

倒扣指導語對測驗品質的影響—— 以物理學科測驗為例

區雅倫¹ 張耿輔² 翁儷禎³ 程暉滢⁴

摘要

選擇題的答錯不倒扣是大多數評量專家所建議的計分方式，但台灣重要的入學考試——指定科目考試採用的是答錯倒扣計分，過去研究甚少就學科測驗進行倒扣計分的相關探討，台灣亦缺乏倒扣與否的評鑑文獻。因此，本研究從測驗編製者的觀點，欲了解倒扣與否的測驗指導語對測驗品質的影響。以物理學科測驗為工具，抽樣台灣 313 名高三學生受測不同指導語但試題相同的題本，分析倒扣或不倒扣指導語對答對題數、測驗分數的內部一致性信度、試題難度、鑑別度與棄答題數之影響。結果顯示指導語倒扣與否對試題鑑別度有影響，對棄答題數亦有影響，但對答對題數、測驗分數內部一致性信度與試題難度則沒有影響。如果測驗的主要目的在於區分考生學科能力的高低，本研究結果建議倒扣是可行方法，但若不希望增加考生的答題壓力，不倒扣可能較為適合。然此結論可能僅適用於物理學科測驗，未來可增加不同學科之研究，以瞭解本研究結論在跨學科上的穩定性。

關鍵字：倒扣計分、能力測驗、猜答、測驗品質、測驗指導語、學科測驗

1. 區雅倫，國立台灣大學心理學研究所博士班研究生

2. 張耿輔，國立台灣大學心理學研究所碩士班研究生

3. 翁儷禎，國立台灣大學心理學研究所教授

4. 程暉滢，大學入學考試中心學科研究員

收件日期：2007.03.05；完成修改：2007.08.15；正式接受：2007.09.28

通訊作者：翁儷禎；E-mail：ljweng@ntu.edu.tw

地址：台北市羅斯福路四段 1 號 國立台灣大學心理學系

Effects of Formula Scoring and Number-Right Scoring Instructions on Test Quality: Taking a Multiple-Choice Physics Test as an Example

*Ya-Lun Ou*¹ *Ken-Fu Chang*²
*Li-Jen Weng*³ *Wei-Ying Cheng*⁴

Abstract

The purpose of this study is to compare the effects of instructions of formula scoring and number-right scoring on test quality on a multiple-choice physics test. Three hundred and thirteen 12-grade students were administered the same test with different instructions. The results indicated that the number of omitted items and the item discrimination index might vary with instructions given. If the main purpose of test use is to discriminate among students of different levels of ability, instructions of formula scoring as used by the Department Required Test for college entrance in Taiwan is recommended. Future research is encouraged to evaluate the generalizability of the present findings to tests on other subjects.

Key words: ability test, formula scoring, guessing, subject test, test instructions, test quality

-
1. Ya-Lun Ou, Graduate Student, Department of Psychology, National Taiwan University
 2. Ken-Fu Chang, Graduate Student, Department of Psychology, National Taiwan University
 3. Li-Jen Weng, Professor, Department of Psychology, National Taiwan University
 4. Wei-Ying Cheng, Researcher, College Entrance Examination Center
- Received: 2007.03.05; Revised: 2007.08.15; Accepted: 2007.09.28
Corresponding Author: Li-Jen Weng; E-mail: ljweng@ntu.edu.tw
Address: 1, Section 4, Roosevelt Road, Taipei, Taiwan
Department of Psychology, National Taiwan University

壹、緒論

在進入大學接受高等教育之前，世界各國的高中生都有一個重要生涯關卡，就是接受不同形式的入學考試，以評斷是否具有進入大學校系就讀的能力，如台灣的大學指定科目考試（指考）與學科能力測驗（學測）、美國大學入學測驗 SAT Reasoning Test 與 ACT Assessment、歐洲各國的畢業會考、日本大學入學考試中心考試、中國大陸的普通高校招生入學考試等（林佩寬，2007）。為配合大量考生受測的需要，電腦可讀的選擇題成為前述各考試之主要題型（Koretz & Hamilton, 2006），計分採答對給分、答錯不給分，然而台灣的大學指考可能是唯一的例外。指考命題者認為考生要進入理想學系競爭非常激烈，在志願分發上分分計較，為避免考生投機猜題，造成考試不公平，故在選擇題設計了答錯則倒扣分數的機制（區雅倫，2005）。答錯不倒扣雖是目前大多數評量專家所建議的方法（Downing, 2003），然台灣的指考卻採用答錯倒扣，且鮮少對此議題進行探討，有關倒扣計分對測驗品質的影響研究亦闕如，因此對指考答錯倒扣的適切性進行評鑑缺乏立論依據。基於大學指考每年影響十萬以上莘莘學子的發展，本文以實徵研究來瞭解答錯倒扣對測驗品質的影響，期能做為未來決策的參考。

考生在開始作答能力測驗前，一定會先行閱讀測驗指導語，瞭解計分方式後以某種因應策略進行答題。選擇題型常用的計分指導語有二：倒扣計分指導語（formula scoring instruction）與不倒扣計分指導語（number-right scoring instruction）（Budescu & Bar-Hillel, 1993; Lord, 1975; Rowey & Traub, 1977）。倒扣計分指導語（簡稱倒扣指導語）在測驗作答前清楚地告訴考生，他所得的分數是答對題數再減去答錯題數的部分比例，再乘以配分；不倒扣計分指導語（簡稱不倒扣指導語）說明僅以答對題數乘以配分。多年來，雖然國外已有多篇研究從考生答題習慣、測驗分數信度與效度的角度，討論不同指導語對考生答題之影響（Aiken, 1987; Diamond, 1975），然而過去研究多在詞彙或性向測驗上進行，而非針對學生最常用的學科測驗（Bliss, 1980; Fujimori, & Nakano, 1994; Wood, 1976）。故在此即針對入學學科考試進行研究，並冀望其結果能彌補過去之不足。

一份優良的學科測驗，多數學者認為須包括信度、效度、參照性與客觀性（余民寧，1997；Airasian, 1994）。一種適宜的測驗指導語，積極面有助

於測驗品質的提升，消極面則不破壞原有的測驗品質。多年來大學指考之考生皆在倒扣指導語下答題，無法單由指考答題結果來分析指導語倒扣與否對測驗品質的影響。本研究以內容相同的一卷物理學科測驗，分採倒扣與不倒扣指導語兩種版本，同時施測於兩個隨機考生群樣本，以考生在不同指導語下之答對題數做為能力表徵，以棄答題數來檢視測驗的客觀性。在信度方面，倒扣指導語是否影響測驗分數的信度；在效度方面，倒扣指導語是否影響考生的答對題數、試題難度與鑑別度；在客觀性方面，倒扣指導語是否抑制作答意願而影響棄答題數。這樣的結果不但可與過去研究進行比較，且能對評鑑指考答錯倒扣的適切性提出建言。

然而，在探討倒扣與不倒扣計分指導語前，需先說明不同指導語下的計分方式，以瞭解考生的答題策略為何可能因指導語而有差異。

一、倒扣與不倒扣計分指導語

選擇題型在能力測驗雖擁有相當優勢（余民寧，1997），但考生在面對不會作答或不確定正確選項時，可藉隨機猜題（random guessing）的方式得到分數，這將使測驗誤差加大，影響考試公平性與試題鑑別度（簡茂發，2001; Diamond & Evans, 1973; Mattson, 1965）。為解決選擇題猜答的問題，測驗學者設計出倒扣指導語，相對就是不倒扣指導語。倒扣指導語的特點是考生對不會的試題不敢亂猜答案，但也同時抹殺了考生使用部分知識（partial knowledge）來排除某些選項再進行猜答的行為（Diamond, 1975）；不倒扣指導語的特點是簡單易懂，鼓勵考生對不確定正確選項的試題予以猜答（guessing），但這對有能力正確回答該題的考生卻造成不公平（Lord, 1975）。考生在閱讀各種不同指導語後，通常會採取不同的答題策略，以獲取最滿意的分數（Diamond & Evans, 1973; Downing, 2003）。

其實，倒扣或不倒扣是頗具爭議的議題，正如 Lord（1975）所認為，倒扣與宗教或政治的問題一樣，正反雙方都堅持己見；Rowley 與 Traub（1977）則認為支持倒扣與否乃是個人的價值判斷。主張倒扣者認為倒扣可以去除猜答的因素，正確估計考生真正的知識成就，維持考試的公平性（Lord, 1975）；反對倒扣的理由則是倒扣的基本假設不符合事實，部分知識的答題策略應被考量，倒扣雖然可以防堵猜答，但對不同人格特質與作答態度有不均衡的影響，且倒扣無法使考生的作答趨於一致，導致測驗分數的信度降低（簡茂發，2001）。

考生會因指導語的不同而有不同的因應方式，以獲取最滿意的分數。單一選擇題之答案通常只有一個，考生的答題情況只有答對、答錯與棄答三種。假設在一份每題有 k 個選項的選擇題試卷上，一個考生答對 R 題，答錯 W 題，棄答 M 題， R 、 W 、 M 三者之和即為試卷題數。最簡單的計分方法就是不倒扣，若以 $S1$ 代表不倒扣計分，在 $S1$ 下考生的得分即等於 R 。因為答錯與棄答均得 0 分，但猜答有機會答對，故從決策理論的觀點來看，在指導語採不倒扣計分時，全部選答是考生重要的策略之一（Budescu & Bar-Hillel, 1993）。

距今八十年前，倒扣計分的概念與公式就已經發展完成（Holzinger, 1924; Thurstone, 1919），以 $S2$ 代表之，在 $S2$ 下考生之分數等於 $R - [\frac{W}{(k-1)}]$ 。這個公式的基本假設是，考生對不知正確答案的試題，均隨機猜答，故答錯的試題完全是隨機答錯所致。至於答對的總題數則包括兩個部分，即確知答案而答對者與不知正確答案而猜對者，後者應從答對的題數中扣除之。因猜對題數無法直接獲知，但能推測得知猜對題數為猜錯題數的 $\frac{1}{(k-1)}$ 倍，故應自 R 中減去 $\frac{W}{(k-1)}$ （簡茂發，2001; Thorndike Ed., 1971）。Traub、Hambleton 與 Singh（1969）發展出第三種計分方法，以 $S3$ 代表，在 $S3$ 下考生之得分為 $R + (\frac{M}{k})$ ，這種方法與 $S2$ 都不鼓勵考生猜答，只是 $S3$ 給予棄答題部分分數，把答錯處罰的方式改成答錯不給分， $S2$ 與 $S3$ 兩種計分方式下的得分排序是相同的。雖然 Waters 與 Waters（1971）的研究顯示考生喜歡 $S3$ 勝於 $S2$ ，但目前選擇題主要的計分方式多採用 $S1$ 與 $S2$ ，故本研究僅關注於這兩種計分方式之指導語。

二、指導語倒扣與否對測驗品質之影響

（一）信度

測驗分數的信度為對測驗評價之重要指標，但指導語倒扣與否對於信度的影響並無一致結論（Burton, 2004）。內部一致性信度是整份測驗所有試題反應一致的程度，隨機猜答可能會擾亂一致性，因此有學者認為不倒扣指導語會降低測驗分數的信度。但相反的也有學者認為，不倒扣指導語鼓勵考生運用部分知識作答，考生在試題上之表現應較為穩定，故可提高內部一致性

(Aiken, 1987)。有些研究結果顯示有無倒扣指導語下，考生測驗分數的信度並沒有顯著差異 (Sabers & Feldt, 1968; Traub et al., 1969)，但有些研究結果發現在倒扣指導語下，考生測驗分數之信度會較高 (Diamond, 1975)。Frery、Cross 與 Lowry (1977) 則發現，在不倒扣指導語下，因猜測被鼓勵，信度會稍微升高。因為信度與受試者能力的變異有關，受試者間能力變異愈大則信度可能愈高，Burton (2002, 2004, 2005) 建議測驗編製者如果想要提高信度，可以使用倒扣計分，因為倒扣會增加分數的變異性。

從提高信度的觀點考量，認為要增加分數的變異程度，倒扣指導語與計分方式是值得建議的。

(二)考生能力：答對題數

反對倒扣的學者認為，倒扣指導語讓有部分知識的考生不敢猜而棄答，以致影響其真實能力之表現；主張倒扣者認為，倒扣可以去除猜答的因素，藉以正確估計考生真正的知識成就 (Bliss, 1980; Cross & Frery, 1977)。但何者是考生的真實能力？且如何評量考生的真實能力？想必這是測驗理論長期所關心但永無共識的話題。若學科評量專家能編製出一份品質優良的學科測驗，該測驗能測出所欲測量之學科能力、呈現適切的問題類型、達到知識領域比率的平衡、有該領域知識才會作答、有高的內部一致信度、良好的外在關聯效度、適當的難度與良好的鑑別度等，這樣的測驗在合適的環境中受測，考生在該測驗的答對題數應可表示考生在該學科的能力 (Ebel & Frisbie, 1991)。在測驗中面對相同的試題，能力較佳的考生答對之可能理應高於能力較差的考生，由此觀之，答對題數或可為考生能力之表徵。

Frery (1989) 發現在適當難度的考題下，倒扣指導語會讓考生不敢猜答而造成成績降低。Albanese (1988) 的研究亦呈現，相同考生應試不倒扣指導語下的成績高於倒扣指導語，原因在於前者每題均答。Quereschi (1974) 發現倒扣指導語降低了答對題數，增加了答錯題數。Traub 等人 (1969) 的研究顯示，在倒扣指導語下受測之考生，其答對與答錯的題數均低於無倒扣指導語的考生。總而言之，諸多研究顯示倒扣指導語會降低考生的答對題數。

(三)試題難度與鑑別度

試題參數如難度（通過率）、鑑別度等是試題分析的重要指標。過去倒

扣指導語的研究聚焦於考生觀點，多考量指導語對考生答題反應的影響。然測驗編製者關心的是測驗品質，理當關注試題參數的議題。Quereshi (1974) 發現，在倒扣指導語下考生答對題數顯著下降，答錯題數與棄答題數則增加；換言之，倒扣指導語下通過率可能下降。至於試題之鑑別度，在倒扣指導語下，不確定答案而隨機猜題的考生可能改為棄答，使得試題更具鑑別力；但考生也可能因倒扣計分而放棄使用部分知識進行合理猜答，致使試題的鑑別度降低。然而，在過去的研究中，鮮少由此一角度來探討倒扣指導語的影響，故本研究將以實徵資料來檢視倒扣與否的指導語對難度與鑑別度的影響。

(四)棄答題數

考生棄答題數的多寡，嚴格講起與測驗品質沒有直接關係，但棄答表現出考生因畏懼倒扣分數，寧願放棄作答而不敢嘗試的心理，也影響了測驗的客觀性。在倒扣與否的指導語下，考生關心的是要不要猜答，測驗編製者關心的是倒扣是否會對不知答案的考生遏止其猜答 (Budescu & Bar-Hillel, 1993)。Aiken (1987) 曾綜合分析當時有關倒扣計分與猜答的研究，他認為考生猜答並不是隨機發生的，而是排除某些選項後才進行，倒扣指導語時猜答行為會減少，這種猜答行為雖無法從答題對錯顯示，卻可表現在棄答題數上。過去許多研究亦顯示，在倒扣與不倒扣指導語之下，考生之棄答題數會有所差異 (Aiken, 1987; Lord, 1975; Traub & Hambleton, 1973)。如 Waters 與 Waters (1971)、Traub 與 Hambleton (1973) 的研究顯示，於倒扣指導語下考生的棄答題數較高。Lord (1975) 亦發現，兩種指導語下考生答題的差別僅在棄答題數，既然倒扣可以防止隨機猜答，故他建議測驗編製者採用倒扣指導語。Fujimori 與 Nakano (1994) 在日本對大學生之施測則發現，在倒扣指導語、不倒扣指導語與無說明倒扣指導語三種情境下，考生棄答題數有顯著的差異。

有關倒扣指導語影響棄答的現象，可從台灣的大學入學學測與指考略窺一二。這兩種考試的考生能力分布相近，但測驗目的不同，指考的測驗目標是為大學選才，考生成績分分計較，測驗範圍包括高中三年所學，試題比學測難；學測是為大學入學做初步的檢定，瞭解考生是否學會高中基本學科知識與技能，測驗範圍僅限高一高二課程，試題比指考容易 (曹亮吉、程暉滢, 2003)。從分析兩種考試各科的平均棄答率，略可察覺倒扣指導語對答

題反應可能具相當影響。表 1 是 2005 年學測與指考各科單選題平均棄答率。學測採不倒扣指導語，幾乎所有考生每一題均作答；指考採倒扣指導語，全體考生在各科的平均棄答率少至 8%，多達 24.4%。既然倒扣與不倒扣指導語出現在兩種不同考試，考生雖能力相似，但試卷卻不同，其結果難以釐清學科考試中倒扣指導語是否影響棄答，而需要設計相同內容但不同指導語之試卷，其結果才能探討指導語倒扣與否對棄答的影響。

表 1 2005 年學測與指考各科單選題平均棄答率

指考	國文	英文	數甲	數乙	歷史	地理	物理	化學	生物
棄答率	10.47	12.50	14.00	14.00	9.97	8.00	24.40	14.64	10.10
學測	國文	英文	數學	社會	自然				
棄答率	0	0	0	0	0				

註：平均棄答率為該科各單選題試題棄答率之平均。單選題之棄答率為該題未作答人數除以全體人數，再乘以 100。

資料來源：大學入學考試中心（2005）

三、研究目的

綜合上述倒扣研究文獻可知，不倒扣指導語與計分法簡單易懂，鼓勵考生以部分知識來進行猜答，獲得近代各項大型客觀測驗所採用。考生在作答無倒扣指導語的試卷時，較敢對不確定答案試題猜答，但答錯題數卻也增加。相反的，倒扣指導語與計分法能讓較少知識背景之考生有答錯倒扣的壓力，使其對答案不明確的試題，由於害怕猜錯而不敢隨意猜答，改採棄答的保守策略。過去實徵研究中所採用的測驗工具，內容多以詞彙測驗或性向測驗為多，較少需要累積知識來推理的學科測驗，且大多數研究著重考生答題策略，較少從測驗品質（如難易度與鑑別度等）之觀點進行分析（Bliss, 1980; Wood, 1976）。雖有研究以不同難度的試卷進行測試與分析，但探討倒扣指導語及計分對不同能力考生影響之文章亦闕如。

因此，本研究採用物理學科專家認為難易適中的高中物理學科能力試卷，設計倒扣與不倒扣兩種指導語，於同一時間施測於兩群能力相同的考生。本研究欲從考生答題反應結果探討下列問題：(1)指導語有無倒扣與考生能力對考生答對題數與棄答題數，有無顯著影響或交互作用？(2)測驗的品質如內部一致性信度、難度、鑑別度，是否因有無倒扣指導語而有顯著不同？

貳、研究方法

一、題本與考生樣本

本研究所選用的測驗是大學入學考試中心物理科研究用試卷，該試卷內容以高中三年所研習之物理知識為主，包括力學、流體性質、熱學、波動、電磁學、電子學、光學、近代物理與實驗方法。測驗目標在評量考生知識記憶、基本觀念的瞭解、方法與原理的綜合運用、分析數量關係等能力。該試卷題型依序有單選題（每題五個選項中僅有一個正確答案）、多選題（每題有一個以上正確答案）與非選擇題，全卷受測時間為 80 分鐘。

本研究依研究目的僅選擇 20 題單選題進行分析，因單選題配置於本卷最前段，考生應有充裕的時間完成所有單選題。命題者在設計單選題時，儘量讓每題僅評量單一知識或概念，物理學科專家認為，此 20 題難度為中等偏易。考生的作答結果雖受學科知識之影響，然而本研究僅針對有無倒扣指導語所造成的作答反應加以探討，而不涉及試題在物理學科知識的質性分析。

試卷依不同指導語分為兩種，不倒扣指導語是：「本試卷的選擇題不倒扣，每題答對得該題題分；每題得分最少為 0 分，而不會是負值」；倒扣指導語則為：「本試卷計分方式（含倒扣）與目前指定科目考試相同，答錯倒扣 $\frac{1}{4}$ 題分，倒扣到單選題實得分數 0 分為止，未答者不給分亦不扣分」。每題均有五個選項，考生擇一選項，將答案劃記在答案卡上。

本研究施測的時間為 2006 年 5 月中旬，受測考生抽樣方法採「高中生抽樣方法」（林光賢，1990），以 2005 年指考物理科全體考生為母群體，隨機組合五所高中應試考生，其指考物理科成績分布與母群體相似，採此五所高中 2006 年高三自然組學生為受試。

倒扣與不倒扣指導語兩種題本乃依序交錯疊放，施測前監考老師直接發放題本與答案卡。監考老師在教室內不宜讀題本指導語，而由考生自行閱讀，故考生並不知道鄰座同學之試卷指導語是否均同。受試共計 465 名，其中受測倒扣指導語的考生有 230 人，受測不倒扣指導語的有 235 人。由於抽樣時以指考物理科成績分布為取樣標準，而 465 名中有部分受試沒有參加指考物理科考試，故本研究待 2006 年 7 月指考成績公布後，僅以實際參加指考

的 313 人為樣本，其中男生 236 名，女生 77 名。受測倒扣指導語的考生有 167 人，受測不倒扣指導語的考生有 146 人。

二、資料分析

為比較不同物理科能力考生在兩種指導語下的答題反應，將 313 名考生依指考物理科成績高低排序，令前三分之一為高能力組考生，共 104 人（倒扣 52 人，不倒扣 52 人）；後三分之一為低能力組考生，共 108 人（倒扣 64 人，不倒扣 37 人）；餘 101 人為中能力組考生（倒扣 51 人，不倒扣 57 人）。

本研究首先以二因子變異數分析（2-way ANOVA），探討考生的答對題數是否受有無倒扣指導語、考生能力及其交互作用之影響，並計算各題答對率（該題答對人數除以全體考生數）與鑑別度（高能力組考生與低能力組考生答對率差距）。難易度以大學入學考試中心物理科所採用的分類，將各題依答對率換算為難中易三級（程暉滢，2005），比較有無倒扣指導語之難中易的題數分布何者與學科專家所預估的較為接近；鑑別度以 Crocker 與 Algina（1986）建議之評鑑標準給予優、良、可、劣四等級，而後比較有無倒扣指導語下各題鑑別度分類是否一致。探討有無倒扣指導語下測驗分數的內部一致性信度 Cronbach's α 是否具顯著差異時，則採用 Feldt 等人所發展之方法進行檢定（Feldt, Woodruff, & Salih, 1987）。

參、研究結果

表 2 呈現不同指導語情境下，考生答對、答錯與棄答题數之描述統計量，其中答對、答錯題數在不同指導語下平均數雖略有不同，但由偏態、峰度視之，分配大體均符合常態。

一、指導語倒扣與否是否影響答對題數

針對考生答對題數，進行指導語情境（有倒扣、無倒扣）與考生能力（高、中、低）之二因子變異數分析的結果呈現於表 3。由表 3 可知，指導語情境與能力二因子對考生答對題數沒有交互作用，有無倒扣之指導語亦對考生的答對題數沒有顯著影響，僅考生之能力高低對其答對題數有顯著影響。表 4 為各組答對及棄答的平均題數與標準差，由表 4 亦可發現，隨著考

表 2 不同指導語情境下，考生答對、答錯與棄答題數之描述統計量

作答結果	倒扣 (N = 167)				不倒扣 (N = 146)				全體 (N = 313)			
	平均數	標準差	偏態	峰度	平均數	標準差	偏態	峰度	平均數	標準差	偏態	峰度
答對題數	9.07	4.45	0.31	-0.45	9.80	4.13	0.21	-0.63	9.41	4.31	0.25	-0.54
答錯題數	8.36	4.03	0.13	-0.61	10.01	4.25	-0.17	-0.71	9.13	4.21	0.01	-0.73
棄答題數	2.57	3.26	1.46	1.77	0.19	0.62	4.14	19.49	1.46	2.69	2.32	5.54

表 3 指導語情境與考生能力對「答對題數」之變異數分析結果

變異來源	SS	df	MS	F	η^2	p
指導語	6.87	1	6.87	0.67	.00	.413
能力	2467.34	2	1233.67	120.78**	.44	<.001
指導語×能力	43.20	2	21.60	2.12	.01	.122
誤差	3135.85	307	10.21			
總變異	5653.26	312				

** $p < .001$

表 4 各組答對與棄答之平均題數及標準差

		考生能力							
		高		中		低		全體	
答對題數	有倒扣	13.08	(3.63)	8.78	(3.25)	6.03	(3.22)	9.07	(4.45)
	無倒扣	13.62	(3.04)	8.07	(3.21)	7.11	(2.55)	9.80	(4.13)
	全體	13.35	(3.35)	8.41	(3.23)	6.34	(3.02)	9.41	(4.31)
棄答題數	有倒扣	1.81	(2.30)	2.18	(2.90)	3.52	(3.94)	2.57	(3.26)
	無倒扣	0.21	(0.70)	0.19	(0.52)	0.16	(0.69)	0.19	(0.62)
	全體	1.01	(1.87)	1.13	(2.25)	2.29	(3.55)	1.46	(2.69)

註：各格數值為平均值，括弧中為標準差。

生能力愈高，其平均答對題數從低能力組的 6.34 題，漸漸提升至中能力組的 8.41 題，乃至高能力組的 13.35 題，亦呈現增加之趨勢。

二、指導語倒扣與否是否影響內部一致性信度

全體考生測驗分數之內部一致性信度 $\alpha = .783$ ，在指導語情境為有倒扣

時，其測驗分數之內部一致性信度 α 為 .805 ($N = 167$)；倘若指導語為無倒扣，則信度 α 下降至 .762 ($N = 146$)，兩者間不具顯著差異 ($F(167, 146) = 1.22, p = .111$)。由此可見，採用有無倒扣之指導語，對測驗分數之內部一致性信度無顯著影響。

三、指導語倒扣與否是否影響試題難度與鑑別度

(一)試題難度

於不同指導語下，各題之通過率列於表 5，為探究兩種指導語對各題難度的影響，將各題依據通過率分為難中易三種程度。全部 20 題的難度在不同指導語情境下均一致，難度中等有 16 題，難題有 2 題，易題有 2 題，與學科專家預估相同的有 12 題，而指導語有無倒扣對試題難度不具影響。

(二)試題鑑別度

有關各題鑑別度及其是否受有無倒扣指導語所影響之結果，請見表 6。全部 20 題中，共有 9 題鑑別度之分類在不同指導語情境下不一致，其中 6 題於倒扣指導語下較佳，餘 3 題則是於無倒扣指導語下較佳。以此觀之，試題之鑑別度傾向於倒扣指導語下較為優良。

四、有無倒扣指導語對棄答題數之影響

表 7 呈現對考生棄答題數，進行指導語情境（有倒扣、無倒扣）與考生能力（高、中、低）之二因子變異數分析結果。由表 7 中可發現，考生棄答題數確實會受指導語之不同影響；考生能力高低及能力與指導語間交互作用之影響雖未達顯著之標準，但其效果量已達 Cohen (1988) 所建議之小效果，故其影響亦不宜全然漠視之。表 4 各組棄答平均題數與標準差的結果顯示，有倒扣指導語的考生平均棄答題數 (2.57) 高於不倒扣指導語 (.19)，顯示不倒扣指導語對考生之作答有鼓勵效果。



表 5 有無倒扣指導語下各題之通過率及其難易度差異比較

題號	有倒扣 指導語 (N = 167)	無倒扣 指導語 (N = 146)	專家 預估
1	0.58 (中)	0.60 (中)	中
2	0.22 (難)	0.27 (難)	中
3	0.50 (中)	0.49 (中)	易
4	0.56 (中)	0.57 (中)	易
5	0.53 (中)	0.41 (中)	中
6	0.53 (中)	0.45 (中)	中
7	0.60 (易)	0.72 (易)	中
8	0.41 (中)	0.51 (中)	易
9	0.30 (中)	0.36 (中)	中
10	0.57 (中)	0.56 (中)	易
11	0.69 (易)	0.76 (易)	易
12	0.42 (中)	0.47 (中)	中
13	0.40 (中)	0.41 (中)	中
14	0.37 (中)	0.41 (中)	中
15	0.56 (中)	0.56 (中)	易
16	0.54 (中)	0.58 (中)	易
17	0.22 (難)	0.22 (難)	難
18	0.31 (中)	0.50 (中)	中
19	0.39 (中)	0.50 (中)	中
20	0.37 (中)	0.45 (中)	中

註：答對率大於或等於 .6 為容易題；答對率在 .3 到 .6 之間為中等題；答對率小於或等於 .3 為難題（程暉滢，2005）。

表 6 有無倒扣指導語下各題鑑別度及其差異比較

題號	有倒扣 指導語 (N=167)	無倒扣 指導語 (N=146)	等級 差異
1	0.33 (良)	0.12 (劣)	倒扣 > 不扣
2	0.32 (良)	0.42 (優)	倒扣 < 不扣
3	0.54 (優)	0.36 (良)	倒扣 > 不扣
4	0.14 (劣)	0.27 (可)	倒扣 < 不扣
5	0.32 (良)	0.35 (良)	-
6	0.29 (可)	0.35 (良)	倒扣 < 不扣
7	0.51 (優)	0.41 (優)	-
8	0.28 (可)	0.26 (可)	-
9	0.50 (優)	0.53 (優)	-
10	0.44 (優)	0.33 (良)	倒扣 > 不扣
11	0.35 (良)	0.20 (可)	倒扣 > 不扣
12	-0.06 (劣)	-0.12 (劣)	-
13	0.34 (良)	0.32 (良)	-
14	0.47 (優)	0.56 (優)	-
15	0.46 (優)	0.49 (優)	-
16	0.55 (優)	0.55 (優)	-
17	0.21 (可)	0.13 (劣)	倒扣 > 不扣
18	0.22 (可)	0.22 (可)	-
19	0.40 (優)	0.31 (良)	倒扣 > 不扣
20	0.47 (優)	0.42 (優)	-

註：倒扣為倒扣指導語，不扣是不倒扣指導語。0.4 ≤ 鑑別度：優；0.3 ≤ 鑑別度 < 0.4：良；0.2 ≤ 鑑別度 < 0.3：可；鑑別度 < 0.2：劣（Crocker & Algina, 1986）。

表 7 指導語情境與考生能力對「棄答題數」之變異數分析結果

變異來源	SS	df	MS	F	η^2	p
指導語	406.44	1	406.44	72.38**	.19	<.001
能力	37.22	2	18.61	3.32	.02	.038
指導語×能力	41.54	2	20.77	3.70	.02	.026
誤差	1724.05	307	5.62			
總變異	2209.25	312				

** $p < .001$

若探究考生能力與指導語情境之交互作用對棄答題數之影響，則在無倒扣指導語下，各能力層考生之棄答題數並無差異（低 .16；中 .19；高 .21； $F(2, 307) = .00$ ， $p = .995$ ）；然而，在倒扣指導語下，棄答題數則隨考生能力升高而漸少（低 3.52；中 2.18；高 1.81； $F(2, 307) = 8.49$ ， $p < .001$ ）。對於能力較佳之考生，與不倒扣指導語情境下相較，倒扣指導語對其棄答題數之影響並不大。然而考生能力愈差，有無倒扣指導語對其棄答題數之影響就愈趨明顯。

肆、結論與討論

大學指考是目前國內外大型入學測驗中，可能是唯一選擇題採答錯倒扣的測驗，由於缺乏學科測驗在倒扣議題上之實徵研究成果，對指考答錯倒扣的適切性進行評鑑缺乏立論依據。本研究以入學考試研究用的物理科測驗為工具，從考生答對題數、內部一致性信度、試題難度、鑑別度與棄答率，分析有無倒扣指導語對測驗品質之影響。

指導語有無倒扣孰較能反應考生能力，是本研究在效度上關心的議題。有關考生能力的表徵，除了由學科專家確認試卷的良好內容效度外，全體樣本 313 人之測驗分數內部一致性信度 $\alpha = .783$ ，尚稱理想。考生答對題數與指考物理科成績的相關為 .71，效標關聯效度亦屬良好。全體樣本在全卷 20 題中有 16 題屬中等難度試題，與學科專家的評估相近；鑑別度 D 值在 0.2 以上的亦有 16 題。這些數據顯示，整份試卷具良好信度與效度，考生的答對題數應可反應考生的物理學科能力。

Angoff (1989) 認為，採不倒扣指導語時，對高能力考生應較為有利，

因為他們運用部分知識成功猜對的機率比低能力考生來的高。然而，於本研究中無論採用的指導語有無倒扣，在答對題數上，考生的表現均無差異。因此，若以考生答對題數視之，指導語倒扣與否並不會在反映考生能力上有所偏頗，此點可供專研考試公平性的學者專家參考。支持不倒扣指導語的學者認為，在不倒扣下考生能夠使用部分知識來排除某些選項（Diamond, 1975），由此推論，不倒扣下考生答對的機會應該會有所提升。但本研究卻未顯現如此之結果，是學科特性造成部分知識無法發揮？抑或台灣地區考生已長期受到倒扣指導語答題訓練的影響？關於此點，應於未來研究中藉選項分析或深度訪談進行探究。此外，本研究發現，考生在無倒扣指導語下，其平均棄答題數明顯減少（有倒扣：2.57 題；無倒扣：0.19 題； $t(311) = 9.25$ ， $p < .001$ ），顯示考生之作答意願在無倒扣情境下被鼓勵，然而答對題數卻未因此增加，故可推估考生在無倒扣指導語下，雖然猜答表現增多，但考生之猜答並未能達到效果，即「猜的多，錯的也多」。

過往研究對測驗分數之內部一致性信度，是否受指導語有無倒扣影響的結論多所歧異，本研究則是得到與 Sabers 與 Feldt（1968）及 Traub 等人（1969）相同之結果，顯示無論在何種指導語下，測驗分數之內部一致性信度皆無顯著不同。然而過去研究所針對之測驗，多為詞彙或性向測驗，其本質與目的均與本研究所採之學科測驗不太相同，因此指導語倒扣與否對於學科測驗分數之內部一致性信度是否確無影響，此尚待未來持續研究釐清。

對於試題參數，本研究結果發現，在不同指導語之下，試題之難度並未因而有所變化，顯示對考生而言，答對試題之可能並未因指導語而有所不同。試題難度與測驗成績之間應有相關，所以兩種指導語下的試題難度與答對題數無歧異。而近半數之試題鑑別度因倒扣指導語與否而有變化，其中有倒扣指導語下之試題鑑別度較佳。此結果可建議學科測驗編製者，若想提高試題鑑別度時，倒扣是可考慮的方式。

有關試題的棄答題數，本研究發現，當採無倒扣指導語時，確實會使試題棄答題數下降。當考生面對不確定正確答案的試題時，他可選擇棄答或猜答，猜答可能答對，亦可能答錯。在無倒扣指導語的情境下，由於「棄答」與「猜答但答錯」所造成之結果相同（均得 0 分），但猜答卻有答對之可能，因此考生的決策多採猜答而不會選擇棄答，致使棄答題數下降。但在有倒扣指導語之情境下，考生若猜答錯誤，非但未能得分，反而會因被倒扣而失去既得的分數，此無疑對考生形成一種「懲罰」，其結果會抑制考生的猜

答行為，故有倒扣指導語時棄答題數會較高。而在無倒扣指導語之情境下，考生的猜答行為形同被鼓勵，這樣的鼓勵猜答，卻並未對試題之通過率造成顯著變化；意即，縱使考生的猜答行為被無倒扣之指導語所鼓勵，但大多數猜答之考生均未能猜對。此一假設，亦反映在前述之答對題數並無顯著差異上。故若以考生答對題數評估，指導語倒扣與否並不會對考試之公平性造成干擾。

綜合上述分析，指導語倒扣與否對試題鑑別度有影響，有倒扣指導語下的試題鑑別度較高；棄答題數亦有影響，有倒扣指導語下考生棄答題數較高；但指導語倒扣與否對答對題數、內部一致性信度與試題難度則沒有影響。建議學科測驗編製者在面對倒扣與否的決策上，倘若測驗的主要目的在於區分考生的學科能力，倒扣為可行的方法之一；但若不希望增加考生的答題壓力，採用不倒扣可能較為適合。

因本研究是倒扣計分相關議題的初階基礎研究，僅針對學科測驗考生答題反應與試題參數進行分析，在考生之答題策略與決策歷程等部分並未深入研究。未來可在施測同時，以問卷或深度訪談方式，瞭解考生在面對不同難度與指導語之試題的作答策略與猜測行為，且運用選項分析結合學科專家之試題質性分析，瞭解部分知識如何在作答中發揮效用。本研究之結論僅限於物理學科測驗，未來尚可增加不同學科之研究，以瞭解本研究結論在跨學科上的穩定性。

參考文獻

中文部分

- 大學入學考試中心（2005）。**九十四年度指定科目考試工作報告**。台北市：作者。
- 余民寧（1997）。**教育測驗與評量——成就測驗與教學評量**。台北市：心理。
- 林光賢（1990）。**高中生抽樣方法**。大學入學考試中心研究計畫報告。台北市：大學入學考試中心。
- 林佩霓（2007）。歐洲高中畢業會考與高中課程分組的關係——探討法國 Bac 與英國 A-level 的改革。**考試學刊**，2，39-60。
- 區雅倫（2005）。**探討倒扣機制**。大學入學考試中心內部專題報告。台北市：大學入學考試中心
- 曹亮吉、程暉滢（2003）。學科能力測驗與指定科目考試的命題理念與方向。**文教新潮**，2，12-18。
- 程暉滢（2005）。**九十三年指定科目考試試題分析——物理科**。台北市：大學入學考試中心。
- 簡茂發（2001）。**心理測驗與統計方法**。台北市：心理。

英文部分

- Aiken, L. R. (1987). Testing with multiple-choice items. *Journal of Research and Development in Education*, 20, 44-58.
- Albanese, M. A. (1988). The projected impact of the correction for guessing on individual scores. *Journal of Educational Measurement*, 25, 149-157.
- Angoff, W. H. (1989). Does guessing really help? *Journal of Educational Measurement*, 26, 323-336.
- Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.
- Bliss, L. B. (1980). A test of Lord's assumption regarding examinee guessing behav-

- ior on multiple-choice tests using elementary school students. *Journal of Educational Measurement*, 17, 147-153.
- Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 30, 277-291.
- Burton, R. F. (2002). Misinformation, partial knowledge and guessing in true/false tests. *Medical Education*, 36, 805-811.
- Burton, R. F. (2004). Multiple choice and true/false tests: Reliability measures and some implications of negative marking. *Assessment and Evaluation in Higher Education*, 29, 585-595.
- Burton, R. F. (2005). Multiple-choice and true/false tests: Myths and misapprehensions. *Assessment and Evaluation in Higher Education*, 30, 65-72.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: Holt, Rinehart, and Winston.
- Cross, L. H., & Frary, R. B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple-choice tests. *Journal of Educational Measurement*, 14, 313-321.
- Diamond, J. J. (1975). A preliminary study of the reliability and validity of a scoring procedure based upon confidence and partial information. *Journal of Educational Measurement*, 12, 129-133.
- Diamond, J., & Evans, W. (1973). The correction for guessing. *Review of Educational Research*, 43, 181-191.
- Downing, S. M. (2003). Guessing on selected-response examinations. *Medical Education*, 37, 670-671.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93-103.
- Frary, R. B. (1989). The effect of inappropriate omissions on formula scores: A simulation study. *Journal of Educational Measurement*, 26, 41-53.
- Frary, R. B., Cross, L. H., & Lowry, S. R. (1977). Random guessing, correction for

- guessing and reliability of multiple-choice tests scores. *Journal of Experimental Education*, 46, 9-15.
- Fujimori, S., & Nakano, K. (1994). Effect of test directions upon examinees' test-taking behaviors and performance. *Japanese Journal of Educational Psychology*, 42, 455-462.
- Holzinger, K. J. (1924). On Scoring Multiple Response Tests. *Journal of Educational Psychology*, 15, 445-447.
- Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Washington, DC: American Council on Education.
- Lord, F. M. (1975). Formula scoring and number-right scoring. *Journal of Educational Measurement*, 12, 7-11.
- Mattson, D. (1965). The effects of guessing on the standard error of measurement and the reliability of test scores. *Educational and Psychological Measurement*, 25, 727-730.
- Quereshi, M. Y. (1974). Performance on multiple-choice tests and penalty for guessing. *Journal of Experimental Education*, 42, 74-77.
- Rowley, G. L., & Traub, R. E. (1977). Formula scoring, number-right scoring, and test-taking strategy. *Journal of Educational Measurement*, 14, 10-21.
- Sabers, D. L., & Feldt, L. S. (1968). An empirical study of the effect of the correction for chance success on the reliability and validity of an aptitude test. *Journal of Educational Measurement*, 5, 251-258.
- Thorndike, R. L. (Ed.) (1971). *Educational measurement*. Washington, DC: American Council on Education.
- Thurstone, L. L. (1919). A method for scoring tests. *Psychological Bulletin*, 16, 235-240.
- Traub, R. E., & Hambleton, R. K. (1973). Note of correction on the article entitled "The effect of scoring instructions and degree of speededness on the validity and reliability of multiple-choice tests". *Educational and Psychological Measurement*, 33, 877-878.
- Traub, R. E., Hambleton, R. K., & Singh, B. (1969). Effects of promised reward and threatened penalty on performance of a multiple-choice vocabulary test. *Educa-*

tional and Psychological Measurement, 29, 847-861.

Waters, C. W., & Waters, L. K. (1971). Validity and likability ratings for three scoring instructions for a multiple-choice vocabulary test. *Educational and Psychological Measurement*, 31, 935-938.

Wood, R. (1976). Inhibiting blind guessing: The effect of instructions. *Journal of Educational Measurement*, 13, 297-308.

