

論文名稱：以最大測驗訊息量決定通過分數之研究

校所系別：國立政治大學教育學系

指導教授：余民寧博士

研究生：謝進昌

## 論文摘要內容

本研究目的，乃在運用試題反應理論中最大測驗訊息量的概念於精熟標準設定上作為探討的主軸，透過其歷史的演進與發展，衍生出詮釋本研究最大測驗訊息量法的三個面向，分別為：元素的搭配組合與調整、廣義測驗建構流程、多元效度等，並以此概念賦予解釋運用最大測驗訊息量於精熟標準設定時的合理性與適切性。同時，確立最大測驗訊息量法於公式意涵、試題選擇與統計考驗力面向的合理性，建立其於精熟標準上的理論基礎，而後，再輔以精熟/未精熟者分類一致性信度值以期提供多元效度證據。最後，探討測驗分數的轉換方法、差異能力描述，期能同時獲得量與質的測驗結果解釋。

綜整分析，可發現以下幾點結論：

- 一、運用最大測驗訊息量法於精熟標準設定時，在分類的信度指標上，顯示由此求得精熟標準，經交叉驗證後，大致可獲得滿意的結果，皆有高達九成以上的精確分類水準，且藉由區間的概念亦能充分顯現出，以最大測驗訊息量法求得之標準，可作為專家設定精熟標準時參考、判斷的優勢。而在分數轉換上，不論搭配換算古典測驗分數法或測驗特徵曲線構圖法時，其分類精熟/未精熟者的一致性表現，大致可獲得滿意的結果，乃是值得參照的組合策略。
- 二、在運用定錨點以解釋由最大測驗訊息量法於國中基本學力測驗求得之精熟標準時，可發現未精熟者乃僅需具備學科基礎知識與簡易圖示理解能力，而對於精熟者而言，則需進一步擁有對於廣泛學科知識的了解；複雜問題、資料與圖表詮釋；邏輯推理、分析實驗結果以獲得相關論點等能力，或者更高階之具備進階學科知識；綜合、評鑑資料、情境傳遞之訊息的能力。
- 三、探討測驗長度因素時，分析結果顯示不論採行最大測驗訊息量法、換算古典測驗分數法或是測驗特徵曲線構圖法，皆受此因素的影響，顯示測驗長度愈長，分類一致性愈高，此項結果乃與過去大多數的研究一致。另，由本資料分析結果乃建議測驗長度 20 題時，會是必備的基本題數要求值。此外，若從細部精確錯誤分類人數角度分析時，於實務用途上，可發現對於影響轉換分數時，產生差異分數的因素，決策者並不容易掌握與控制，但卻可藉由增加測驗長度，分散分數點的人數，以彌平錯誤分類的影響。

四、探討測驗異質性因素時，最大測驗訊息量法因具有因試題參數而調整估計受試者能力的特性，使得在異質測驗時，分類一致性仍能維持在不錯的水準之上。反觀換算古典測驗分數法與測驗特徵曲線構圖法，在固定精熟標準下，則有明顯的錯誤分類比率，此現象也反應出現行以固定 60 分作為及格(精熟)標準的缺失。

五、探討採用簡易測驗、困難測驗或常態測驗間於轉換分數上之效果時，由換算古典測驗分數法或測驗特徵曲線構圖法轉換來自最大測驗訊息量法之精熟標準時，資料分析結果顯示，不論於何種測驗難度類型中，採用何種轉換方式，並不會嚴重影響轉換分數間一致性分類的效果。另，若從細部精確錯誤分類人數角度分析時，本研究所採之最大測驗訊息量法，因具備隨測驗難易程度來決定門檻的特性，於簡易測驗中求得之精熟標準較低，而於困難測驗中求得之精熟標準相對較高，使得於轉換分數上，即使有較大的差異分數，亦不會造成嚴重的錯誤分類人數。

六、在探討測驗長度、測驗異質性因素與定錨點題目篩選間互動關係時，分析結果顯示，測驗長度與測驗異質性，並非是絕對影響定錨點題目篩選的因素，更重要的在於最大試題訊息量所對應之最適能力值是否能與定錨點相搭配。

綜整之，本研究所採最大測驗訊息量法，經檢驗後，於分類一致性上乃具有不錯的表現，且搭配相對強韌、嚴謹的理論支持與適切測驗結果解釋方法等，是最適合用於大型考試上使用。因此，乃建議未來政府單位或實務工作者於進行大型證照、資格檢定考試時，可考慮使用本策略。

關鍵字：最大測驗訊息量法、換算古典測驗分數法、測驗特徵曲線構圖法、定錨點、精熟標準設定、精熟測驗

# **Study of the Standard Setting by the Maximum Test Information**

**Jin-Chang Shieh**

**Department of Education**

**National Chengchi University**

**Min-Ning Yu, Advisor**

## **Abstract**

The purpose of this study is to adopt the concepts of IRT maximum test information to standard setting. At first, we are trying to discover three facets of interpretation in using the maximum test information to standard setting through the historical movement of standard setting. The three facets are component combination and adjustment, generalized test construction processes and multiple validities. Depending on these three concepts, we can easily explain the reasonableness and appropriateness of maximum test information approach. After that, we further investigate the reasonableness from the dimensions of definition of formula, item selections and statistical power to establish the basic theory of the maximum information approach in standard setting. In addition, we also examine the effects on exact classification of master/non-master in expectation to provide multiple evidences for validity. Finally, the method of classical test scores transformation and difference ability description are discussed to provide quantitative and qualitative test result interpretation simultaneously.

In sum, some conclusions are proposed.

1. In applying the maximum test information approach to standard setting, the effect on exact classification of master/ non-master may come to a satisfying result. We may have at least 90% exact classification performance. At the same time, we also find that the mastery standard deriving from the maximum test information approach may have some advantages being a starting reference point for experts to adjust on the basis of the view of confidence interval. In the aspect of classical test scores transformation, no matter what approach you take, the transformed classical test scores approach or test characteristic curve mapping method, the consistency of exact classification of master/ non-master may hold. We may suggest the combination strategy is really worthy to take into consideration in standard setting.
2. In applying the anchor point to interpret Basic Competency Test result, we may

find non-master only has basic academic knowledge and simple graph understanding ability, but for the master, he may need extensive academic knowledge; ability of complicated problems, data and graph interpretation; logic reasoning analyzing experimental result to get related issues. Moreover, advanced academic knowledge; ability of synthesizing and evaluating information from data and surroundings are also included.

3. In the aspect of test length, the result of this research shows no matter what approach you take, maximum test information approach, transformed classical test scores approach or test characteristic curve mapping method, they are all influenced. It shows the longer test length, the higher consistency of exact classification of master/non-master. This result is consistent to most of the studies in the past. On the other hand, we suggest the 20 items is a fundamental value. Moreover, from the view of exact number of error classification, we can find that the real factor affecting the difference scores in transforming classical test score is unable to control in practical usage, but we can just disperse the numbers of people in each test score point to reduce the influence of error classification by increasing test length.
4. In the aspect of diverse test difficulty, because the maximum test information approach possesses the characteristic of examinees' ability adjustment depending on item parameters, it is less influenced to maintain a acceptable level of consistent classification. In contrast with the maximum test information approach, the transformed classical test scores approach and test characteristic curve mapping method may have obvious high ratio of error classification under the fixed mastery standard. This also reflects the deficiency of current fixed 60 points passing scores.
5. In the aspect of analyzing the effect of score transformation between easy, hard and normal test, this research shows no matter what approach you take in any type of test difficulty, they may not severely influenced. Furthermore, from the view of exact number of error classification, because the maximum test information approach possesses the characteristic of deciding passing level depending on the degree of test difficulty (the lower mastery standard in easy test and the higher in hard test), it may not lead to a severe error classification even if there exists a large difference score in classical test score transformation.
6. In the aspect of interaction between test length, diverse test difficulty and anchor

items selection, this research shows that test length and diverse test difficulty are not the real factors affecting anchor items selection. The more accurate cause is if the mastery standard deriving from the maximum test information approach may coordinate with the anchor point or not.

In sum, the maximum test information approach may not only lead to a satisfying exact classification performance after analysis, but also be supported by strong and strict theory and accompany proper test result interpretation method. It is the most proper method in standard setting for large-sized test. Finally, we suggest the government or practitioners may consider adopting this strategy for future usage.

**Key words:** maximum test information approach, transformed classical test scores approach, test characteristics curve mapping method, anchor points, standard setting, mastery test