本文接下去，最不可能討論哪一個主題？

- (1) 舉例說明實作評量的類別
- (2) 說明實作評量的結構與表現
- (3) 討論實作評量與標準測驗的差異
- (4) 討論真實評量的類別

28、依照本文，有關實作評量的類別，下列何者最接近事實？

- (1) 一般都將實作評量的問題分為三或五個層次
- (2) 一般都將實作評量的問題分為五或七個層次
- (3) 實作評量的分類系統仍模糊不清
- (4) 一般都將實作評量的問題分為兩個層次

29、依照本文，有關實作評量的時間，下列何者最適當？

- (1) 實作評量的時間經常長達數天
- (2) 實作評量的時間經常長達數週
- (3) 要評量學生在評量項目中的思考過程，時間就必須十分有彈性。
- (4) 要評量學生在評量項目中的解題表現，時間就必須十分有彈性。

30、依照本文，有關實作評量的結構，下列何者最適當？

- (1) 實作評量的結構通常比較清楚
- (2) 實作評量的結構通常比較模糊不清
- (3) 實作評量的結構與品質無關
- (4) 實作評量的結構與目標無關

31、依照本文，有關實作評量的結構，下列何者最適當？

- (1) 問題結構愈強，則真實性與情境性愈低
- (2) 問題結構愈弱，則真實性與情境性愈低
- (3) 問題結構愈強，則真實性愈高，情境性愈低
- (4) 問題結構愈弱，則真實性愈高，情境性愈低

- 32、依照本文，有關實作評量的結構，下列何者最適當？
（1）問題結構愈強，則解題所需訊息愈不足
（2）問題結構愈弱，則解題所需訊息愈不足
（3）問題結構愈強，則解題所需訊息愈模糊
（4）問題結構愈弱，則解題所需訊息愈模糊
- 33、依照本文，有關實作評量的結構，下列何者最適當？
（1）問題結構愈強，則高層次思考愈模糊
（2）問題結構愈弱，則高層次思考愈模糊
（3）問題結構愈強，則高層次思考愈多
（4）問題結構愈弱，則高層次思考愈多
- 34、依照本文，有關實作評量的結構，下列何者最適當？
（1）問題結構愈強，則問題定義愈模糊
（2）問題結構愈弱，則問題定義愈模糊
（3）問題結構愈強，則問題定義愈長
（4）問題結構愈弱，則問題定義愈長
- 35、依照本文，專家和生手的最大差異為何？
（1）知道如何解題（2）知道如何判斷（3）知道如何應用（4）知道如何辨別
- 36、依照本文，有關作文，下列何者最適當？
（1）重視文意切合（2）重視用字遣辭（3）重視過程（4）重視結果
- 37、依照本文，有關樂器演奏，下列何者最適當？
（1）重視表現形式（2）重視技巧精熟（3）重視過程（4）重視結果
- 38、依照本文，有關科展競賽，下列何者最適當？
（1）結果與過程並重（2）重視內容廣度
（3）重視說明清晰（4）重視成果深度
- 39、依照本文，實作評量情境與標準測驗最大的差異為何？
（1）結構強弱（2）社會的互動（3）過程性（4）多向度
- 40、依照本文，有關實作評量，下列何者最適當？
（1）要求標準真實性愈高，學生的動機與興趣愈強
（2）要求標準真實性愈高，學生的動機與興趣愈弱
（3）問題真實性愈高，學生的動機與興趣愈強
（4）問題真實性愈高，學生的動機與興趣愈弱

- 41、依照本文，有關實作特質，不包含下列何者？
（1）執行任務 （2）表現能力 （3）學科應用 （4）學習成就
- 42、依照本文，有關實作評量，下列何者最適當？
（1）問題定義愈不清，結構愈強 （2）問題定義愈不清，結構愈弱
（3）問題定義愈不清，自由程度愈低 （4）問題定義愈不清，自由程度愈高
- 43、依照本文，有關實作評量，下列何者最適當？
（1）問題定義愈不清，解題條件愈繁多
（2）問題定義愈不清，解題方法愈複雜
（3）問題定義愈不清，解題結果愈未知
（4）問題定義愈不清，解題結果愈複雜
- 44、依照本文，有關實作評量，下列何者最適當？
（1）國內中小學教師非常熟悉實作評量
（2）國內中小學教師並不熟悉實作評量
（3）國內很多大學教師在研究實作評量
（4）國外很多大學教師在研究實作評量
- 45、本測驗的試題皆為標準測驗題，請問它對能力的測量最偏重於那一種能力？
（1）推論能力
（2）有關事實的知識能力
（3）組織能力
（4）閱讀能力

三、問答題

- 1、何謂測驗偏差(test bias)？請先說明其定義，再以台灣實際發生的新聞事件舉例說明(5%)，並說明可以事先預防以及事後補救的各種具體方法(5%)。