從難度觀點探討國中基測新制量尺

黃國清
國立高雄師範大學教育學系博士候選人

摘　要

國中基測為了呼應各界的要求，從2009年起採用的國中新制基測量尺分數計算方式，在實施的第一年仍然引起社會對於量尺分數的諸多質疑。本文從測驗的難度觀點，採文獻分析法進行探討，針對該新制的量尺分數計算方式的相關議題，例如：試題難度在新制基測量尺計分公式之重要性、如何區分後端考生的能力、量尺分數之轉換是否為線性轉換、提升題庫難度參數的精確度等面向作深入地剖析。本研究結果發現：1.試題難度在98新制量尺的計分公式，扮演重要的角色。 2.須考量測驗難度的條件下，每科加20分才能區分後端考生的能力。 3. 在考慮測驗難度與試題數的情況下，量尺分數之轉換才完全為線性轉換。4. 改善題庫難度參數的精確度，能提高等化的正確性。

最後本文根據上述研究的結果，分別向教育部及國中基測工作推行委員會提出具體之建議。

關鍵詞：國中基測、量尺分數、入學考試

Examining the New Formula of Calculating Scale Score for the BCTEST 2009 in Taiwan from the Perspective of Item Difficulty

Guo-Chin Huang
Doctoral Student, Department of Education, National Kaohsiung Normal University

Abstract

The formula for calculating scale score of the BCTEST(Basic Competence Test for Junior High School Students)was changed in 2009. After the launch in June, the score calculated by the new formula was doubted in Taiwan. This study discusses some topics about the formula and tries to find the ways promoting from a view of item difficulty. The research findings as follows: 1. The item difficulty plays an important role for this formula. 2. The discrimination between low achievement examinees is obvious under some circumstances. 3. The linear translation is effected by the difficulty and the number of item. 4. Calibrating the item difficulty in bank is good for equating. Finally, this study also provides some suggestions to the Ministry of Education for future revisions of the BCTEST.

Keywords: BCTEST, scale score, admission examination

壹、緒　論

一、研究背景

教育部為緩解考生全錯或只答對少數題目時，量尺分數皆對應到1分，及單科答錯1題時所得到的分數，與全對的考生所得的60分差異過大之現象，已於2008年 6月5日以新聞稿正式宣布，2009年國民中學學生基本學力測驗採用新量尺計分方式，同時函知各主管教育行政機關轉知轄屬國中、高中職及五專學校，每年攸關全國約30萬考生的國中基測，隨著2008年8月11日的榜單公布，也宣告從2001年到2008年國中基測所採用的舊制量尺分數計算方式，正式走入歷史！

國中基測量尺分數的初始設計精神為「作為門檻使用」，各考科試題皆以中等偏易為組題原則。然而由於分數使用始終無法依照原始設計目標使用，導致雖然原先量尺分數設計、正確及符合學理，但一般民眾因考生全錯或只答對少數題目時，量尺分數皆對應到1分，及單科答錯1題時所得到的分數，與全對的考生所得的60分差異過大等問題而有所質疑（國中基測工作推行委員會，2008b），也引起一些測驗與統計方面專家學者的關心，更有學者的研究結果顯示一些值得深思及改進之處，例如量尺分數的估算與等化的方法與轉化過程、題庫難度指標的準確度…等等，存在許多令人質疑的問題（余民寧，2006；林妙香，2007）。為因應社會各界對國中基測量尺分數的關切，教育部委請國立臺灣師範大學心理與教育測驗研究發展中心邀集統計及測驗學者專家等，深入研議國中基測量尺計分方式，基於分數使用短期內難以改變，與維持各科測驗難度中等偏易及兩次測驗分數須擇優使用等前提下，透過實徵分析，比較各種可能的計分方式之特性及優缺點，透過多次諮詢會議徵詢專家學者及相關團體針對不同量尺計分方式分析結果的看法，大多數支持下列的量尺分數改變（教育部，2008）：

（一）將深奧難懂的「非線性轉換」方式，改為「線性轉換」，減少全對與錯一題間的分數差距。

（二）每科加20分以墊高底部分數，對原先分數皆為1 分者，新量尺分數則可以看出其學習成果的差異，以區分低成就考生的能力

從2009年開始使用的國中基測新制量尺計分公式，乃是教育部對社會大眾的善意回應，將量尺計分方式採行另一種學理，以符合大多數家長、老師們所認知的傳統測驗理論（Classical Test Theory, CTT），而這個新制量尺計分公式中，每位考生只要知道每科的「平均答對題數」後，便可自行計算出該科的量尺分數了，因此「平均答對題數」也就是這個公式的最主要變數了，就CTT而言，「平均答對題數」相當等同於一份測驗的「難度」（difficulty）（郭生玉，1990；Anastasi, 1990），因此這個新制的量尺計分公式便隱含測驗難度的重要性。

新制的量尺計分方式既然是為了達成兩項主要目標：（一）將量尺計分方式，改為「線性轉換」。（二）每科加20分以墊高底部分數，以區分低成就考生的能力。依前述所言，測驗的難度乃為新制量尺計分公式之關鍵所在，故測驗難度勢必會影響上述目標之達成與否！再進一步來說，如何控制測驗的難度才能達成上述之目標呢？這些都是值得深入探討議題。

另外，對於兩次基測量尺分數之間的等化問題，新制量尺計分方式仍然採用舊制的方式，也就利用Rasch模式的能力值來轉換，其中考生的能力值估計亦牽涉到試題的難度，如何增進能力值估計的精確度，亦值得加以討論。

二、研究目的

綜上所述，本文將針對國中基測新量尺計分公式，從「難度」的觀點加以探討，本研究目的如下：

（一）了解「測驗難度」在新制國中基測新量尺計分公式之重要性。

（二）探討每科加20分能否區分後端考生的能力並提出最佳模式。

（三）剖析量尺分數之轉換是否必為線性轉換以及尋求改善方法。

（四）增進兩次量尺分數等化之精確度。

貳、文獻探討

一、2001年舊制的量尺計分公式

舊制的量尺計分理論是依據Lord（1965）所提出的強真分數模式（strong true score model），其計分過程如下（涂柏原，2002）：

假定一個測驗有K個試題，一個考生的原始分數（raw score）是這K個二元計分（dichotomously scored）試題得分的和。假定X代表原始分數，那麼在一個母群體X的值等於 i (i =0, 1, ...,K)的機率函數是

$Pr(X=i) = \int Pr(X=i \mid \tau)\, g(\tau)\, d\tau$

其中 τ 是母群的答對率真分數（proportion-correct true score），其機率密度函數為$g(\tau)$，而$Pr(X=i \mid \tau)$則是以真分數為 τ 的考生所形成的原始分數之條件分配（conditional distribution）。

而其中$g(\tau)$與$Pr(X=i \mid \tau)$的分配可依照強真分數模式（strong true score model）來加以估計，測驗界的專家學者大多接受強真分數模式（strong true score models）；也就是母群的答對率真分數的分配$g(\tau)$是服從四參數的beta分配，而以真分數為 τ 的考生所形成的原始分數之條件分配$Pr(X=i \mid \tau)$，則為複合二項誤差模式（compound binomial error model）。因此強真分數模式就被許多學者稱為four-parameter beta compound binomial model，或是直接叫做beta-binomial model。

Carlin與Rubin（1991）提出了廣義的（extended）beta-binomial分配以簡化整個模式的估計。為了解決條件測量標準誤不等的問題，國中基測推委會也就是利用Freeman和Tukey (1950)所描述的正弦反函數轉換(arcsine transformation) 來穩定（stabilize）誤差變異數。這個轉換的型式是

$$c(i)=\frac{1}{2}\left\{\sin^{-1}\sqrt{\frac{i}{K+1}}+\sin^{-1}\sqrt{\frac{i+1}{K+1}}\right\}$$

其中$\sin^{-1}$是正弦反函數。這個轉換能將靠近中間的量尺加以壓縮（compressing），並將兩端的量尺拉長開來（stretching）。首先我們得先將原始分數用上述Freeman和Tukey (1950)所提的公式加以轉

換以得到c(ｉ)，（ｉ=1，2，….，Ｋ），然後將c(ｉ)、Pr（X=ｉ）、g（τ）及Pr（X=ｉ｜τ）之估計數代入公式來求量尺分數的平均數與變異數、量尺分數的條件平均數與變異數，然後再求得考生的平均量尺分數誤差變異數以及量尺分數的信度係數等。到此，所算出來的量尺分數以及其平均數或變異數等，再用下面的公式來轉換：

$$s^* \equiv s^*[c(i)] = \frac{\sigma_s{}^*}{\sigma(Ec)}\{c(i) - \mu[c(x)]\} + \mu_s^*$$

其中μ_s*及σ_s*分別是平均數（30）及測量標準誤（30÷4=7.75），s*即是我們最後想要得到的量尺分數。

由上面的說明，那就是說明必須等到收集考生作答的資料之後，分數轉換公式才能確定，這也說明了每一年基本學力測驗分數的轉換公式可能會有些許的不同。

二、2009年新制的量尺計分公式

新制量尺分數係將考生在各科的答對題數以線性轉換方式來建立，公式如下：

$$s^* \equiv s^*[x] = A\{x - \mu(X)\} + \mu_s^* \quad ,$$

$$A = \frac{60 - \mu_s^*}{K - \mu(X)} \quad ,$$

其中 μ_s^* 是預設的測驗分數平均數(定為30)；K 為各科測驗的題數；而 $\mu(X)$ 為答對題數的平均數。上述轉換後的分數再加上常數20，加以裁截和四捨五入後即得新量尺分數，因平移20 分的效果，各科新量尺計分結果平均分數從30 變為50，最高分數由60 變為 80。

各科新量尺計分方式改為線性轉換，轉換後分數（包含負分）全部加上20 分，分數區間為1-80，各科新量尺計分結果平均分數從30 變為50，最高分數由60 變為 80。新計分方式可減緩一般民眾因現行量尺考生全錯或只答對少數題目時，量尺分數皆對應到1分(負分皆算1分)的問題，因為加上

20 分後，原先的負分多數可以得到正值，因此考生的表現可以充分呈現。新量尺分數也可以解決以往舊量尺分數單科答錯1 題時所得到的分數，與全對的考生所得的60 分差　過大等問題之疑慮。

經由上述，可以將新、舊兩種計分的公式（第一次學測之量尺分數所採用）約略區分為，2001年舊制是以「試題反應理論」為基礎的計分方式；2009年新制則是採行「古典測驗理論」來計算量尺分數的。

參、研究方法

本研究係採文獻分析法（documentary analysis），以2008年第一次國中基測之資料為研究文獻，並運用測驗理論及數學計算之演繹推理（王文科，2008），推論出符合研究目的之模式，其研究對象、工具、設計與程序如下：

一、研究對象

本研究之研究對象為參加2008年第一次國中基測之考生，報名參加考試總人數共有321,190人（國中基測推委會，2008a）。

二、研究工具

本研究之工具為2008年第一次國中基測之各科試題，計有國文（共48題）、英語（共45題）、數學（共34題）、自然（共58題）及社會（共63題）。

三、研究設計與程序

本研究之程序如圖1所示，研究設計如下述：

（一）分析2008年第一次國中基測之資料，並觀察比較以新制與舊制的計分公式所得量尺分數之差別。

（二）根據新制量尺分數，驗證是否達成新制量尺計分方式之目的。

（三）以CTT之難度為基礎，運用數學演算推論達到新制度目的最佳模式 。

（四）分析兩次基測量尺分數等化之程序及其與提升能力值估計之精準度。

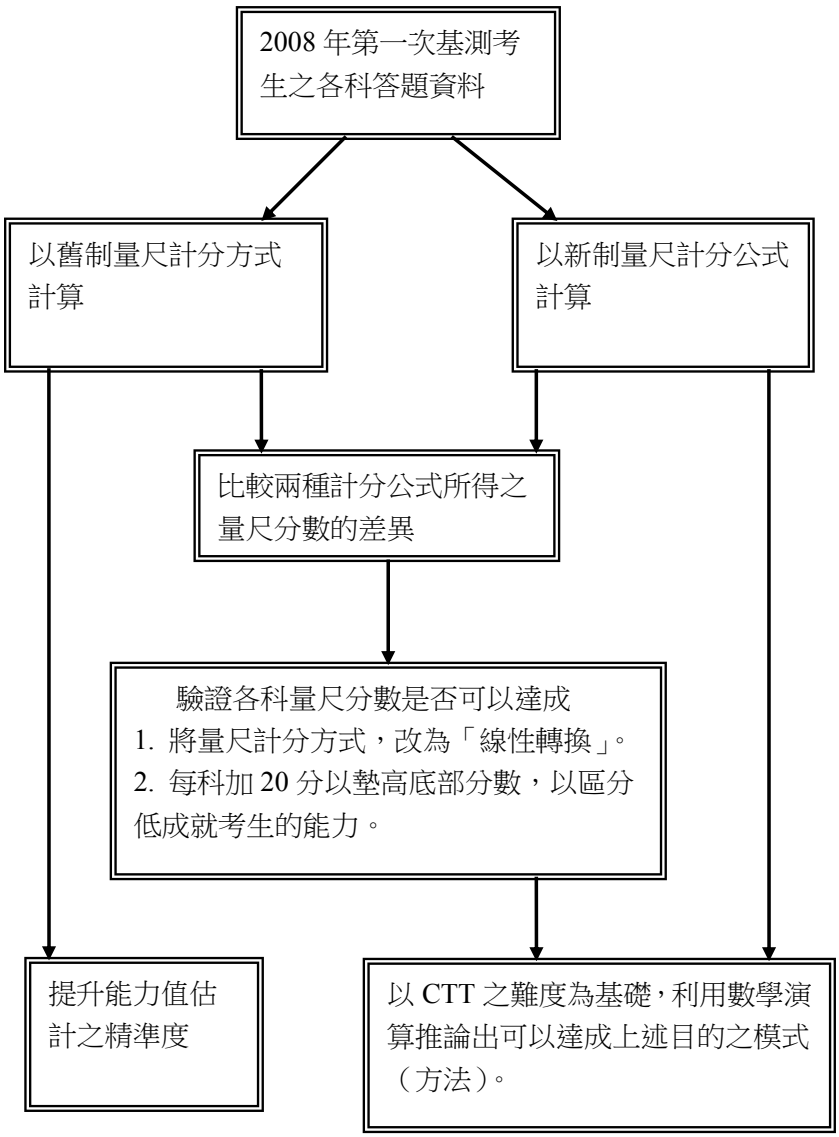

圖1 研究設計與流程圖

肆、過程與結果

一、新制量尺公式與測驗難度

首先，對於新制量尺計算的公式，教育部新聞稿解說如下（教育部，2008）：

> 新量尺分數係將考生在各科的答對題數以線性轉換方式來建立，公式如下：
>
> $$s^* \equiv s^*[\mathrm{x}] = \mathrm{A}\{\mathrm{x} - \mu(X)\} + \mu_s^* \qquad (1)$$
>
> $$\mathrm{A} = \frac{60 - \mu_s^*}{K - \mu(X)} \qquad (2)$$
>
> 其中 μ_s^* 是預設的測驗分數平均數(定為30)；K 為各科測驗的題數；而 $\mu(X)$ 為答對題數的平均數。上述轉換後的分數再加上常數20，加以裁截和四捨五入後即得新量尺分數。

以上新制的量尺分數之計算式子，吾人可以改寫成如公式（3）：

$$\text{新制量尺分數} = 30 \times \left(\frac{\text{考生個人答對題數} - \text{全體平均答對題數}}{\text{測驗總題數} - \text{全體平均答對題數}} \right) + 30 + 20 \qquad (3)$$

從公式（3）可以清楚看出：如果考生想要計算出自己本身的量尺分數，考完自行核對答案，「考生個人的答對題數」是已知的，另外「測驗的總題數」也是已知的，所以就整個公式而言，只有「全體的平均答對題數」是未知的，而「全體的平均答對題數」的多寡與「測驗的難度」有關，所以可以說明「測驗的難度」是整個新制量尺計分公式之關鍵；再從另外一個觀點來討論，就國民中學學生基本學力測驗工作推動委員會（以下簡稱：國中基測推委會）的立場來說，「考生個人的答對題數」是無法左右的，但是「測驗的總題數」以及「全體的平均答對題數」是可控制的，而這兩者也就是「測驗的難度」的概念，因此「測驗的難度」在新制量尺計分公式之重要性，就不言可喻了。

二、區分後端考生能力與測驗難度設計

（一）墊高量尺分數底部20分的意義與目的

以2008年第一次基測為例，依據舊制量尺的計算分數，如果自然科答對7題以下、國文科答對11題以下、英文科答對13題以下、社會科答對17題以下，這些情形在舊制量尺分數的計算下，都是以1分計（有可能分數是負的），如表1所示亦，即答對這些題數以下的考生能力或學習成就是無法區分的；而採用2009年所實施之新制基測量尺計分的特點，其中之一就是試圖將量尺分數底部墊高20分，如此一來便可將程度屬於後段的考生能力區分出來。

以舊制量尺計分方式而言，考生必須在某一科至少答對若干的題數後（例如自然科須答對7題以上），其能力或學習成就始能以量尺分數來作區分者，吾人稱這『若干題數』為『區辨題數』；從表1觀察得知，在新制量尺的計算方式下，自然科只要答對1題以上，每種答對題數的分數都會得到不一樣的分數，這是符合新制量尺計分的最理想模式，如此一來就達到新制量尺計分墊高底部分數的目的了，但是反觀在社會科、國文科以及英語科，新量尺方式卻還是只能改善少部分而已，例如社會科仍然要答對14題以上，國、英兩科也要答對7題以上才能區分低學習成就考生的量尺分數（國中基測推委會，2008a）。

表1 97年第一次國中基測各科答對題數與新舊量尺分數對照表

	社會科		國文科		英語科		自然科		數學科	
答對題數	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺
0	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	6	1	1
2	1	1	1	1	1	1	1	7	1	1
3	1	1	1	1	1	1	1	8	1	4
4	1	1	1	1	1	1	1	10	1	6
5	1	1	1	1	1	1	1	11	1	8

表1 97年第一次國中基測各科答對題數與新舊量尺分數對照表 (續)

	社會科		國文科		英語科		自然科		數學科	
答對題數	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺	舊量尺	新量尺
6	1	1	1	1	1	1	1	12	1	11
7	1	1	1	2	1	2	1	13	1	13
8	1	1	1	3	1	4	2	15	3	16
9	1	1	1	5	1	6	3	16	5	18
10	1	1	1	7	1	8	4	17	7	21
11	1	1	1	9	1	10	6	19	9	23
12	1	1	2	11	1	12	7	20	11	26
13	1	1	3	13	1	14	8	21	13	28
14	1	3	4	15	3	16	9	23	15	31
15	1	4	6	17	4	18	10	24	16	33
16	1	6	7	19	6	20	11	25	18	36
17	1	8	9	21	7	22	12	26	20	38
18	2	9	10	23	8	24	13	28	22	41

由上述可知，以2008年第一次國中基測的實徵資料來驗證新制量尺計分公式，其在社會科、國文科及英語科結果，似乎很難說服社會大眾相信「底部墊高20分可以紓解低分端考生得到相同量尺分數」的目的了！

因此，思考如何藉由測驗理論性質的運用，使新制量尺計分方式的『區辯題數』降到最低，如此才符合這次新制量尺計分方式改革之主要內涵。

（二）『區辨題數』與測驗難度、題數之關係

新制量尺計分公式，乃訂定當考生答對所有考題時，可得滿分80分、當考生答對題數剛好等於「全體平均答對題數」者，得50分，但是這個計分公式如果不考慮任何測驗的難度下，卻變成是一個「無法確定答對0題一定是0分」以及「答對不同題數可能都是0分」的設計而已，因此如何改善上述兩種情形，則需要從測驗的編製技術上加以調控，首先觀察由公式（1）改寫成公式（4）

$$30\times\frac{X-u(X)}{K-u(X)}+30+20 \qquad (4)$$

其中X表示「考生答對題數」、u(X)表示「全體考生的平均答對題數」、K表示「該科總題數」。公式（5）中u(X)「全體考生的平均答對題數」與某科的測驗難易度有關，若某科難度低、較簡單，則該科的平均答對題數就會較高；反之，難度高、較困難，則某科的平均答對題數就會較少而公式（1）亦可改以公式（4），而公式（4）最後加上去的20，就是所謂的每科墊高的20分 ，因為教育部與國中基測推委會一再強調基測試題的難易度都是中間偏易，大多設定答對率介在於 .5至 .75之間（教育部，2007），所以假設某一科的測驗難度落於 .5到.75之間，再根據該科目的總題數，將上述兩個條件帶入公式（4）計算，可以得到量尺分數如表2，舉例而言：

社會科總題數有60題，該科測驗難度為.70，故「全體考生的平均答對題數」為42題，假如考生答對0到12題，代入上述公式，則其新制量尺分數為「－20分」到「0分」，計算過程如下：

答對0題者得 $30\times\frac{0-42}{60-42}+30+20=-20$ ；

答對12題者得 $30\times\frac{12-42}{60-42}+30+20=0$

所以答對0到12題，新制的量尺分數還是都為1分（基測量尺分數0分以下，皆以1分計），答對13題後才會有不同的量尺分數。

亦即當公式（4）中 $30\times\frac{X-u(X)}{K-u(X)}<(-50)$ 時，新制的量尺分數就會成為負數，也就是量尺分數皆為1分，所以假若吾人可以在編製某一科測驗時，決定總題數之後，便可以根據表2來決定測驗的難度，例如為緩解低分端考生得分對應到相同量尺分數的情況，則『區辨題數』要儘量少，無論總題數為何，若將測驗難度控制在.60以下，則可以真正達到緩解低分端考生得分對應到相同量尺分數的目的，但是在實務上，整份測驗之難度

要在.60以下，似乎有違國中基測推委會的目的，所以可以考慮設定在.60到.65之間，除了測驗編製的技術上應該不會有問題外，亦有舒解後端考生產生同分的效果。

表2 區辨題數與測驗試題數、難度對照表

題數＼難度	p=.75	p=.70	p=.65	p=.60	p=.50
64	22	13	5	0	0
60	21	13	4	0	0
56	19	12	4	0	0
48	17	10	4	0	0
36	12	8	3	0	0
32	10	7	3	0	0

三、試題總數、測驗難度與線性轉換

（一）「直線轉換」與「非直線轉換」之意義

新制基測量尺分數計算方式，也已經將深奧難懂的IRT能力值及反正旋函數法的「非線性轉換」方式（涂柏原，2007），改為「線性轉換」，本文先以淺顯的圖解方式來說明「非線性轉換」與「線性轉換」，圖2 即為分數的「非線性轉換」，相對而言，圖3便是「線性轉換」。

舉例而言，「非線性轉換」就是某科考生「甲答對15題、乙答對16題」與「丙答對30題、丁答對31題」，兩者雖然都是只差了一題，可是甲、乙兩人的量尺分數可能只差1分甚至同分，相對的丙、丁兩人可能就會差了3、4分，如圖2所示；另外「線性轉換」的意思就是上述情形中，甲乙兩人的量尺分數差異，會跟丙丁一樣，如圖3所示，「線性轉換」也就是說明在同一門考科內，每一題的重要性都是一樣，亦即『題題等值』。

再者，就新制量尺計分公式來討論，從線性圖形之方程式（或稱線性函數）來看，上述之公式（1）及公式（2）中

$$s^* \equiv s^*[\mathrm{x}] = \mathrm{A}\{\mathrm{x} - \mu(X)\} + \mu_s^* \qquad (1)$$

$$A = \frac{60 - \mu_s^*}{K - \mu(X)} \qquad (2)$$

對照直線方程式（函數）的一般式 $y = ax + b$ 來看，公式（1）中的A即可視為斜率a，而 μ_s*可看作是截距b，s* 為應變數y ，x- u(X) 意即自變數 x。

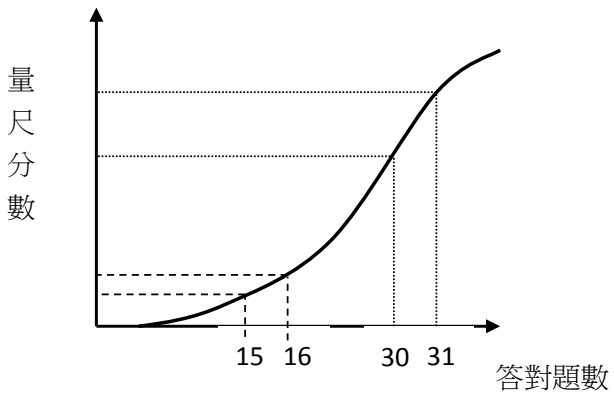

圖 2 非線性轉換

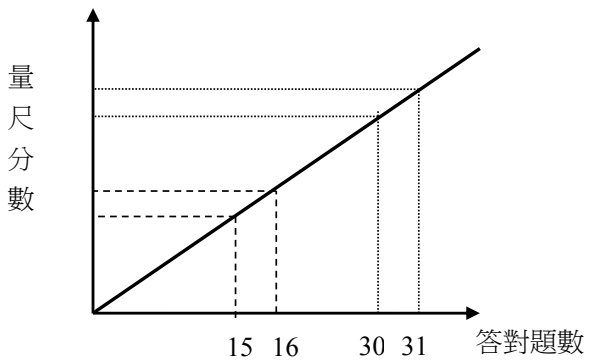

圖 3 線性轉換

（二）測驗的每題分數與測驗難度之關係

新制的量尺計算，雖然『題題等值』，不過在上述數學公式（4）的計算結果，應該都會有小數點的產生，所以若是經過四捨五入後，每題差距的量尺分數可能就不一定一樣了；甚至也可能出現答對不同題數，還是會有同分的情形。

由公式（4）可知每答對一題的分數為 $\frac{30}{K-u(X)}$ （舉例說明，比如考生答對16與15題之量尺分數差距的計算過程為：

$$\left[\ 30\times\frac{16-u(X)}{K-u(X)}+30+20\ \right]-\left[\ 30\times\frac{15-u(X)}{K-u(X)}+30+20\ \right]=\frac{30}{K-u(X)}\ )$$

又考慮基測的測驗難度p須介於 .5到 .75之間， 所以「全體平均的答對題數」u(X)（難度×題數） 介於0.5K到0.75K之間，且K－u(X)介於0.5K到0.25K之間，亦即0.5K＞K－u(X)＞0.25K，因此

若30≦K－u(X)，則 K ≧61；若30≧K－u(X)，則K≦61

因此故，必須要舉題數大於61題的科目來加以討論，而歷屆基測試題各科題數最多者，同時也超過60題者，就只有社會科而已，故舉某年基測社會科共有68題為例，同時考慮測驗難度介於.5到 .75之間，討論如下：

1. $\frac{30}{K-u(X)}$ ＝1 時，例如社會科共有68題，測驗難度p≒.56時，K=68、u(X)=38（≒68×0.56），每一題分數都為1分。

2. $\frac{30}{K-u(X)}$ ＜1時，亦即 K－u(X)＞30，例如社會科共有68題，測驗難度p=0.5時，K=68、u(X)=34（=68×0.5），每一題分數為0.88分，每一題差距會低於1分。

3. $\frac{30}{K-u(X)}$ ＞1時，亦即 K－u(X)＜30，例如社會科共有68題，測驗難度p=0.65時，K=68、u(X)=44（≒68×0.65），每一題分數為1.25，每一題差距會高於1分。

以上在三種不同的測驗難度時，若只考慮考生「答對題數」超過「平均答對題數」的情形（「答對題數」小於「平均答對題數」的情形，則以負數表示），答對題數」每超過「平均答對題數」一題，則每題差距分數為 $\frac{30}{K-u(X)}$ ，超過N題，即為 $\frac{30N}{K-u(X)}$ ；但是因為量尺分數是以整數形式標記，因此可能因為「每題差距分數」的加總後，四捨五入而產生答對不同題數，卻得到相同的量尺分數，表3是「答對超過平均答對題數N題」與「每題分數差距」之量尺分數對照。

表3 超過平均答對題數、每題分數與量尺總分對照表

每題分數 / 超過平均題數N	0.6	0.8	1	1.2	1.4
1	50.6 （**51**）	50.8 （51）	51	51.2 （51）	51.4 （51）
2	51.2 （**51**）	51.6 （**52**）	52	52.4 （52）	52.8 （53）
3	51.8 （**52**）	52.4 （**52**）	53	53.6 （54）	54.2 （54）
4	52.4 （**52**）	53.2 （53）	54	54.8 （55）	55.6 （56）
5	53 （53）	54	55	56	57
6	53.6 （**54**）	54.8 （55）	56	57.2 （57）	58.4 （59）
7	54.2 （**54**）	55.6 （**56**）	57	58.4 （58）	59.8 （60）
8	54.8 （**55**）	56.4 （**56**）	58	59.6 （60）	61.2 （61）

由表3可歸納出：

當「每題差距分數等於1分」時，若以「超過平均答對題數」為橫座標、量尺分數為縱座標，可以畫出其圖形應為斜率為1的直線，其答對題數與量尺分數為線性轉換，如圖4所示。

當「每題差距分數低於1分」時，若以「超過平均答對題數」為橫座標、量尺分數為縱座標，畫出之圖形為折線，其答對題數與量尺分數為非線性轉換，如圖5所示。

當「每題差距分數高於1分」時，若以「超過平均答對題數」為橫座標、量尺分數為縱座標，畫出之圖形亦為折線，惟較圖5之折線高陡，其答對題數與量尺分數並非線性轉換，如圖6所示。

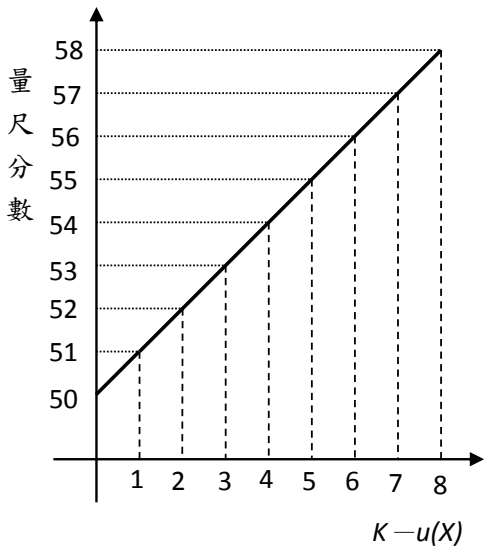

圖4 線性轉換

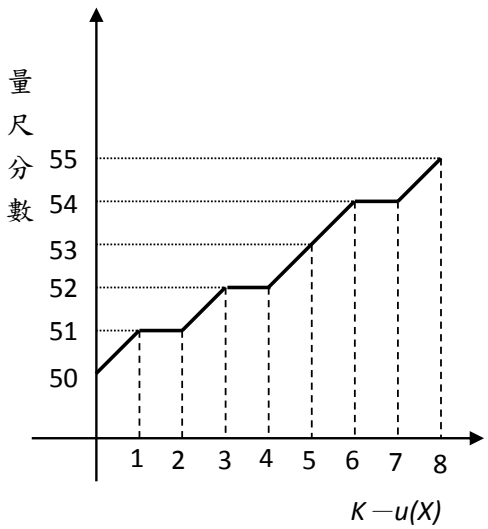

圖5 非線性轉換

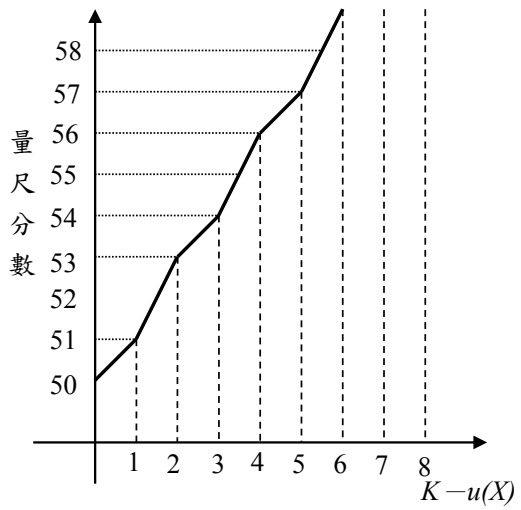

圖6 非線性轉換

（三）試題總數、測驗難度與線性轉換

因此，國中基測的量尺分數若要符合線性轉換的理想、避免以往答對不同題數卻同分的現象，則答對每一題之間的分數差距就一定要為整數，如此不管如何加總，就沒有四捨五入而同分的情形了；為達此目的，吾人可以再從公式（4）加以探討，亦即 $\frac{30}{K-u(X)}$ ＝整數，既可滿足上述情形，又考慮測驗的實務情形，K－u(X)＝30、15、10較為可行，所以當K為已知，只要能掌控測驗難度p（依據國中基測之原則難度p值介於0.5到0.75之間），則 u(X) 亦可獲得掌握，進而控制 $\frac{30}{K-u(X)}$ 為整數，題數K、K－u(X)及難度p三者之間關係如表4所示。

表4 題數K、K－u(X)及難度p三者之關係表

題數K \ K－u(X)	**30**	**15**	**10**
68	0.56		
60		0.75	
48		0.69	
40		0.63	
33		0.55	0.70

註：表格內斜體字表示為測驗難度p

四、試題難度與分數等化

教育部在2008年的新聞稿中，雖然提及了上述公式（1）、（2）之量尺計算方法，但是未強調這個公式僅適用在第一次基測，第二次基測的量尺分數計算方法並不會依據該公式來估計，而是採行舊制的能力值等化。

（一）考生能力值與試題難度之關係

國中基測的考生能力值是依照試題反應理論(IRT)計算出來，並且依據Rasch（1960）所建議，採用省略猜測參數c與鑑別力參數a的單參數對數模式(one-parameter logistic modal)，亦即若已知學生能力θ，試題難度b，可得該名學生答對此題的機率為：

$$P(\theta)=\frac{e^{(\theta-b)}}{1+e^{(\theta-b)}} \qquad (5)$$

在各科的題庫中，已有各道題目的難度，因此可以根據學生的答題反應，估算其能力參數。我們定義概似函數(likelihood function)如下：

$$L(u/\theta)=\prod_{i=1}^{Q}P_i^{ui}Q_i^{1-ui} \qquad (6)$$

其中P_i表答對第i 題的機率，$Qi=1-Pi$表答錯該題的機率，ui只有兩種可能，若該題答對則ui=1，答錯則ui=0； (6)式即代表能力為θ的學生對全部Q題的聯合機率。當我們取得某一位學生的答題反應，而某一能力值θ可以產

生最大概似函數，則此能力值可視為該學生的能力，稱作最大概似能力估計(maximum likelihood estimator of θ)。

（二）兩次國中基測量尺分數之等化過程

國中基測推委會為了使兩次基測分數能比較，也就是兩次分數放在同一把兩尺上來比較，所以還是沿用舊制的IRT能力值的插補法（涂柏原，2003），舉上述數學科為例，其等化過程如下：

1. 國中基測推委會在考試前會依據考題難易度，先行計算出每種答對題數之能力值θ，例如答對22題θ＝1.54（56分）、答對23題θ＝1.61（58分）等各種答對題數之能力值及其量尺分數。

2. 第一次基測考完後，國中基測推委會可以計算出各科每種答對題數之量尺分數，例如答對22題得56分、答對23題58分….等。

3. 在第二次基測考前，國中基測推委會也會依據第二次考題難易度，事先計算出數學科每種答對題數之能力值θ，但不再計算量尺分數。

第二次基測考題較難，某甲答對21題，其能力值θ＝1.59，再對照第一次基測，某甲的能力在第一次基測應該答對題數介於22到23題之間（1.54＜1.59＜1.61），亦即量尺分數介於56到58分之間，利用差補法原理，計算（58－56）×【（1.59－1.54）÷（1.61－1.54）】＋56≒57.42，四捨五入後，某甲的量尺分數為57分。

由上述公式（5）、（6）可知，雖然Rasch模式所估計之能力值，最後似乎只與答對題數有關，但其能力值之估計仍與試題難度b有著密切的關係，尤其在選擇同一科目兩次基測試題時，會從題庫中選擇難度b相近之題目組合，而在編製某科測驗時，其測驗難度的計算，也是經由每一道試題難度及其組合才能算出，所以當某一科測驗難度已知時，各種答對題數之能力值就可以事先計算得知了，因此可知掌握題庫中試題難度的精確度是很重要的，所以進行題庫之試題校準（item calibration ）工作，對於估計試題難度之精確度的提升，是有相當的助益的（余民寧，2008、吳裕益，2008；Kolen, & Brennan, 1995）。

但對於兩次量尺分數之等化能否精準，其最大關鍵在於題庫之難度的

正確性（涂柏原，2003；Wright & Stone，1979），因此，以單參數b難易度為指標的題庫試題，其正確性、精準性更顯得十分重要，也因此有學者質疑以3百人預試的難易度，是否適合30萬人使用（林妙香，2007）。

伍、結論與建議

綜合本文上述之探討結果，無論是在量尺計分公式的探討、區別低學習成就者、量尺分數之線性轉換、能力值估計等，這些本文所探討的面向，總括來說，吾人都可以經由測驗（試題）的難度來調控。

因此，針對本文研究之目的，擬分別陳述以下之結論與具體建議：

（一）測驗難度在新制量尺計分公式的重要性

在「考生個人的答對題數」是已知的，「測驗的總題數」也是已知的情形下，考生若想計算出自己本身的量尺分數，所以就整個公式而言，只有「全體的平均答對題數」是未知的，而「全體的平均答對題數」的多寡與「測驗的難度」有關，所以可以說明「測驗的難度」是整個新制量尺計分公式之關鍵。

從國中基測推委會的立場來說，「考生個人的答對題數」是無法左右的，但是「測驗的總題數」以及「全體的平均答對題數」是可控制的，而這兩者也就是「測驗的難度」的概念，因此，國中基測推委會及相關教育單位應該了解「測驗的難度」在新制量尺計分公式上，扮演一個重要的角色。

（二）底部加20分無法全然區分後端考生能力

教育部與國中基測推委會對於新制量尺分數計分方式，其主要目的之一是藉由底部墊高20分，可以紓解低分端考生得到相同量尺分數，可是國中基測推委會以97年第一次國中基測的實徵資料所驗證的結果，除了自然科符合目的外，其餘各科均未達到理想，尤其社會科差距頗大。因此，除了底部墊高20分外，應該思考如何藉由測驗理論性質的配合運用，才能符

合這次新制量尺計分方式改革之主要內涵。

本文所提及之『區辨題數』，可作為編製某科測驗時之參考指標，控制試題數及測驗難度便可以得到『區辨題數』，『區辨題數』越少，考生得負分數的機會越低，越可以紓解低分端考生得到相同量尺分數的狀況，甚至當『區辨題數』為0時，便無負分情況，就可完全達到理想的模式。

（三）新制量尺分數之轉換並非一定為線性轉換

新制量尺公式確為一對一函數的直線轉換，但是可能因為計算上的四捨五入而產生答對不同題數，卻得到相同的量尺分數， 此產生「答對題數」與「量尺分數」之間的轉換並不是線性轉換，例如本文圖4及圖5所示，其圖形為折線。

為了能解決因四捨五入後而產生的同分現象，建議參酌本文所提，「每題差距分數」宜為整數，因此編製測驗時，善用「測驗難度」及「試題數」之調配，使得「每題差距分數」為整數而達到最佳的直線轉換模式。

（四）兩次基測量尺分數等化之能力值估計

第一、二次的量尺分數計算方式是不同的，第一次是根據新制的量尺分數計分方式，然而第二次是採用IRT之能力值，對照到第一次相同能力值的量尺分數，作直線轉換，亦即第二次的量尺分數計算與舊制的計分方式相同，並無改變。

建議國中基測推委會對於基測題庫的試題難度（b值），宜每年在考完基測後即校準試題參數，以提高基測之能力值估計的精準度；同時也建議國中基測推委會在考慮試題的曝光度的前提下，也能考慮提高每道題目的預試人數，以提高試題難度估計之效度。

參考文獻

王文科、王智弘（2008）。**教育研究法**（12版）。台北：五南圖書公司。

余民寧（2006）。 國中基本學力測驗使用量尺分數之檢討與省思。**教育資料與研究**，68 ，63-77。

余民寧（2008）。**試題反應理論的介紹（十一）--題庫的建立**。檢索日期：2008年11月21日，取自： http://www.edutest.com.tw/e-irt/irt11.htm

吳裕益（2008）。**測驗理論**。國立高雄師範大學教育學系博士班未出版之上課講義。

林妙香（2007）。90-93年度國中基測量尺及等化程序之個案研究。**中國統計學報，45**（4），402-436。

涂柏原（2002）。**國中基本學力測驗量尺分數的說明（中）**。檢索日期：2008年8月15日，取自http://www.bctest.ntnu.edu.tw/flying/flying11-20/flying18-5.htm

涂柏原（2003）。**國中基本學力測驗量尺分數的說明（下）**。檢索日期：2008年8月15日，取自http://www.bctest.ntnu.edu.tw/flying/flying11-20/flying19-9.htm

涂柏原（2007）。國中生基本學力測驗量尺分數轉換之實徵研究。**教育研究學報，41**（1），61-77。

國民中學學生基本學力測驗推動工作委員會（2008a）。**97年第一次國中基測各科答對題數與新舊量尺分數對照表**。檢索日期：2008年9月15日，取自http://www.bctest.ntnu.edu.tw/

國民中學學生基本學力測驗推動工作委員會（2008b）。**精打細算談量尺**。檢索日期：2008年9月15日，取自http://www.bctest.ntnu.edu.tw/documents/98QA_score.pdf

教育部（2007）。**「國中基測貫徹一綱」說帖**。檢索日期：2008年9月15日，取自http://www.bctest.ntnu.edu.tw/flying/flying41-50/44-p2.pdf

教育部（2008）。**98年新量尺計分方式問與答**。檢索日期：2008年8月15日，取自http://www.edu.tw/files/list/B0037/98年新量尺計分方式問與答.doc

郭生玉（1990）。**心理與教育測驗**。台北：精華書局。

Anastasi, A.（1990）. *Psychological Testing.(6th ed)*.New York, NY: Macmillan.

Carlin, J. B., & Rubin, D. B. (1991). Summarizing multiple-choice tests using three informative

statistics. *Psychological Bulletin, 110*, 338-349.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and square root. *The Annals of Mathematical Statistics, 21*, 607-611.

Kolen, M. J. & Brennan, R. J. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag.

Lord, F. M. (1965). A strong true score theory, with applications. *Psychometrika, 30*, 239-270.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: The Danish Institute for Educational Research.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.