輔 導 與 諮 商 學 報
民 99 年，32 卷 1 期，01-16 頁

# 統計顯著性在班級團體輔導研究中之意義

吳耀明

## 摘 要

推論統計的顯著水準概念，一般都訂在.05 或.01，而這種限定是否眞能應用在班級團體輔導或一般諮商的研究上？研究者得考慮採用某一種顯著水準可能導致研究結果決定之錯誤程度，更需深思統計檢定的考驗力高低程度。統計考驗力是考驗假設的一個重要指標，同時它亦有助於決定適當的樣本數目。本文針對第一類型錯誤、第二類型錯誤、統計考驗力、P 值，及樣本大小有關之誤用與迷思提出說明；其次，探究統計考驗力的重要性及其影響因素，接著說明透過 Cohen 氏查表法與統計考驗力分析軟體 G＊Power 進行樣本規劃並實際分析 30 篇碩士論文研究假設的統計考驗力；最後，提供一項實際可行的假設考驗的步驟。

**關鍵詞**：統計考驗力、效果值、量化研究

吳耀明　屏東縣麟洛國小（通訊作者，a7261143@yahoo.com.tw）

# 緒　論

## 一、研究動機

猶記民國八十七年六月，研究者在撰寫碩士論文時，研究參考相關書籍文獻，編擬十次現實治療取向(Reality Therapy Approach)團體輔導活動課程，並經二位有團體輔導活動實務經驗之國小教師提供寶貴意見與修訂，提高其內容效度，最後再由指導教授認可同意，由研究者擔任團體領導者，進行爲期十週的團體輔導活動課程。最後爲了使實驗組與控制組原來存在之誤差獲得控制，以實驗組和控制組之前測分數爲共變量，進行獨立樣本雙因數共變數分析。研究結果顯示，實驗組學生在「國小學童生活適應量表」上的「社會適應」、「功課與遊戲調適」兩分量表上，有顯著性之立即實驗效果；而在「親和力」、「社交技巧」、「社會適應」、及「功課與遊戲調適」等四個分量表上，具有顯著的持續性輔導效果。

當初認爲，團體輔導其偏重於人文感性的研究，似乎較難達到純粹實驗的要求。一般假設團體輔導的效果要視「帶領者」、「個案的特質」和其「問題性質」而定。所謂「個案特質」是指個案的年齡、背景、教育程度和性別等；而「問題性質」是指人際交往、個人困擾、情感或認知失調等。所以其效果是否顯著，應考慮許多變項；和其他心理學研究不太相同，譬如反應時間測量、迴避學習等的研究，其顯著水準可要求高一些，但是團體輔導研究就不能這麼要求了。而且，一般實驗設計只告訴我們輔導或諮商前後，許多人平均表現(average performance)的差異，而不能幫助我們瞭解個體在許多交互影響的因素下，所產生的複雜現象。此外，當初研究者的目的在於比較何種輔導方法對學習者較有幫助，然而，閱讀相關研究文獻時，卻有些不易解釋的情況。其一是出現「虛無的結果」(null result)，指實驗處理對學生的生活適應並無明顯之影響，亦即實驗組與控制組之間沒有顯著差異；其次是出現紛歧的結果，例如，在對同一年級的研究中，同樣的現實治療取向卻在不同的實驗中產生不同的成效，也就是有些實驗發現效果的存在，有些則否。實驗出現虛無結果或者出現紛歧結果誠然是令人困惑的問題。一般而言，研究者在形成假設的過程中預期研究結果是有差異的，但此差異若未達統計顯著也會造成解釋上的困難，因爲這可能意味沒有實驗效果存在，或有實驗效果但研究本身的敏感度(sensitivity)不足以偵測到效果存在。因此實驗的處理是否產生組別間的差異，需經實徵性的評估才能加以確認。

職是之故，研究方法的選擇，往往限制了我們研究的對象，所以，利用傳統的物理或生物的實驗方法，將把社會科學帶入一個迷思—認爲「科學就是統計」。有鑑於此，是不是需要找出一些適合團體輔導的研究方法？再者，在推論統計的顯著水準概念，一般都訂在.05或.01，而這種限定是否眞能應用在班級團體輔導或一般諮商的研究上？一般精確的物理或化學的實驗，其標準可能限定在.01 以下，但面對一種感性，且情緒起伏的團體輔導過程中，其顯著水準的標準，應該就不能如同實驗室的要求一般嚴苛了。

統計的顯著水準($\alpha$)本應是一項極爲重要的觀念，卻已爲一般研究者的疏忽，使其變成一種陳腔濫調，隨意採用的一項檢定統計假設所必須的一種限制而已。事實上，統計的顯著水準($\alpha$)是指研究者所「訂定」的一項決定拒絕虛無假設(null hypothesis)與否的最低的機率。這項機率是「必須」在擬定研究的統計假設之後而在眞正檢定樣本資料結果之前就先審愼訂定的，而不是視手頭上資料的結果達到何種顯著程度而論。如果我們以目前刊登在學報的研究論文來看，我們很少發現研究者在事前訂定統計顯著水準($\alpha$)的跡象，幾乎研究者只是在事後視樣本結果所獲得之 Z、t、F、$\chi^2$值，比照查閱在任何一本統計書末所附的各該項統計檢定分配表上的理論上之數值，以確定是否

達到一般最低限度的.05 或其他程度的顯著水準而已。這種作法值得商榷。雖已成習慣，爲了方便起見，一般研究者採用.05 顯著水準幾已成必然之事。至於.05 的傳統是如何產生的？一般的理解是 Fisher 在發展變異數分析時所建立的傳統，而且帶有一些武斷的(arbitrary)成分在內(Cowles & Davis, 1982)。但事實上，研究者是有其獨立自主權的，在權衡可能影響研究結果的輕重後，可以採用適當的機率爲統計的顯著水準，而不必受制於一般傳統習慣非將顯著水準訂在.05 或.01 不可。

所以研究者以前所思考的問題「.05 顯著水準是否眞能應用在班級團體輔導或一般諮商的研究上？」並不是一項「眞能應用」與否的問題，而是使用是否「恰當」的問題。能否採用恰當機率的顯著水準，端賴研究者是否審愼地考慮到一切可能影響統計假設檢定的因素。研究者得考慮採用某一種顯著水準可能導致研究結果決定之錯誤程度。因此，當母群之間存在著差異時，研究者要能偵測到組別之間的差異才能正確的拒絕虛無假設；而要瞭解研究能偵測到多少的組別間差異程度，則牽涉到統計考驗力(statistical power)的問題。

## 二、研究目的

本研究的目的有三：

（一）嘗試對統計顯著性考驗與統計考驗力做一省思與探究。

（二）以三所研究所的教育研究論文爲例，分析其平均統計考驗力，並與國外研究結果比較是否有差異。

（三）根據研究結果對班級團體輔導教學實驗的研究提出相關建議。

## 三、名詞釋義

本研究所謂「班級團體輔導」，乃指輔導活動是以班級教學方式進行。以班級爲單位進行輔導活動，不僅可提供學生有關生活適應與生活態度之訊息，而且其推行之效果也易落實在眞正的教學情境中。而班級課程教學方式，也是一種既經濟且有效之方式，相當適合目前無專任輔導教師編製而輔導人力又嚴重不足之國小實際教學場域中。

# 文獻探討

## 一、統計考驗力及其重要性

Cohen(1962)調查 1960 年變態心理學期刊(The Journal of Abnormal Psychology)內各篇論文的統計考驗力，發現小、中、大效果值(effect size)的平均統計考驗力分別僅爲.18、.46、.83，此意味著這些刊出論文的中效果值能達到統計顯著水準的機率還不到一半，統計考驗力分析開始受到重視。而 Sedlmeier 與 Gigerenzer(1989)再次針對變態心理學(The Journal of Abnormal Psychology)調查 54 篇論文的統計考驗力，發現僅有兩篇提及統計考驗力分析，中效果值的平均統計考驗力僅爲.44。

最近，教育與心理學界再度關心統計考驗力分析的重要性，一些學者(Daniel,1993; Deng, 2000; Maddock, 2000; Tener, 2000)以期刊或博士論文的統計考驗力爲研究，得知要偵測出小效果值的研究結果，其機率大約在.13 至.41 間，而中效果值的統計考驗力介於.53 至.81 之間，實屬偏低。職是，雖然 Cohen(1990)認爲統計考驗力分析要成爲研究者主要的例行工作，或要普遍成爲教科書的重要題材，但要達此目的，尚須假以時日。

長期以來，研究者特別注意犯第一類型錯誤，而忽視了統計考驗力，如此似乎是捨本逐末，也可能導致追逐統計上的意義甚於在教育上或科學上意義的偏差，鑑於此，美國心理出版協會(APA)第六版的出版手冊中，有關「結論章節中應包括些什麼」即要求在稿件中應提供樣本大小的分析及效果值與統計考驗力的計算(Wilkinson, 1999)，這項做法除能增加研究者及參閱者對研究成果的信心外，研究的成果將更具實用價值。因此，爲使研究水準更趨嚴謹，研究報告中闡明統計考驗力與效果值應是必要

的。

統計考驗力(statistical power)是指能正確拒絕虛無假設的機率，一種判定做出正確決定的準確度之參考依據，例如研究者想探求某一現象或某個實驗效果，當這些現象或是效果眞的存在的機率即表示統計考驗力。在推論統計的條件機率下，一般行爲科學研究，研究者對研究假設進行考驗時，並非要考驗對立假設爲眞的可能性，而針對虛無假設進行考驗，以「否證」某個現象或效果可能是存在的，這種對立假設的觀念被納入 Neyman-Pearson 的假設考驗模式中，統計考驗力的概念才得以發展出來（張漢宜，2002），因此要瞭解統計考驗力時，應自有關假設考驗觀念著手。

依據前述的定義，統計考驗力(1-β)是避免犯第二類型錯誤的機率，如果統計考驗力低，代表犯第二類型錯誤的機率相對的高，因此若不進行統計考驗力的檢核，那研究結果的眞確性即很難辨明；另 α、β 是互爲消長的，而 β 與統計考驗力(1-β)也是互爲消長，因此可以說 α 與(1-β)成正比。緣此，當設定 α 值愈小的時候，則統計考驗力也愈小。防範產生第二類型錯誤的可能，在推論統計上稱之爲統計考驗力(power of test)。換言之，統計考驗力也就是「當虛無假設不正確時，正確拒絕它的機率」。當 α 愈小時，β 將愈大，統計考驗力(P)也就變得愈小。反之，當 α 較大時，β 將變小，統計考驗力(P)也就較大。從 β 和 P 的關係，我們可以知道 β 和 P 大抵同屬於跟虛無假設不同的抽樣分配裡。以數學公式示之，β 和 P 的關係是：P＝1-β。

著重追求虛無假設的顯著性考驗，忽略統計考驗力與效果值大小對研究品質的影響，或許是一般研究者過於擔心犯第一類型錯誤，或時下教科書未能給予明確闡述所造成。統計考驗力能正確地拒絕虛無假設的機率，也是決定一項研究結果可信度的重要指標，而量化研究品質的提昇也端賴統計考驗力的分析與效果值的分析（李茂能，2002），因統計考驗力較高時，若研究結果達到顯著，研究者可以抱持著比較高的信心去接受研究假設，而 Rossi(1990)指出，統計考驗力的訊息能告訴研究者統計考驗能達到的統計顯著的機率、幫助研究者解釋虛無結果及助於研究者洞察領悟整體的文獻。相對的，許多學者（李茂能，2002；張德榮，1982；謝季宏、塗金堂，1998；Bezeau & Grave, 2001）認爲忽視統計考驗力可能的危害包括：

（一）統計考驗力偏低，顯示犯第二類型錯誤的可能性極高。

（二）忽視統計考驗力，僅重視統計顯著水準，可能導致研究結果的誤判。

（三）在樣本數量與對象選擇上，容易造成資源的浪費或對受試者的傷害。

職是，量化研究品質的提升，除了取樣需具代表性與蒐集資料的工具信效度外，端賴統計考驗力與效果值分析。

## 二、教育研究假設獲得支持的迷思

統計考驗力是當研究假設是眞實時，預期的結果將可在樣本上呈現的機率的一種評定。在研究事實既定後（事後），統計考驗力的值就代表著如果重新做相同的研究時，能發現類同結果的可能性。

Cohen 指出，現行的統計檢定模式是依賴統計顯著性爲求證確實性的指數(an index of certainty)，這種例行作法至少有四點迷思之處：

（一）要求研究者報導統計顯著性，其主要只是假定在事前研究者已選擇檢定規則，以效果值大到某種程度方認爲具有科學上顯著或教育上的顯著。然而誠如 Cohen(1977)的調查顯示，大多數的研究者並未在事前作統計考驗力的分析。

（二）即使事後的結果達到統計上的顯著，其效果程度常常是並不顯著的。

（三）即使事後的結果沒有達到統計上的顯著，效果程度或方向卻常會顯著。

（四）許多達到統計上的顯著差異是「沒有其重要性的」，而許多重要的差異卻未達到統計上的顯著水準。

Rojewski(1999)與 Thompson(1998)認爲，現在研究中虛無假設的顯著性考驗，一直給人的印象是只談第一類型錯誤$\alpha$，避談第二類型錯誤$\beta$與統計考驗力$\rho$；只關心 p 值是否不大於$\alpha$，不在乎效果值是否具有應用價值。這些似是而非的迷思，導致不少的困惑與誤解：

（一）許多研究者常把虛無假設考驗的統計顯著性視爲研究價值的指標，當研究結果達到預期的統計顯著水準，即下結論說研究結果在臨床上或實際應用上，具有顯著效果。

（二）不少研究者認爲犯第一類型錯誤較嚴重必須加以控制，犯第二類型錯誤可以不管，而使用了統計考驗力很低的統計考驗，許多矛盾的研究結論因而產生。

（三）對於實驗效果的考驗，因襲統計力的假設考驗，忽視了區間估計的效用性，而導致臨床應用上二分法（有效或無效）之誤導。

（四）很多研究者於解釋 p 值時，常把 p 值看成效果值的代名詞而有不當之詮釋，導致在解釋 p 值時，產生很多的誤用與迷思。

（五）使用大樣本永遠比小樣本好，以追求統計的顯著性。

上述這些迷思與誤解易導致統計顯著考驗的誤用，以致於研究結果的誤判（李茂能，2002）。研究者渴望其虛無假設($H_0$)達到.05 的顯著水準的程度已成爲研究的陳腔濫調。事實上，傳統的統計檢定法是對虛無假設有所偏估，復以一般驗證的理論離理想的深度尚遠，而研究者對其研究設計之有欠周密，致使其研究成爲荒謬的假實驗 (Pseudo-experiments) (Cohen & Hyman, 1979)，是故達到.05 或.01 顯著水準的研究結果屢見不鮮。姑不論否定虛無假設是否涉及第一類型錯誤，遑言統計檢定的結果是否具有辨別虛無假設眞僞的能力。研究者徒從研究假設之獲得支持，「似乎以顯著水準來評定一項研究之成敗了」（張德榮，1982）。

決定一項研究（使用推論統計者）結果之可信度，端視統計考驗力之高低而定。是故，統計考驗力爲考驗假設過程中最重要的一環，重視研究的統計考驗力不但可使研究者避免徒勞無功，且可使行爲科學研究的消費者對非科技研究的成果具有信心。

Cohen(1977)認爲推論統計的基本核心要素是：(1)統計考驗力、(2)顯著水準、(3)樣本人數、(4)效果值，這四項要素是互爲影響的，任何三項決定了第四項要素的狀況。亦即，統計考驗力是由顯著水準($\alpha$)、效果值($\gamma$)、及樣本人數(n)所決定的函數。職是之故，影響統計考驗力的因素，主要有第一類型錯誤$\alpha$的大小、效果值的大小、樣本人數的多寡等（謝季宏、塗金堂，1998；Cohen, 1988; Grimm, 1993）。以下分別說明之：

### （一）α、β 對統計考驗力之影響

「研究結果是否達到顯著水準？」是研究者所關切的問題。一般研究似乎也以顯著水準來評鑑其研究之成敗。Cohen(1962)分析了 70 篇達到.05 顯著水準的研究論文；Brewer(1972)分析了 373 篇達到.05 或.01 顯著水準的研究論文，結果都發現這些研究論文的統計考驗力是很微弱的。而 Chase 與 Chase(1977)針對 1974 年應用心理學雜誌，調查 121 篇論文的統計考驗力，發現小效果值的平均統計考驗力僅爲.25，中效果值的平均統計考驗力僅爲.66，大效果值的平均統計考驗力僅爲.84，統計考驗力偏低的情況並未改善多少。國內張德榮（1982）以台灣師大教育心理系、高雄師大教育系、及彰化師大輔導系之學術研究期刊（62 年至 70 年），隨機選取 50 篇爲樣本做統計考驗力的分析，其結果與 Cohen(1962)早期的研究結果頗爲類似，亦即犯第二類型錯誤的可能性極高。

Neyman 和 Pearson（1933）將顯著水準和否定正確的或接受錯誤的虛無假設混爲一談。他們介紹了統計假設檢定的兩項極重要的觀念，即所謂第一類型錯誤(Type I error, α)和第二類型錯(Type II error, β)。他們主要以這兩種類型錯誤、臨界區域和統計考驗力，作爲檢定統計假設時所應考慮的依據。「當虛無假設是正確的，而研究結果否定了他的眞實性」，就產生

第一類型錯誤.05 或.01 的顯著水準，亦即研究者所願冒犯第一類型錯誤的機率；因此第一類型錯誤也就是 $\alpha$。第二類型錯誤則是「當虛無假設事實上是錯誤的，而研究結果使研究者接受它以爲眞」。第一、二類型錯誤是互相影響的。如果顯著水準訂得愈低，研究者犯第一類型錯誤之可能性會愈小，因爲臨界區域愈小，否定虛無假設之可能性也愈小。是故，否定虛無假設的可能性就並不全是由於機會使然。但當顯著水準訂得低時，跟臨界區域相鄰的「接受區域」(region of acceptance)就相對地變大了。其影響所及，會使得研究者接受事實上錯誤的虛無假設而信以爲可能。因此，第二類型錯誤和錯誤的虛無假設有密切關係。

防範產生第二類型錯誤的可能，在推論統計上稱之爲統計考驗力。換言之，統計考驗力也就是「當虛無假設不正確時，否定它的機率」（林清山，1993）。當 α 愈小時，β 將愈大，統計考驗力 P 也就變得愈小；相反地，當 α 愈大時，β 將愈小，統計考驗力 P 也就變得愈大。以數學公式式之，β 和 ρ 的關係是：P=1-β。從統計考驗力分析的角度來看，決定顯著水準的大小是一種兩難(dilemma)的問題，這是因爲統計考驗力與第二類型(β)錯誤率是一種互補的(complement)關係。

有些統計考驗方法在牽涉到拒絕虛無假設時會有方向性的問題（如 t 考驗），在進行雙尾考驗時，拒絕虛無假設的臨界顯著區會位於考驗統計值抽樣分配的左右兩端，用於計算統計考驗力的 $\alpha$ 也只有原來的一半，計算出來的統計考驗力相對也較小；而在進行單尾考驗時，臨界顯著區則集中在於某一端，因此，單尾考驗的統計考驗力會比雙尾考驗來的高。

### （二）效果值對統計考驗力之影響

研究者欲估計其統計檢定的第二類型錯誤的機率，首先得決定其對立假設的特定值及研究樣本人數。對立假設的特定值與虛無假設的差異稱爲母群體的「效果值」，亦即指虛無假設與對立假設之間差距情形，也就是因自變項的不同而導致依變項的差異程度。效果值通常以希臘字母 γ 代表，每種統計方法都有其各自的效果值計算公式，有興趣讀者可參考 Cohen(1988)的著作。一般在 SPSS 或 SAS 的變異數分析報表中，則以 eta-squared($\eta^2$) 或 partial$\eta^2$ 來代表樣本效果值的大小，以 $\omega^2$ 表示母群效果值，$\omega^2$ 永遠比 $\eta^2$ 來的小。另外，Schwarzer(1989) 所研發的免費整合分析(Meta-analysis)軟體，可協助研究者計算與轉換各種效果值，堪稱便利，值得讀者下載參考。

效果值愈大，統計考驗力就愈高。研究者在蒐集資料前，所預估的效果值，乃是他認爲比此值更大的效果才具有應用價值的臨界值，亦即最起碼的效果值。由於影響統計考驗力大小的主要因素爲效果值、犯第一類型錯誤機率，及樣本人數等三個，也就是統計考驗力是效果值、犯第一類型錯誤機率，與樣本人數的函數。當效果值、犯第一類型錯誤機率，與樣本人數均爲已知時，就可以計算出統計考驗力的大小。因此，計算統計考驗力時，必先確定效果值的大小。然而要求研究人員在進行研究之前就先估算其可能獲得的效果值，卻是件不易的事，爲了解決事先估算效果值的難題，Cohen(1988)根據經驗法則所提出的權宜措施，訂定了大、中、小效果值之判斷標準（摘述如表 **1**）亦可作爲社會科學研究者參考之依據。他將效果值大致分成三類：小效果值約爲.20、中效果值約爲.50、大效果值約爲.80。Cohen(1990)發現大多數的研究結果其效果值爲.05；研究結果屬於大效果值的，通常是研究設計能有效操控控制變項，例如實驗心理學或生理心理學等；至於屬於小效果值的研究，通常是對實驗變項缺乏良好的控制，例如試探性研究或實地研究。研究者在決定效果值時，除了可以根據本身研究控制的程度外，而參考 Cohen(1988)所建議的小效果值.20、中效果值.50，與大效果值.80 外，亦可根據過去的研究結果，或依理論根據對效果值的上下限間，做明智之推估或進行前導性研究後再歸納出可能的臨界效果值。

表 1　Cohen 氏大、中、小效果值之判定標準

| 統計檢定 | 指標 | 效果值（以標準分數計） | | |
|---|---|---|---|---|
| | | 小 | 中 | 大 |
| 平均數的 t 考驗 | d | .20 | .50 | .80 |
| 相關係數的 t 考驗 | r | .10 | .30 | .50 |
| ANOVA 的 F 考驗 | f | .10 | .25 | .40 |
| MCR 的 F 考驗 | $f^2$ | .02 | .15 | .35 |
| $\chi^2$ 考驗 | w | .10 | .30 | .50 |

註：採自「量化研究的品管：統計考驗力與效果值分析」，李茂能（2002）。**國民教育研究學報，8**，頁 4。

### （三）樣本人數 n 對統計考驗力之影響

在推論統計中，由於母群母數的值未知，因此研究者需以 $\delta/\sqrt{n}$ 做爲標準誤來推估母群的性質，樣本愈大時，標準誤愈小，則拒絕虛無假設的可能性愈高，因此可說，當其他條件不變時，n 遞增時，統計考驗力遞增；當 n 遞減時，統計考驗力同樣遞減。亦即統計考驗力與樣本大小具有非常密切的關係，如在 t 考驗中假如分子保持恆定，樣本平均數的抽樣標準誤會因 n 逐漸增大而使 t 值變大。由是觀之，當樣本大小趨於無限大時，這些統計量不管其實際之效果值有多小，勢必會達到統計上的任何顯著水準；當研究者使用的樣本很小時，這些統計量不管其實際之效果值有多大，勢必永遠無法達到統計上的任何顯著水準。也因此，推論考驗經常被批評爲只是一種數字遊戲罷了。

因此，在從事研究之前，透過統計考驗力而估算出所需要的適當樣本，是相當重要的事。在教育或心理學的研究研究中，許多的研究都需要藉由測量樣本來調查或推論母群的某些現象。在此過程中，受試者人數的多寡是一個重要的問題，因爲當受試者增加時，樣本能夠有效代表母群的機率也會相對地提高 (Heppner, Kivlighan, & Wampold, 2008)。然而，在設計教育研究時，樣本應該有多大，是許多統計學者經常被問到的問題 (Green, 1991)。因此，如何決定一個適當的樣本大小，使得虛無假設被拒絕時，不僅在統計上是顯著的，而且具有實用顯著(practical significance)，自然是一個值得探討的問題。爲了兼顧統計顯著與實用顯著，而且考慮到研究資源的有限性，研究者需要找出一個合理的最小樣本大小。目前在決定樣本大小的研究方面，是以統計考驗力分析作爲主要的方法（張德榮，1982；謝季宏、涂金堂，1998；Cohen, 1992; Green, 1991）。底下介紹 G＊Power 統計考驗力分析軟體，可快速呈現出所需規劃樣本之大小。

## 三、G＊Power 統計考驗力分析軟體簡介

統計考驗力是顯著水準、樣本大小與效果值的函數，因此只要知道這三個因素大小，研究者就能進行統計考驗力的分析。統計考驗力分析的軟體除了「Statistical Power Analysis」之外，目前也有許多的統計考驗力軟體可以使用，如 G＊Power、PASS、PC-SIZE、PowerPlant、PS、SamplePower、SIMSTAT、STPLAN 等軟體，尤其 G＊Power(Erdfelder, Faul, & Buchner, 1996)係免費共用軟體，最爲大家所樂用。（其下載網址爲：http://www.psycho.uni-Duesseldorf.de/aap/prjects/gpower/index.html）

G＊Power 可同時分析三種不同類別的統計考驗力分析。使用時要先在視窗的最右側 Analysis 欄與 Test is 欄中確認研究者要進行哪種考驗力分析，再先點選功能表單上的「Tests」，選擇待考驗的統計量（含 t 考驗、F 考驗與 $\chi^2$ 考驗等七種），接著填入效果值大小、α 大小與所需之 Power，再按視窗「Caculate」，電腦即會將你所規劃的樣本大小，顯示在視窗中央，而整個計算參數、過程與結果都會顯示於視窗的下方 Protocol 中。

統計考驗力、效果值的分析與樣本大小、α 和 β 具有密切的函數關係。讀者亦可應用 G＊Power 探求它們之間的關係。例如，過去最常見研究者利用統計考驗力、效果值與 α、β 的大小去估計所需要的樣本大小（選擇 A Prior Analysis 或 Compromise Analysis）。其次，研究者亦可利用樣本大小、效果值與 α、β 的大小，去估計統計考驗力（選擇 post Hoc Analysis）。同樣地，研究者亦可利用統計考驗力、樣本大小與 α、β 的大小，計算效果值大小（按「Calc Effect size」），以進行研究結果的資料分析。

## 研究方法

### 一、研究對象

本研究分析對象爲國立台北教育大學（北）、嘉義大學（中）、及屏東教育大學（南）等三校，自 92 至 97 學年度有關教育研究方面的碩士論文，採用 t 考驗研究假設的論文爲主，經審閱後，共選取 30 篇碩士論文，檢定 594 項假設。另 F 考驗檢定、卡方考驗檢定及有關二級考驗，如內部一致性、因素分析、事後比較與單純主要效果並未在本研究分析範圍中。

### 二、分析步驟

其分析步驟如下：1.分析每篇研究論文研究假設之顯著水準、樣本大小及單側考驗或雙側考驗等訊息；2.以 G＊Power 軟體作爲統計工具，在該軟體視窗中，依序輸入選定的小、中、大三項效果值及步驟 1.之各項數值，以計算各項研究假設之統計考驗力值；3.依照選定的小、中、大三項效果值，計算統計考驗力的平均值。

### 三、分析與資料處理的工具

本研究利用以 G＊Power 軟體計算各篇研究論文中有關 t 考驗各項研究假設的統計考驗力。之後，再以 Excel 軟體對 t 考驗計算其小、中、大三項效果值的平均考驗及百分次數分配情形，包括百分比、檢定樣本大小、中數、眾數、變異數等描述性統計資料。

在 30 篇論文中，總共檢定 594 項研究假設，在小、中、大三項效果值的統計考驗力平均值依次爲.175、.928 及.973，其相關描述統計資料如表 2。

## 結果與討論

### 一、小效果值結果的分析

在小效果值上，t 考驗的研究假設統計考驗力屬於偏低的，僅有.175。此意謂著有近乎 82％的機率犯第二類型錯誤（近八成的機率接受不正確的虛無假設）。在整體的 t 考驗中，有 94％的研究假設未達.80 的傳統標準(Cohen, 1988)，而根據國外研究統計，各相關行爲科學領域研究論文在小效果值的統計考驗力平均數大約在.13 至.41 間；另張漢宜（2002）在其博士論文中偵測小效果值的研究結果爲.24。與之相較，本研究對象在 t 考驗的小效果值統計考驗力上與國外有關報告相近，但就理論觀點而言，此研究對象有關 t 考驗研究假設的小效果值統計考驗力仍屬偏低。職是，當一研究報告的統計考驗力數值偏低或是未交代其統計考驗力的大小時，那該文獻內容的統計顯著性的可靠性是應該被懷疑的，因爲，低的統計考驗力隱含著犯第二類型的錯誤機率會增高；同樣地，統計考驗力數值低落也會增加犯第一類型錯誤的機率。

### 二、中效果值結果的分析

在中效果值上，t 考驗的研究假設統計考驗力平均值達.928，已達.80 的傳統標準，也明顯高於國外的.64 及國內的.63 平均值（張漢宜，2002；張德榮，1982）；同時，也有 93％的 t 考驗研究假設統計考驗力達.80 以上。在此情況下我們自可有相當的信心接受對立假設，

表 2　t 考驗研究假設的統計考驗力摘要表

| | 小效果 | | 中效果 | | 大效果 | |
|---|---|---|---|---|---|---|
| 統計考驗力 | 次數 | 累積百分比 | 次數 | 累積百分比 | 次數 | 累積百分比 |
| .91 至 1.0 | 7 | 100 | 472 | 100 | 570 | 100 |
| .81 至.90 | 34 | 99 | 80 | 21 | 4 | 5 |
| .71 至.80 | 77 | 94 | 6 | 7 | 9 | 4 |
| .61 至.70 | 81 | 81 | 9 | 6 | 1 | 2 |
| .51 至.60 | 112 | 67 | 8 | 4 | 3 | 2 |
| .41 至.50 | 61 | 48 | 6 | 3 | 2 | 1 |
| .31 至.40 | 84 | 38 | 6 | 2 | 1 | 1 |
| .21 至.30 | 93 | 24 | 2 | 1 | 0 | 1 |
| .11 至.20 | 39 | 8 | 5 | 1 | 3 | 1 |
| .01 至.10 | 6 | 1 | 0 | 0 | 1 | 0 |
| 檢定樣本數 | 594 | | 594 | | 594 | |
| 篇數 | 30 | | 30 | | 30 | |
| 平均數 | .175 | | .928 | | .973 | |
| 中數 | .501 | | .981 | | 1 | |
| 眾數 | .201 | | 1 | | 1 | |
| 標準差 | .193 | | .120 | | .093 | |
| Q1 | .283 | | .911 | | .972 | |
| Q3 | .674 | | 1 | | 1 | |
| 平均數 95% 信賴區間 | .195 至.164 | | .923 至.902 | | .971 至.964 | |

但對中效果值欄的 Q1 與 Q3 稍加計算，即可發現，在這些檢證的樣本中，有相當數量的離群樣本與極端樣本點，致統計考驗力有膨脹現象。

## 三、大效果值結果的分析

在大效果值上，t 考驗的研究假設統計考驗力平均值達.973，已達.80 的傳統標準，也明顯高於國外的.86 及國內的.85 平均值（張漢宜，2002；張德榮，1982）；同時，也有 96% 的 t 考驗研究假設統計考驗力達.80 以上。在此情況下我們自可有相當的信心接受對立假設，但對中效果值欄的 Q1 與 Q3 稍加計算，即可發現，在這些檢證的樣本中多數爲極端值，致統計考驗力有膨脹現象。

上述中、大效果值的統計考驗力值是相當的高，然一旦高於.95 時，在解釋上要特別小心。因前述及統計考驗力是樣本數、效果值與顯著水準所決定的函數，而此四個要素中，又以樣本人數最易受研究者所控制，而本研究分析是採理論的效果值而非實得效果值，且顯著的效果值也是一固定值，顯而易見的是本研究中統計考驗力膨脹的因素可推斷是受樣本人數因素所影響，因此以下茲就對本研究有關 t 檢定研究假設樣本人數分配情形做討論。

中樣本人數在 300 人以上的比率爲 70%，此可說明何以在中、大效果值中，統計考驗力呈現高於.95 的原因，再與 Cohen(1988)推估的樣本大小估計數相較，顯然各研究假設樣本人數是過多的。從另一角度來看，非常大的樣本幾乎常導致統計上的顯著性(Cohen, 1977, 1988)，因此，對達「顯著性」的研究假設，但爲報告效果值統計考驗力其眞確性可能有待進一步的考驗。

表 3　t 檢定樣本人數分配表

| 樣本人數 | 次數 | % |
|---|---|---|
| 300 以上 | 413 | 70 |
| 271 至 300 | 20 | 3 |
| 241 至 270 | 2 | 0 |
| 211 至 240 | 42 | 7 |
| 181 至 210 | 19 | 3 |
| 151 至 180 | 48 | 8 |
| 121 至 150 | 34 | 6 |
| 91 至 120 | 6 | 2 |
| 61 至 90 | 5 | 1 |
| 31 至 60 | 3 | 0 |
| 1 至 30 | 2 | 0 |
| 檢定樣本數 | 594 | |
| 樣本中數 | 410 | |

綜上所述，本研究 t 檢定上，中、大效果值的統計考驗力皆高達.9 以上，但也發現在 t 檢定中有 70%樣本屬離群值或極端值，致造成統計考驗力有膨脹現象，另分析 t 檢定樣本人數分配情形，發現在 t 檢定中使用樣本人數的平均值（中數）爲 410 人，遠高於 Cohen(1992)推薦在中、大效果值的樣本人數 21 至 177 人。

## 四、討論

由於筆者碩士論文爲現實治療取向的班級輔導活動，是一種教學實驗研究，而本研究中所採 30 篇有關教育研究之碩士論文，則屬於非實驗教育研究。一般而言，教學實驗研究受限於樣本取得困難，受試者的流失(experimental mortality)等問題，樣本大小通常都較其他類型的研究來得少，就檢視兩者的樣本訊息後可發現，當初筆者教學實驗的樣本大小平均數爲 55 人，而本研究中非實驗教育研究的樣本大小平均數爲 410 人。因此，如果教學實驗的樣本大小受限於現實教育場境之取樣困難與樣本流失等問題，不易取得足夠之參與者，則在其他條件相等的情況下，教學實驗的統計考驗力有可能會低於非實驗教育研究的統計考驗力(Sawyer & Ball, 1983)。

Cohen(1988)提到，交互作用的考驗力會比主要效果的考驗力來得低。造成統計考驗力降低的理由在於，交互作用的考驗力是使用格的人數計算而來的，因此交互作用的層級越高，則格人數就越少。而在筆者的班級團體輔導教學實驗中最主要的統計方法是變異數分析與共變數分析，且使用 2×2 多因子設計，因子數的增加會導致交互作用的層次提高，而隨著交互作用層次的提高，交互作用的統計考驗力往往會降低筆者該篇研究的統計考驗力平均值。

此外，筆者之實驗教學由於使用實驗控制及排除前測共變數之統計控制來降低誤差，用以避免實驗效果被混淆，因而，其效果有可能會大於非實驗教育研究之效果。因此，即使本研究中非實驗教育研究的樣本大小大於筆者的教學實驗的樣本，但是兩者的統計考驗力未必會有差異。不過由於目前並沒有研究者針對各類型研究的效果值進行深入之調查，所以教學實驗與非實驗教育研究的「估計效果值」各是如何，仍不清楚。這是未來可以繼續研究的一個方向。

# 結論與建議

## 一、結論

### （一）小中大效果值之統計考驗力表現不一

本研究分析對象爲國立台北教育大學（北）、嘉義大學（中）、及屏東教育大學（南）等三校，自 92 至 97 學年度有關教育研究方面的碩士論文，採用 t 考驗研究假設的論文爲主，經審閱後，共選取 30 篇碩士論文，檢定 594 項假設。分析結果發現，在 t 檢定中：

1.小效果值的統計考驗力略爲低落

在本研究中，小效果值上的研究假設的統計考驗力只有.175，從其平均數 95%信心區間來看，與國內相關領域研究的統計考驗力平均值.18（張德榮，1982）相仿，但低於國外平均值.24（張漢宜，2002)。國外的許多研究者普遍覺得，在他們研究中小效果的考驗力有些低落，而本研究偵測小效果的統計考驗力還低於國外統計考驗力研究的平均數。此意謂國內教學實驗偵測小效果的統計考驗力偏低。而此偏低的考驗力將造成教學實驗者較難偵測到實驗組與控制組之間的實驗差異。

2.中大效果的統計考驗力尚稱充足

另在 t 檢定中，中、大效果值的統計考驗力分別爲.928、.973，基本上，這個研究結果還算可以接受。但也發現在 t 檢定中有 70%樣本屬離群值或極端值，造成統計考驗力有膨脹現象，另分析各項檢定樣本人數分配情形，發現在 t 檢定中使用樣本人數的平均值（中數）爲 410 人，遠高於 Cohen(1992)推薦在中、大效果值樣本人數 21 至 177 人。因此，對於這個情況我們需要審慎地看待，整體教育研究的統計考驗力仍有向上提升的空間。

### （二）要先排除或削弱足以影響研究之障礙因素，再考慮顯著水準之訂定

現在再回到研究者先前所提出顯著水準限度在班級輔導或一般諮商研究上是否眞能應用上的問題。Winer(1971)認爲「假如研究是極其嚴謹，則無論採用何種顯著水準(α)，其研究假設的統計考驗力也不失其當；反之，如研究是在有限度的情況下（意指並非十分嚴謹之研究）執行，那種唯有盡可能使研究設計趨於完善。」從 Winer 的話，我們可以知道顯著水準(α)之訂定雖然影響一項研究甚鉅，但其影響力還遠不如嚴謹的研究設計來的重要。研究者覺得在班級輔導或一般諮商的研究上，影響研究結果的外在因素(external factors)甚多，研究者應事先謹愼考慮如何控制那些可能影響研究的外在因素，要能排除或削弱這些障礙因素，再考慮顯著水準之訂定，方爲妥善之法。若能控制或排除那些可能影響研究的外在因素，研究者方能信心十足地宣稱某一特殊團體或個別諮商技術對某個個案有顯著成效。

## 二、建議

由此可知，一個研究達到統計的顯著水準可能係因樣本過大所致，統計的顯著性不代表臨床上或教育上一定具有應用價值；而統計尚未達顯著水準，可能係統計考驗力太低，或效果太小，但不代表沒有臨床上或教育上的應用價值。至於.05 或.01 顯著水準的限度問題，只是指在某一限制內所達到的統計上的意義而已，很多時候達到統計水準的研究結果，在實質上並不具顯著的意義（教育上的意義）。長期以來，許多研究者特別注意犯第一類型錯誤，而忽略了統計考驗力，可說是捨本逐末，也將導致追逐統計上的意義甚於在教育上或科學上意義的偏差。因此，如果在這層上有所體認，就不至於過份強調統計顯著水準了。

量化研究的品管從研究考驗前的樣本抽樣與工具設計即應開始，在量化研究的過程中，講求的是客觀性與推論性，研究者如能抽取適當的代表性樣本，其推論性與複製性必更高；如能講究測量工具的信、效度以減少測量誤差，其內在效度必高（李茂能，2002；Rojewski, 1999）。最後，研究者對於研究結果之解釋時亦應縝密且與過去之研究脈絡相結合，如仍無法定論，需再進行複製研究。如此，量化研究的內、外在效度必能獲得品質保證，所累積的知識才能日臻完善。

綜合先前所做的各種探討與分析，本研究

提出相關建議，分項陳述如下。

### （一）設計具有高統計考驗力的班級團體輔導教學實驗

從本研究可發現，筆者的班級團體輔導實驗教學或一般教育研究的非實驗教學的統計考驗力並不高，因此，即使班級團體輔導的處理具有效果，研究者也未必能偵測到這個效果的存在。統計考驗力低落的研究可能被烙印爲失敗的研究，但其眞正的原因可能是研究者無法敏感出實驗效果之存在，而不是研究本身沒有效果。研究者要設計出具有統計力的班級團體輔導教學實驗，在研究設計階段要考慮幾個因素：

1.彈性調整顯著水準大小

在研究實務上，降低顯著水準的標準（即大於.05）容易招來質疑眼光。當然，此種想法會面臨第一類型錯誤與第二類型錯誤孰輕孰重之權衡問題。然研究者不妨就某些狀況加以考慮，例如 Cohen(1988)認爲在交互作用情況時，由於分子自由度較大且格人數較小，統計考驗力會變得很低弱，因此研究者可考慮降低顯著水準之標準（如.10 而非.05）。如果虛無假設因而被拒絕，研究者固然要付出可信度的代價，但因爲可提高統計考驗力，所以還是值得的。此外，研究者若有理論上或先前研究之依據，使用單側考驗也相當於增加顯著水準的大小。

2.提升效果值

Rossi(1990)認爲，藉由提升實驗的效果來提高統計考驗力是比較實際可行的選擇，主要在於它們能夠降低誤差變異量。如使用共變數分析(ANCOVA)可以控制無關的混淆(extraneous)變異來源；使用多變項統計方法可有效降低各種誤差變異的來源。實驗效果會受到隨機變異的干擾，而經由改進實驗設計來控制不同的變異來源就能擴大效果值。

3.充足的樣本大小

在多因子變異數或共變數分析的情況下，不足的樣本大小對交互作用的統計考驗力會有嚴重之影響，因爲此時計算統計考驗力所使用的樣本大小是根據格人數而來。因此，班級團體輔導教學實驗之設計者應設法獲取充足的樣本大小。研究者可在實驗設計的階段，使用統計考驗力分析來決定所需之樣本大小。

### （二）運用統計考驗力概念協助分析研究資料

當班級團體輔導教學實驗的研究者無法拒絕虛無假設時，不宜明確地宣稱實驗沒有產生效果。因爲如果統計考驗力較爲低落，虛無結果的產生可能是研究變項眞的沒有效果亦有可能是實驗有效果但統計考驗沒有足夠的敏感度偵測出效果的存在。所以除非統計考驗力很高（大於.80），研究者並不宜根據虛無結果來作結論與建議。

另從分析 30 篇的教育論文的結果而言，本研究假設考驗的步驟強調研究者在研究前的規劃準備與研究後的結果報告。研究前的規劃足以顯示一項研究的嚴謹性及研究者的自律態度；研究後的結果交代足以使研究的讀者明瞭問題眞相的可信性及意義所在。以下就研究考驗之時序加以說明：

1.在研究考驗前

根據待答問題擬出對立假設與虛無假設。假設可自理論及其他的假設中，演繹而出，並彙集個人經驗與睿智而對問題的解決之看法。

2.在研究考驗過程中

(1)應在考驗假設前，決定適當之顯著水準 α。顯著水準之選擇並不侷限於.05 及.01。研究者應權衡觸犯第一類型錯誤及第二類型錯誤對研究結果之影響，以此擬定此項機率之準則。在研究報告上，研究者應說明採用特定顯著水準之原因。

(2)審愼擬定對立假設的「預期效果值」。Cohen(1977)推介的三項（小、中、大）的效果值是一般性的指標而已。研究者在無所遵循的情況下採用，是其功用所在。研究者應視其研究的實際情形來擬定效果值的假設。

(3)審愼計畫樣本的選取及人數之多寡。研究者應避免以偶然取樣爲獲得研究對象之法，因代表性的偏估足使研究結果之推論失去外在效

度。班級隨機自以「班」爲樣本之單位，在單位人數上可以斟酌擬定；而隨機學生選擇則以較多爲佳。

(4)在兼顧第一、二類型錯誤機率下，研究者應先決定「統計考驗力」的程度。研究前分析統計考驗力應包括對實際所需樣本人數之分析在內。這是結合統計考驗力、效果值，及顯著水準三項要素對所需樣本人數之估計。其目的在於審視是否實際所需樣本情形與計畫的樣本人數大有差距，是否在執行研究上可行，以爲增減之決定。在研究報告上，研究者應對研究前預測統計考驗立即樣本人數之估計做明確之說明，以獲得適當的統計考驗力。Cohen(1990)有鑑於統計考驗力的分析常大費周章，異常繁複，爲讓讀者能快速查出 Power 等於.80 時所需之樣本數，提供樣本規劃簡表供查用。研究者只要先確定所需用到的統計量、α 水準與效果值大小，即可查到雙尾檢定時所要的樣本總人數或各組的樣本人數。

(5)使用良好工具蒐集資料，以減少測量誤差，增加其內在效度。

(6)選擇適當的統計考驗。

(7)比較統計考驗值與臨界值，裁決統計上是否拒絕虛無假設。

3.在研究考驗後：

(1)根據效果值大小、效果值離散情形、成本等相關因素，判定決策上之應用價值。

(2)研究者應做事後統計考驗力分析。事後統計考驗力分析需使用實際之樣本大小、α 與理論上的效果值而非實得效果值去估計事後統計考驗力，才有助於研究結論的正確解釋(Thomas, 1997)。Cohen(1988)主張統計考驗力至少要有.80 以上，但也不要過高，以致過當而使研究結果沒有實質效益。一個研究因統計考驗力高才會達到統計上之顯著水準，因統計考驗力低才不能達到統計上之顯著水準。因此，在進行事後統計考驗力分析時，必須以理論上的效果值而非實得效果值去估計事後統計考驗力，才更具實質意義。

(3)考驗假設的結果不宜只報告達到統計顯著水準與否。建議研究結果之報告應以「效果值」爲中心。研究者就樣本效果值之報告，可以證明考驗的結果。顯著水準僅是幕後的一環。更確實的作法是研究者可以樣本效果值的「信賴區間」(confidence interval)來解釋母群相對現象的可能性。

以上所提進行統計顯著性考驗的步驟是強調研究前以統計考驗力之分析爲重心，研究後則以效果值之分析爲主。而我國量化研究正值高度產量期，要導正量化研究的方向與品質，應要求研究者於研究計畫或論文的方法論一節中，規劃樣本大小、統計考驗力分析，於結果分析一節中，除報告 p 值外，尚需報告實得統計考驗力與效果值分析（李茂能，1998，2002；Cohen, 1990; Thompson, 1998），以進行量化研究品質的最後把關。

## 參考文獻


李茂能（1998）。統計顯著性考驗的再省思。**教育研究資訊，6** (3)，103-115。

李茂能（2002）。量化研究的品管：統計考驗力與效果值分析。**國民教育研究學報，8**，1-24。

林清山（1993）。**心理與教育統計學**。臺北：東華書局。

張漢宜（2002）。**教學實驗中的考驗力分析**。國立高雄師範大學教育研究所博士論文，未出版，高雄。

張德榮（1982）。教育心理研究的統計考驗力分析。**輔導學報，5**，119-137。

謝季宏、塗金堂（1998）。T 考驗的統計考驗力之研究。**教育學刊，14**，93-113。

Bezeau, S., & Grave, R. (2001). Statistical power and effect size of clinical neuropsychology research. *Journal of Clinical & Experimental Neuropsychology, 23*(3), 399-406.

Brewer, J. K. (1972). On the power of statistical

tests in the American Educational Research Journal. *Journal of American Educational Research, 9*(3), 391-401.

Chase, L. J., & Chase, R. B. (1977). A statistical power analysis of applied psychological research. *Journal of Applied Psychology, 61*(2), 234-237.

Cohen, J. (1962). Statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*(3), 145-153.

Cohen, J. (1977). *Statistic power analysis for the behavioral sciences*(rev. ed.). New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*(2nd ed.). Hillside, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned so far. *American Psychologist, 45*, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Buelletin*, *112*(1), 155-159.

Cohen, S. A., & Hyman, J. S. (1979). How come so many hypotheses in educational research are supported? *Educational Researcher*, *8*(11), 12-16.

Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist, 37*, 553-558.

Daniel, T. D. (1993). *A statistical power analysis of the qualitative techniques used in the journal of research in music education, 1987 through 1991*. Unpublished doctoral dissertation, Auburn University, Alabama.

Deng, H. (2000). *Statistical power analysis of dissertations completed by students majoring in educational leadership at tennessee universities*. Unpublished doctoral dissertation, the East Tennessee State University.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavioral Research Methods, Instruments, & Computers, 28*, 1-11.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research, 26*, 499-510.

Grimm, L. G. (1993). *Statistical applications for the behavioral science*. New York: John Wiley & Sons.

Heppner, P. P., Kivlighan, D. M., & Wampold, B. E. (2008). *Research design in counseling* (3rd ed.). Thomson Brooks/Cole.

Maddock, J. E. (2000). Statistical power and effect size in the field of health psychology. *Dissertation Abstracts International, 60*(9), 44939B. (UMI No. 2000-95006-499)

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A, 231*, 289-337.

Rojewski, J. W. (1999). Five things greater than statistics in quantative educational research. *Journal of Vocational Research, 24*(1), 63-74.

Rossi, J. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of consulting and clinical psychology*, *58*, 646-656.

Sawyer, A. G., & Ball, A. D. (1983). Statistical power and effect size in marketing research. *Journal of Marketing Research, 18*, 275-290.

Schwarzer, R. (1989). *Statistics Software for Meta-Analysis[Computer software and manual]*. Retrieved from http://www.yorku.ca/faculty/academic/schwarzer/meta_e.htm

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309-316.

Tener, M. A. (2000). *A post hoc statistical power analysis and survey of the research published in the journal of athletic training*. Unpublished doctoral dissertation, Middle Tennessee State University.

Thomas, L. (1997). Retrospective power analysis. *Conservation Biology*, *11*, 276-280.

Thompson, B. (1998, April). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas*. Paper presented at the annual meeting of American Educational Research Association, San Diego, CA.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, *54*(8), 594-604.

Winer, B. J. (1971). *Statistical principles in experimental design*(2nd ed.). New York: McGraw-hall Book Co.

# The Effects of the Appliance of Statistical Significance on Classroom Guidance

Yao-Ming Wu

## Abstract

The idea about significance level in Inferential Statistic is generally at 0.05 or 0.01. However, whether this definition would be applicable to the research of class and group guidance or general counseling? The researcher should consider that the application of one kind of significance level might result in a mistake, which would cause the failure of the study. Therefore, the researcher should not be too considerate of the level of statistical check.

The power of test is an important index to test the hypothesis, and that can help to decide an adequate size. This paper addresses several misuses and misconceptions related to Type-I error, Type-II error, power, p-values, and sample size . Then, makes a study on the power issue. Next, computing power for any specific study via Cohen's power table and a computer program called G ✳Power. Finally, A modest hypothesis testing model is also proposed.

**Keywords: power of test, effect size, quantitative studies**

Yao-Ming Wu Pingtung Lin Luo Elementary School (a7261143@yahoo.com.tw)