

美國身心障礙學生考試調整策略成效之後設分析

張萬烽

高雄市明華國中資源班教師

鈕文英

高雄師範大學特教系教授

本文旨在後設分析美國 1999 到 2008 年間，38 篇考試調整策略對一般和全體身心障礙學生的成效；接著分析考試調整的效果是否會因學科、策略類型，以及身心障礙學生的障礙特性、教育階段和人數而有差異。結果顯示考試調整對一般學生為些微效果；對身心障礙學生為小效果量，符合「差別增長」和「互動」假設。未接受考試調整之一般學生的測驗成績，高於接受調整之身心障礙學生，二者間呈現「中」效果量的差異，顯示考試調整能維持「公平與權益的平衡」。考試調整對其他學科的效果最佳，其次分別為閱讀和數學。於調整策略上，以延長時間的效果較佳，報讀較差。在考試調整實施的教育階段上，以國中最佳，國小次之，而高中職效果最差。考試調整對 100 人以下身心障礙學生人數的效果最佳，人數愈多，效果愈差。最後，考試調整對學障學生的效果，優於其他身心障礙學生。

關鍵詞：考試調整、身心障礙學生、後設分析

緒論

一、研究背景與動機

Schnirman (2005) 指出在 1997 年之前，身心障礙學生被排除在全州或全學區的評量之外，然而隨著「標準本位改革」(standard-based reform) 的興起，情況隨之改變。在過去 10 年之間，美國特殊教育重視「績效責任」(accountability)，關注身心障礙學生在攸關晉級，及未來教育和生涯計畫決策之「高利益考試」(high-stakes testing) 上的表現，因此如何讓身心障礙學生不因其障礙阻撓其參與考試，提升考試的公平和有效性，讓學生獲得「有效」的成績，成爲一項重要挑戰 (Phillips, 1994; Tindal & Fuchs, 2000; Thompson, Blount, & Thurlow, 2002)。Thompson 等人認爲「考試調整」(test accommodation) 是使身心障礙學生參與大型考試的有效方式之一。美國 1997 年《身心障礙者教育法案》(*Individual Disabilities with Education Act, IDEA*)、2001 年《不讓孩子落後法案》(*No Child Left Behind, NCLB*)，以及 2004 年《身心障礙者教育增進法案》(*Individual with Disabilities Education Improvement Act, IDEIA*) 均強調提供考試調整，協助身心障礙學生參與「全州或全學區的評量」(Elbaum, 2007)。隨著考試調整的提供，參加考試的身心障礙學生也漸增 (Bielinski, Ysseldyke, Bolt, Friedebach, & Friedebach, 2001)；Thurlow、Moen 和 Altman (2006) 調查發現在國小到高中階段間，有 61% 到 66% 的身心障礙學生接受考試調整服務。

對教育人員和研究者來說，考試調整是否能符合身心障礙學生的需求，讓他們獲得「有效」(是指與一般學生在未調整下的得分具有相同意義)，而非「理想」的成績一直備受關切 (Thurlow & Bolt, 2001; Thurlow, House, Boys,

Scott, & Ysseldyke, 2000; Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliott, 2000; Tindal, 2002; Tindal & Fuchs, 2000)。相關研究指出學校不願意提供考試調整，尤其是當學校認爲調整花費時間、經費和人力，又無法確切掌握成效時，因此更需要提供調整對身心障礙學生是否有助益的證據以改變其迷思 (Bolt & Thurlow, 2004; Elliott, Thurlow, & Ysseldyke, 1996)。另外，教育人員在決定考試調整策略時，亦需實證資料爲依據；正如 Tindal、Helwig 和 Hollenbeck (1999, p.12) 所言：「考試調整應是基於個人需求而非盲目提供。」相關研究指出缺乏實證本位的資料，學校可能基於學生障礙特質或是擔心家長訴訟，而以個人的主觀意見決定調整策略。由此可知非常需要考試調整成效的研究資料 (Fuchs, Fuchs, & Capizzi, 2006; Hollenbeck, 2005; Lindstrom, 2007; Lindstrom & Gregg, 2007; Schnirman, 2005)。

關於考試調整的成效，相關研究發現調整策略對身心障礙學生多數有正面的效果，但亦有部分研究顯示效果有限；而哪些考試調整策略較具成效，研究結果亦不一致 (Bolt & Thurlow, 2004; Bolt & Ysseldyke, 2008; Fuchs & Fuchs, 1999; Johnstone, Altman, Thurlow, & Thompson, 2006; Schnirman, 2005; Sireci, Scarpati, & Li, 2005; Thompson et al., 2002; Zenisky & Sireci, 2007)。這些考試調整策略中，以延長時間 (extended time) 和報讀 (read aloud) 居多；Ofiesh、Hughes 和 Scott (2004) 更指出延長時間是最常使用的一項調整策略。由此可知，延長時間和報讀爲探討最多的考試調整策略，卻無一致的結論，值得探究。造成考試調整效果變異的原因包括研究方法和對象不同，以往研究大多採團體比較研究，分析身心障礙學生考試調整前後之表現，學生異質性高，此種方法不符合考試調整之「個別化」原則；除此，部分研究分析不同的考試調整策略

亦是原因；即使調整策略相同，執行過程有變異亦會影響成效（Ketterlin-Geller, Yovanoff, & Tindal, 2007; Sireci, Li, & Scarpatti, 2003）。

其中考試調整效果不一致的研究多以學習障礙（learning disabilities，簡稱學障）學生為對象；例如 Fuchs 和 Fuchs（1999）的研究指出延長時間無法補償學障學生的困難，使其在傳統考試中獲益。Fuchs、Fuchs、Hamlett、Eaton、Binkley 和 Crouch 等人（2000）的研究以 181 和 184 位學障和一般學生為對象，比較延長時間、報讀、放大字體和標準化施測方式（即未調整）的效果，結果發現學障學生從報讀中獲益的分數，明顯高於採用延長時間和放大字體者。然而，Helwig、Rozek-Tedesco 和 Tindal（2002）的研究顯示報讀對於數學計算技能佳，但閱讀能力差的國小中高年級學障學生有顯著的效果；而對於數學計算技能差的學生，不論其閱讀能力高或低，皆無法從報讀中獲益。延長時間對學障學生在數學解題上有助益，特別是學障者本身閱讀理解能力愈佳者獲益愈大（Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000）。Elbaum（2007）分析過去的研究發現報讀對國小學障學生數學的成效優於中學者，這可能是由於中學的課程難度提高，即使報讀後，學障學生還是無法理解題意，以致於無法達到效果。Fuchs 和 Fuchs（2001）表示學障學生在聽、說、讀、寫、算的限制，易與評量（尤其是紙筆測驗）互相干擾，致使教師不容易反映其考試調整的需求；而考試調整確實對部分學障學生有幫助，但由於他們的異質性大，加上不同學科適用的調整策略有異，因此沒有任何單一策略對所有學障學生都有助益。另外，Ketterlin-Geller、Alonzo、Braun-Monegan 和 Tindal（2007）指出學生人數會影響考試調整的實施。由此可知，考試調整實施在什麼學科、運用何種考試調整策略，以及身心障礙學生的人數、障礙特性和教育階段會影

響其成效。Lewandowski、Lovett 和 Rogers（2008）認為要了解考試調整成效的細微差異，有必要尋找「調節變項」（moderator variable）。總之，在分析考試調整成效時，可將學科、考試調整策略類型，以及身心障礙學生的教育階段、障礙特性和人數視為調節變項，注意其影響。

對於目前考試調整效果變異的問題，後設分析（meta-analysis）正好可解決此問題。Cook 等人（1994）指出當各研究的發現不一致時，正是運用後設分析整合研究的時機。鈕文英（2009）表示透過後設分析，可檢視實務是否為證據本位（evidence-based）。美國至 2008 年為止，僅 Chiu 和 Pearson（1999）針對 1986 至 1998 年間身心障礙和英語不精熟（limited English proficient, LEP）學生，以及 Elbaum（2007）針對 1998 至 2004 年間學障學生數學考試調整之後設分析研究。由此可知考試調整後設分析的研究仍非常有限，且未分析 2005 年以後的研究。

我國不管國中基本學力測驗或大學入學學科能力測驗等考試，均依據《特殊教育法》（教育部，2009）提供考場服務，保障身心障礙學生公平參與升學考試的機會。然而，搜尋國科會研究報告、碩博士、期刊和會議論文，我國至 2008 年底前有六篇研究探討考試調整實施現況與問題（李佩蓉，2007；洪靜怡，2008；莊虹姿，2008；黃鳳慈，2006；陳明聰、張靖卿，2004；鍾素香，2004）；僅有四篇研究探討考試調整之成效（林筱汶、陳明聰，2006；黃巧雲、陳明聰、陳政見，2007；陳蓓蓉，2003；劉婷妤、陳明聰、陳政見，2008），四篇均探討數學和自然科的考試調整，沒有研究針對語文科；探討的策略僅陳蓓蓉以國中基測試題的呈現形式，以及延長時間，分析對三位視障學生的成效；其他三篇則討論不同電腦化試題、語音調整策略，以及比

較語音合成與真人錄音對一般學生之成效。由於國內身心障礙學生考試調整策略成效的研究量尚不足，且實施上面臨許多問題，非常需要國外後設分析的研究，引導國內身心障礙學生考試調整實務之發展。

綜上所述，考試調整成效的後設分析實有必要，而 1999 到 2008 年間，分析考試調整對學障學生成效的研究僅有 15 篇，DeCoster (2004) 表示宜有至少 30 篇以上的研究才適合做後設分析，由於篇數較不足，加上僅 Chiu 和 Pearson (1999) 的研究以 1986 至 1998 年間之身心障礙學生為對象，尚缺乏 1999 至 2008 年間以身心障礙學生為對象的後設分析；因此，研究者擴大分析的對象範圍，以身心障礙學生為對象，後設分析美國考試調整對身心障礙學生的成效，在針對身心障礙學生分析時，還會分析考試調整對一般學生的成效，並且比較身心障礙與未接受調整之一般學生的成效差異，以了解考試調整是否能讓身心障礙學生獲得有效成績，又不會損及對一般學生的公平性。接著分析考試調整的效果是否會因學科、策略類型，以及身心障礙學生的障礙特性、教育階段和人數而有差異。

二、研究目的與問題

根據上述背景和動機，本研究有以下兩個目的：

1. 探究考試調整對身心障礙和一般學生的成效。

2. 分析考試調整對身心障礙學生之成效因調節變項造成的差異情形。

配合上述兩個研究目的，探討以下研究問題：

1-1. 考試調整是否能提升身心障礙學生的測驗成績？

1-2. 考試調整是否能提升一般學生的測驗成績？

1-3. 身心障礙學生接受考試調整後的成績，是否低於未接受調整的一般學生？

2-1. 考試調整對身心障礙學生的效果是否會因「學科」而有差異？

2-2. 考試調整對身心障礙學生的效果是否會因「考試調整策略類型」而有差異？

2-3. 考試調整對身心障礙學生的效果是否會因其「教育階段、障礙特性和人數」而有差異？

三、名詞解釋

(一) 身心障礙學生

《IDEIA》2004 年指出身心障礙包括了以下 14 個類別：學障、智能障礙、情緒困擾、說話或語言損傷、廣泛性自閉症、發展遲緩、形體損傷、其他健康損傷、聽覺損傷、聾、視覺損傷（包含全盲）、盲／聾、多重障礙、頭部外傷（引自 Smith, 2007, pp. 11-12）。本研究所指身心障礙乃依據上述定義，在 1999 至 2008 年間於美國接受考試調整的對象。

(二) 考試調整策略成效

調整之英文為「accommodation」，它和「修改」(modification) 不同處在於修改是指改變試題欲評量之構念；而調整則不會改變它們 (Stretch & Osborne, 2005)。Wasburn-Moses (2003) 指出考試調整旨在評量學生的能力，而不是障礙。Byrnes (2005) 則表示它在撤除學生因障礙造成學習和接受考試的阻力，而未改變考試活動的必要目的。考試調整僅改變「試題形式」或「測驗程序」，使身心障礙學生不因其障礙阻撓其參與考試 (Fuchs & Fuchs, 1999; Tindal & Fuchs, 2000)。本研究的考試調整策略成效是指測驗成績，包括了州立標準參照測驗、標準化測驗，以及論文作者自編的評量工具。

(三) 調節變項

調節變項又稱為次級自變項 (secondary

independent variable)，它會影響主要自變項與依變項關係的方向和強度 (Baron & Kenny, 1986)。本研究以學科 (閱讀、數學和其他科目, 其他包含了社會、科學或未特別指明的科目)、考試調整策略類型 (報讀和延長時間兩種), 以及身心障礙學生的教育階段 (國小、國中和高中職)、障礙特性 (全體身心障礙和個別學障學生), 以及人數 (100 人以下、100 至 1000 人, 以及 1000 人以上三組) 為調節變項。

考試調整策略類型中, 報讀是改變試題的呈現方式, 將文字改變為口頭的方式呈現, 其形式包括人工報讀及機器 (含錄音機或電腦軟體) 報讀, 較常運用於閱讀題目有困難之身心障礙學生 (Thurlow & Bolt, 2001)。延長時間係指增加應試時間, 讓因閱讀和訊息處理速度較慢的身心障礙學生, 能有較充裕的時間作答 (Lewandowski et al., 2008)。

文獻探討

一、考試調整之重要理論

為提供本研究後設分析之論述基礎, 以下簡述考試調整之重要理論, 以作為本研究後續討論之依據。

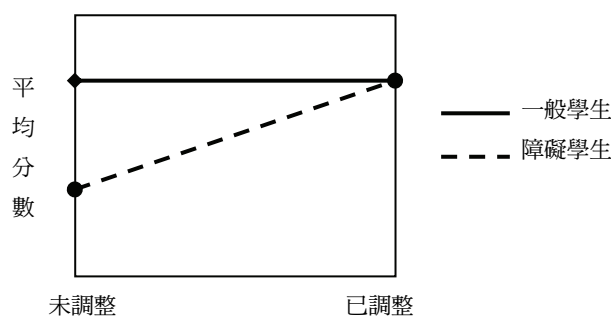
(一) 最大潛能論點 (maximum potential thesis, MPT)

Zuriff (2000) 提出最大潛能論點, 主張一般學生在延長時間的調整中並不會獲益, 因為他們已在有時限的考試中發揮最大能力了; 反觀學障學生由於處理訊息上的緩慢, 導致在時間限制下無法完全發揮潛能。Stretch 和 Osborne (2005) 進一步說明研究如何運用「最大潛能論點」評鑑考試調整的成效, 他們認為研究應包括兩個成果: (1) 學障學生在未限制考試時間的情境下, 其分數會顯著高於有時間限制者; (2) 一般學生在未限制時間的情境中不會有較高的分數。

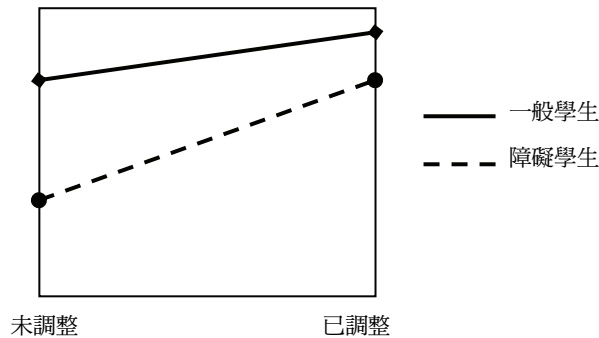
上述論點又稱為「互動假設」(interaction hypothesis), 並擴展此概念至所有類型之考試調整, 假定提供考試調整給身心障礙學生時, 會提升其分數; 而一般學生即使接受調整也不會有較高的分數 (Gregg, 2008; Sireci et al., 2005), 如圖一。

(二) 差別增長假設 (differential boost hypothesis)

由於互動假設界定一般學生的表現不會有任何改變, 過於嚴謹, 因此 Tindal 和 Fuchs (2000) 提出「差別增長」假設, 它強調有效的考試調整為身心障礙學生因調整而進步的分數, 優於一般學生在相同調整下進步者 (Finch, Barton, & Meyer, 2009), 如圖二, 而此觀點也被後續研究者引用並驗證 (Koenig & Bachman, 2004)。互動假設和差別增長的差



圖一 互動假設圖示



圖二 差別增長圖示

異在於：差別增長認為一般學生透過調整分數仍會進步，互動假設則否。本研究將利用這兩項觀點，比較考試調整對一般與身心障礙學生的成效，以驗證之。

(三) 公平與權益的平衡 (leveling of the playing field)

公平與權益平衡的觀點強調考試調整允許身心障礙學生展現其所能，但不會使其超越一般學生的表現，兼顧了身心障礙學生的權益和一般學生的公平 (Elliott, Kratochwill, & McKevitt, 2001, 2002; Fuchs & Fuchs, 2001; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000)。Elliott 等人 (2002) 表示考試調整就如同戴上眼鏡，能讓近視者考駕照路考時看得清楚，而非直接取得駕照；因此身心障礙學生接受考試調整後，無法超越一般學生未調整的表現，符合公平原則。本研究以此觀點討論身心障礙學生接受考試調整後，是否會超越一般學生未調整的表現，以驗證之。

二、考試調整成效的分析研究

有關考試調整成效的分析研究，有三篇採取投票法 (voting method) (Johnstone et al., 2006; Thompson et al., 2002; Zenisky & Sireci, 2007)，一篇採用敘述性的文獻回顧 (narrative reviews) (Sireci et al., 2005)。兩種方法均有其

限制，敘述性的文獻回顧不能整合和比較不同的研究發現；投票法則是容易落入顯著與不顯著研究數量之比較，且容易造成「勝者全拿」的結果，忽略了研究對象人數、顯著水準和研究品質不同等議題，導致結論有偏差 (Kavale & Glass, 1981; Rosenthal, 1991)。目前僅兩篇考試調整之後設分析研究 (Chiu & Pearson, 1999; Elbaum, 2007)。

Chiu 和 Pearson (1999) 分析 1986 至 1998 年間，30 篇考試調整之研究；結果發現延長時間為最常運用的策略，而學障學生為所有障礙中最常被探討的類別；調整能有效增進身心障礙學生成績的效果量為 0.16，一般學生的加權效果量為 0.06，身心障礙學生優於一般學生調整之效果。另外，該研究也分別分析考試調整成效是否會因學生障礙類別和年級，以及調整策略類型而有差異；結果顯示「時間調整」(包含延長時間等) 的加權效果量為 0.31，「呈現方式調整」(包含報讀等) 的加權效果量為 -0.12，「綜合調整」則為 0.1，時間調整策略的效果優於呈現方式的調整。教育階段方面以跨階段的調整效果最佳 (0.67)，國小 (幼稚園到小五) 的加權效果量為 -0.07，國中 (7 到 8 年級) 為 0.14，高中職 (9 到 12 年級) 為 0.22，高中職後則為 0.28。障礙類別中，不分類的效果量為 0.01，注意力缺陷過動症為

0.47，學障為 0.12，肢體障礙為 0.19，視覺障礙則為 0.62；整體來說，以對視障學生的效果最佳，可能與調整策略如報讀、放大字體等方式直接可協助視障學生有效作答所致。對一般學生來說，除了時間調整策略較有效外，其他調整策略並無太大效果。

Elbaum (2007) 分析 1998 至 2004 年間，比較報讀策略對學障和一般學生數學成效的 17 篇研究（其中有兩篇無對照組），總平均效果量為 0.03，國中小階段（8 篇）的效果量為 0.20，95%信賴區間為 0.10 到 0.30；而高中職階段（6 篇）的效果量為-0.12，95%信賴區間為-0.03 到-0.21。然而，Elbaum 的研究未計算加權效果量，無法考量研究對象人數帶來的影響。

總之，兩篇後設分析研究發現：考試調整對於身心障礙學生均有助益，但屬於小效果量；而就教育階段言之，Chiu 和 Pearson (1999) 和 Elbaum (2007) 的研究結果剛好相反，前者發現教育階段愈高效果愈好，後者則顯示國小的效果優於中學階段。而 Kim、Schneider 和 Siskind (2009) 認為雖然考試調整在教育階段上的效果無一致結論，但隨著學生年級增加，調整效果隨之減弱的趨勢似乎難

以避免。

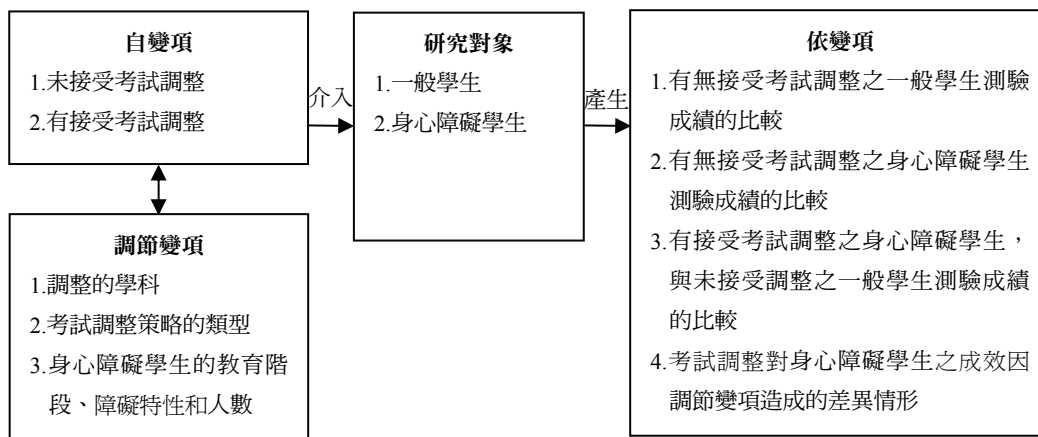
綜合文獻，考試調整成效的後設分析實有必要，而美國至 2008 年為止，僅兩篇考試調整成效之後設分析研究；因此，本研究後設分析美國 1999 至 2008 年間，考試調整對身心障礙學生的成效，在針對身心障礙學生分析時，還會依據互動和差別增長假設，以及公平與權益平衡的觀點，分析考試調整對一般學生的成效，並且比較考試調整對身心障礙與未接受調整之一般學生的成效差異，以了解考試調整是否讓身心障礙學生獲得有效成績，又不會損及對一般學生的公平性。接著分析考試調整的效果是否會因學科、策略類型，以及身心障礙學生的障礙特性、教育階段和人數而有差異。

研究方法

本研究依據 Littell、Corcoran 和 Pillai (2008) 對後設分析之建議，採取以下步驟。

一、確立研究架構

本研究依據目的，確立研究架構如圖三。



圖三 研究架構

二、界定文獻搜尋的來源和範圍

(一) 文獻搜尋的來源

爲了減少出版偏差的問題，研究者利用以下三種方式蒐集發表之期刊論文、未發表的碩博士論文，以及政府和會議報告等：一爲電腦資料庫查詢，搜尋年代從 1999 至 2008 年底爲止，以關鍵字 (test accommodation、oral presentation、read aloud、extended time、disabilities、learning disabilities、large assessment、high-stakes testing) 搜尋 EBSCO、ERIC、PQDT、ProQuest 等電子資料庫系統。二爲人工回溯查詢，美國國家教育成果中心分別於 1999 至 2001 (Thompson et al., 2002)、2002 至 2004 (Johnstone et al., 2006)，以及 2005 至 2007 年間 (Zenisky & Sireci, 2007)，回顧身心障礙學生考試調整的研究，搜尋三篇文獻中所列之研究。三爲從研究中的參考文獻查詢，在蒐集上述研究後，檢視所列之參考文獻，尋找額外的研究。若有兩篇或以上的文獻爲相同資料來源時，採用最早發表之文獻。

(二) 文獻搜尋的範圍

文獻搜尋的範圍包括以下標準：

1. 符合本研究目的，亦即探討考試調整對一般和身心障礙學生的成效。

2. 選取 1999 至 2008 年間的研究，以延續 Chiu 和 Pearson (1999) 的研究，並比較之。

3. 選取研究對象爲一般和身心障礙學生，如僅說明是特殊班學生或 IEP 服務對象則不納入。

4. 選取方法爲實驗或準實驗研究，設計爲有兩組比較 (實驗組和對照組) 或前後測資料之研究，以便於進行後設分析，排除個案、相關、調查和單一受試研究。

5. 選取考試調整策略爲延長時間和報讀兩種之研究。

6. 提供計算效果量所需數據之研究，包括平均數、標準差、 t 、 r 或 F 值等。

依據上述六項標準，最後選取 38 篇研究，研究資料蒐集來源及選用情形如表一。

二、登錄研究資料並計算編碼者信度

本研究從以下三個部分登錄研究資料：

(1) 研究特徵，包括出版年代、作者、資料來源，以及研究對象、目的、方法和設計；

(2) 調節變項的特徵，包括學科、考試調整策略類型，以及身心障礙學生的教育階段、障礙特性和人數；(3) 計算效果量所需數據，包括平均數、標準差、 t 、 r 或 F 值等。

在完成登錄工作後，研究者委請一位特教系博士班研究生，從 38 篇研究中隨機選取 9 篇 (即 24%) 的文獻，登錄資料，之後計算編碼者信度 (coder's reliability)，結果顯示 $Kappa$ 一致性爲 .85。

表一 後設分析研究資料蒐集來源及選用情形

資料來源	篩選篇數	捨棄篇數	選取篇數
碩博士論文	4 (4%)	0 (0%)	4 (11.4%)
期刊論文	41 (66.1%)	17 (68.0%)	24 (60.0%)
政府報告	10 (16.1%)	3 (12.0%)	7 (20.0%)
會議報告	8 (12.9%)	5 (20.0%)	3 (8.6%)
合計	63 (100%)	25 (100%)	38 (100%)

三、決定資料分析和程序

本研究首先以 Hedges 和 Olkin (1985) 之技術計算各研究效果量 d 值。計算過程中，發現部分單一研究因採取不同的方式（多個評量工具、階段、調整策略類型）檢視考試調整成效，造成單一研究多重效果量的問題；本研究依據 Littell 等人 (2008) 的建議，將兩個以上之效果量平均，以獲得單一效果量，避免重複計算導致整體效果量偏高或偏低的現象。

而在實際分析上，由於整體效果量的分析每篇研究僅能提供一個效果量，並且不是每篇

研究都包含一般和身心障礙生，因此篇數不相同；而在調節變項之分析上，由於線性階層模式將研究間的變異程度納入考量，因此一篇研究可針對不同變項分別計算其效果量，例如一篇研究針對國小、國中、高中職三個教育階段做分析，即可分別計算三個效果量，因此各調節變項間的效果量數不同，茲整理整體與調節變項分析之研究篇數和效果量數如表二。

獲得各研究效果量 d 值後，以 Rosenthal (1991) 提出之 *Winer* 合併檢定 (combined test) 計算合併效果量，以了解是否需要合併計算所有研究之 d 值，並計算加權後之 d 值，

表二 整體與調節變項分析之研究篇數和效果量數

分析內涵	研究篇數	效果量數
整體效果量之分析		
一般學生	20	20
身心障礙學生	24	24
身心障礙學生與一般學生之比較	33	33
調節變項之分析		
學科	28	56
閱讀	11	19
數學	16	30
其他	6	7
考試調整策略的類型	28	56
報讀	17	26
延長時間	11	30
身心障礙學生的教育階段	28	56
國小	16	18
國中	13	18
高中職	7	20
身心障礙學生的人數	28	41
100 人以下	15	20
100 至 1000 人	8	16
1000 人以上	5	5
身心障礙學生的障礙特性	28	28
全體身心障礙學生	13	13
個別學障學生	15	15

給予研究對象人數較多者較高的「權數」(weight)，以避免研究偏差。之後計算 Cooper 之 *Fail Safe N* (N_{fs})，以考驗出版偏差可能的影響，另外在出版偏差的檢定上，一般來說，*Fail-Safe N* 值愈大，愈沒有出版偏差的因素在；若 *Fail-Safe N* 值比 Rosenthal 提出的「容忍層次」(tolerance level) ($5K + 10$) 大得多，則此結果具高度有效性(黃寶園、林世華，2002)。最後進行同質性檢定，林邦傑(1987)指出若結果未達顯著水準，則表示各研究間的差異為抽樣誤差所致，則直接合併效果量，計算效果量和 95%的信賴區間；若結果達顯著水準，則須進一步分析調節變項對於效果量之影響。效果量 d 值、*Winer* 合併檢定、*Fail Safe N* 和同質性檢定皆採用吳裕益(2009b)撰寫之 SPSS 語法。一般而言，調節變項之分析大多採用變異數分析或多元迴歸；但由於後設分析的資料本身具有階層結構，因此，Raudenbush 和 Bryk 於 2008 年建議利用「階層線性模式」(hierarchical linear modeling, HLM)，它不但可處理階層結構之需求，同時解決了後設分析中差異成分的問題，較變異數分析或多元迴歸為佳(引自吳裕益，2009a)。因此本研究以階層線性模式，分析考試調整成效因調節變項造成的差異。

本研究以 Cohen 於 1977 年之建議，將效果量 0.20、0.50 和 0.80 界定為小、中、大(引自 Littell et al., 2008)。除此，以 Hunter 和 Schmidt 於 1990 年發展之「 $U3$ 不重疊量數」，作為另一項效果量解釋指標，為效果量 d 的「考驗力分析」(power analysis)，其意義係指「較高平均群組超過較低平均群組多少百分比」(引自吳裕益，2009b)。另外在調節變項的分析中，無條件模式旨在評鑑真實效果參數中的變異情形，可估計真實效果的平均數和標準差；而條件模式則是解釋產生變異的來源，也就是估計真實效果的殘差變異數，在各調節變項中，截距的

選擇為隨意安排(吳裕益，2009a)。

研究結果與討論

於研究結果與討論中，首先針對考試調整對身心障礙和一般學生之成效，以及考試調整對身心障礙學生成效之調節變項做分析與討論，最後討論研究過程發現的問題。

一、考試調整對身心障礙和一般學生成效之分析與討論

(一) 考試調整對身心障礙學生之成效

在比較身心障礙學生有無接受考試調整之差異上，共納入 28 篇研究。*Winer* 合併檢定為 9.65 ($p = .00$)，顯示適合以合併效果量分析， d 為 0.39，加權後平均效果量 (d_w) 為 0.0304，二者間差異很大，檢視研究後發現有四篇效果量為負值，這四篇取樣自參加大型考試的人，為不等組設計，未接受考試調整之身心障礙學生人數較多(超過 1000 人)，而且可能為能力較佳，不須申請調整之學生。為避免此四篇造成研究結果之偏誤，因此刪除這四篇研究後重新分析，結果如表三。*Winer* 檢定值為 11.19 ($p = .00$)，表示適合以合併效果量分析， d 為 0.48，加權後平均效果量 (d_w) 為 0.31，95%信賴區間為 0.258 至 0.326， $U3$ 為 0.65，顯示身心障礙學生接受考試調整的表現，大約高於 65%未接受調整者，且高了 0.31 個標準差，身心障礙學生有無接受考試調整之差異為「小」效果量。另外，同質性檢定 $Q_{(23)}$ 為 485.74 ($p = .00$)，顯示效果量變異大於抽樣誤差，將進行調節變項之分析。出版偏差 $N_{fs(0.05)}$ 為 1019，只要再有 1019 篇研究無效果，合併檢定就無法達到 .05 之顯著水準；而 $N_{fs(0.01)}$ 為 495.9，只要再有 496 篇研究無效果，合併檢定就無法達到 .01 之顯著水準，遠超過 130 篇 ($5K+10$) 的容忍層次。

(二) 考試調整對一般學生之成效

在比較一般學生有無接受考試調整之差異上，共納入 20 篇研究，分析結果如表四。*Winer* 檢定值為 1.85 ($p = .00$)，表示適合以合併效果量分析， d 為 0.12， $U3$ 則為 0.55，亦即一般學生接受考試調整的表現，大約高於 55% 未接受調整者，且高了 0.12 個標準差；然加權後平均效果量 (d_w) 為 -0.003，95% 信賴區間為 -0.119 至 0.126，顯示一般學生有無接受考試調整之差異為「些微效果」。另外，同質性檢定 $Q_{(19)}$ 為 398.19 ($p = .00$)，顯示效果量變異大於抽樣誤差，可進行調節變項之分析，唯並非本研究探討之範圍，故不執行。出版偏差 $N_{J5(.05)}$ 為 4.87，表示只要再有 5 篇無效

果的研究，合併檢定就無法拒絕虛無假設；而 $N_{J5(.01)}$ 為 -7.61，表示如果要達到 .01 之顯著水準，須減少 8 篇沒有任何效果之研究。

(三) 考試調整對身心障礙的成效和一般學生未調整之比較

比較身心障礙學生接受考試調整和一般學生未接受調整之差異，以驗證調整是否能維持「公平與權益的平衡」，共納入 33 篇文獻，結果如表五。*Winer* 合併檢定為 -46.07 ($p = .00$)，達顯著水準，顯示效果量可以合併分析， d 為 -0.75， $U3$ 為 0.23，加權後平均效果量 (d_w) 為 -0.52，95% 信賴區間為 -0.514 至 -0.526，顯示一般學生未接受調整的表現，大約高於 23% 身心障礙學生接受調整者，且高了

表三 考試調整對身心障礙學生成效的後設分析 ($n = 24$)

後設分析內涵	身心障礙學生未調整和身心障礙學生有調整
<i>Winer</i> 檢定 Z_c (未加權)	11.19**
<i>Winer</i> 檢定 <i>Weighted</i> Z_c	9.65**
效果量 d	0.48
加權效果量 d_w	0.31
效果量信賴區間 95%CI	0.258~0.362
不重疊量數 $U3$	0.68
同質性檢定 Q	485.74**
出版偏差檢定 $N_{J5(.05)}$	1019
出版偏差檢定 $N_{J5(.01)}$	495.9

** $p < .01$ **表四 考試調整對一般學生成效的後設分析 ($n = 20$)**

後設分析內涵	一般學生未調整和一般學生有調整
<i>Winer</i> 檢定 Z_c (未加權)	1.85**
<i>Winer</i> 檢定 <i>Weighted</i> Z_c	0.72**
效果量 d	0.12
加權效果量 d_w	0.014
效果量信賴區間 95%CI	0.062~-0.034
不重疊量數 $U3$	0.55
同質性檢定 Q	398.19**
出版偏差檢定 $N_{J5(.05)}$	4.87
出版偏差檢定 $N_{J5(.01)}$	-7.61

** $p < .01$

表五 考試調整對身心障礙的成效和一般學生未調整相較之後設分析 ($n = 33$)

後設分析內涵	身心障礙學生有調整和一般學生未調整
Winer 檢定 Z_c (未加權)	-46.07**
Winer 檢定 $Weighted Z_c$	-43.01**
效果量 d	-0.75
加權效果量 dw	-0.52
效果量信賴區間 95% CI	-0.516~ -0.524
不重疊量數 $U3$	0.23
同質性檢定 Q	24496.79**
出版偏差檢定 $N_{j5(.05)}$	26003.3
出版偏差檢定 $N_{j5(.01)}$	12944.78

** $p < .01$

0.52 個標準差，二者間呈現「中」效果量的差異。另外，效果量同質性檢定 $Q_{(32)}$ 為 24496.79 ($p = .00$)，顯示效果量變異大於抽樣誤差，可進行調節變項之分析，唯並非本研究探討之範圍，故不執行。出版偏差 $N_{j5(.05)}$ 為 26003.3，表示只要再有 26003 篇研究無效果，合併檢定就無法達到 .05 之顯著水準；而 $N_{j5(.01)}$ 為 12944.78，表示只要再有 12945 篇研究無效果，合併檢定就無法達到 .01 之顯著水準，遠超過可容忍層次的 175 篇。

整體而言，考試調整對身心障礙學生為小效果量；對一般學生的成效為些微效果，符合「差別增長」和「互動」假設；比較身心障礙學生接受考試調整和一般學生未接受調整，二者間呈現「中」效果量的差異，顯示考試調整能維持「公平與權益的平衡」，身心障礙學生接受考試調整並不會超過一般學生的表現，而此結果與 Chiu 和 Pearson (1999) 的研究類似。

二、考試調整對身心障礙學生成效之調節變項分析與討論

以學科、考試調整策略類型，以及身心障礙學生的教育階段、人數及障礙特性為調節變項，分析考試調整對身心障礙學生的效果是否會因這些調節變項而有差異。

(一) 學科

以學科為調節變項，共納入 28 篇研究，56 個效果量，如表六。先進行無條件模式檢定，效果量總平均估計值為 0.39，顯示接受調整的身心障礙學生平均數較未接受調整者高 0.39 個標準差，屬於小效果量，而變異數估計值為 0.21，呈現各篇研究真正效果量 0.21，相當於標準差 0.46， x^2 為 1760.38 ($p = .00$)，表示考試調整對身心障礙學生的效果會因「學科」而有顯著差異。另外以條件模式進行檢定，以其他科目作為截距，係數為 0.57， t 為 3.5 ($p = .00$)，達顯著水準；而閱讀和數學兩科分別低於其他科目 0.26 和 0.15 個標準差，且均未達顯著水準，顯示其他學科的調整效果最佳，其次為閱讀（係數為 0.42），最低為數學（係數為 0.31）。這可能由於閱讀與數學兩科須具備認知能力，而身心障礙學生在認知能力上的限制，使他們在閱讀和數學科上的調整效果未明顯呈現；加上其他科目的效果量數僅 7 個，閱讀與數學兩科的效果量數亦有一些差距（分別為 19 和 30 個），恐因比較之效果量數的差距影響結果。

(二) 考試調整策略的類型

以考試調整策略的類型為調節變項，共納入 28 篇研究，56 個效果量，如表七。先進行

表六 無條件和條件模式下考試調整在學科上之成效差異分析 ($n = 56$)

固定效果	係數	標準誤	t	自由度	p
無條件模式下					
學科總平均值, γ_0	0.39	0.07	5.77**	55	.00
條件模式下					
截距 γ_0 (其他學科)	0.57	0.16	3.5**	53	.00
數學	-0.26	0.16	-1.61	53	.11
閱讀	-0.15	0.22	-0.68	53	.50
隨機效果	標準差	變異數成分	自由度	χ^2	p
無條件模式下					
學科真實效果量, U_0	0.46	0.21	55	1760.38**	.00
條件模式下					
排除學科後效果量, U_0	0.46	0.21	53	1804.53**	.00

** $p < .01$ 表七 無條件和條件模式下考試調整在策略類型上之成效差異分析 ($n = 56$)

固定效果	係數	標準誤	t	自由度	p
無條件模式下					
調整策略類型總平均值, γ_0	0.39	0.07	5.7**	55	.00
條件模式下					
截距 γ_0 (延長時間)	0.51	0.11	4.5**	54	.00
報讀	-0.20	0.14	-1.4	54	.15
隨機效果	標準差	變異數成分	自由度	χ^2	p
無條件模式下					
調整策略類型真實效果量 U_0	0.46	0.21	55	1760.38**	.00
條件模式下					
排除調整策略類型後效果量 U_0	0.45	0.21	54	1526.96**	.00

** $p < .01$

無條件模式檢定，效果量總平均估計值為 0.39，顯示接受調整的身心障礙學生平均數較未接受調整者高 0.39 個標準差，屬於小效果量，而變異數估計值為 0.21，呈現各篇研究真正效果量 0.21，相當於標準差 0.46， χ^2 為 1760.38 ($p = .00$)，表示考試調整對身心障礙學生的效果會因「考試調整策略的類型」而有顯著差異。另外以條件模式進行檢定，以延長時間作為截距，係數為 0.51， t 為 4.5 ($p = .00$)，達顯著水準；而報讀低於延長時間 0.20 個標準差（係數為 0.31），未達顯著水準。此結果與

Chiu 和 Pearson (1999) 的研究類似，時間調整優於呈現方式調整，唯值得注意的是：他們研究的調整策略較本研究的範圍廣，解釋上須謹慎。Elbaum (2007) 指出報讀不一定對所有身心障礙學生均有益，它僅對識字有困難的學生有幫助，對於能識字但閱讀理解或書寫表達有困難學生，報讀將增加聽覺處理的負擔，反而會干擾他們的思考，影響作答速度，助益較小。Fuchs、Fuchs、Hamlett、Eaton、Binkley 和 Crouch 等人 (2000) 雖認為報讀的成效優於延長時間，但 Elbaum 指出隨著年級的增

加，由於學障學生在基礎能力上的低落，報讀成效反而不明顯。

(三) 身心障礙學生的教育階段

以身心障礙學生的教育階段為調節變項，共納入 28 篇研究，56 個效果量，如表八。先進行無條件模式檢定，效果量總平均估計值為 0.39，顯示接受調整的身心障礙學生平均數較未接受調整者高 0.39 個標準差，屬於小效果量，而變異數估計值為 0.21，呈現各篇研究真正效果量 0.21，相當於標準差 0.46， x^2 為 1760.38 ($p = .00$)，表示考試調整對身心障礙學生的效果會因「身心障礙學生的教育階段」而有顯著差異。另外以條件模式進行檢定，以高

中職作為截距，係數為 0.29， t 為 2.6 ($p = .01$)，達顯著水準；而國中和國小分別高於高中職 0.23 和 0.08 個標準差，但均未達顯著水準，顯示國中的調整效果最佳（係數為 0.52），其次為國小（係數為 0.37），最低為高中職。此結果與 Chiu 和 Pearson (1999) 的研究部分類似，國中優於國小，可能因為國中課程難度較高，學生對考試調整的需求相對較高，能顯現有無調整之成效差異，但對高中職的成效則不像 Chiu 和 Pearson 的研究高於國中小。

(四) 身心障礙學生的人數

以身心障礙學生的人數為調節變項，共納入 28 篇研究，41 個效果量，如表九。先進行

表八 無條件和條件模式下考試調整在研究對象教育階段上之成效差異分析 ($n = 56$)

固定效果	係數	標準誤	t	自由度	p
無條件模式下					
教育階段總平均值 γ_0	0.39	0.07	5.77**	55	.00
條件模式下					
截距 γ_0 (高中職)	0.29	0.11	2.60*	53	.01
國中	0.23	0.17	1.40	53	.18
國小	0.08	0.16	0.50	53	.61
隨機效果	標準差	變異數成分	自由度	x^2	p
無條件模式下					
教育階段真實效果量 U_0	0.46	0.21	55	1760.38**	.00
條件模式下					
排除教育階段後效果量 U_0	0.46	0.21	53	1860.43**	.00

* $p < .05$; ** $p < .01$

表九 無條件和條件模式下考試調整在身心障礙學生人數上之成效差異分析 ($n = 41$)

固定效果	係數	標準誤	t	自由度	p
無條件模式下					
人數總平均值 γ_0	0.33	0.063	5.19**	40	.00
條件模式下					
截距 γ_0 (100 人以下)	0.55	0.09	6.18**	38	.00
100 人以上到 1000 人以下	-0.30	0.12	-2.61*	38	.01
1000 人以上	-0.65	0.13	-5.10**	38	.00
隨機效果	標準差	變異數成分	自由度	x^2	p
無條件模式下					
人數真實效果量 U_0	0.36	0.13	40	768.35**	.00
條件模式下					
排除研究對象人數後效果量 U_0	0.27	0.08	38	228.85**	.00

* $p < .05$; ** $p < .01$

無條件模式檢定，效果量總平均估計值為 0.33，顯示接受調整的身心障礙學生平均數較未接受調整者高 0.33 個標準差，屬於小效果量，而變異數估計值為 0.13，呈現各篇研究真正效果量 0.13，相當於標準差 0.36， χ^2 為 768.35 ($p = .00$)，表示考試調整對身心障礙學生的效果會因「身心障礙學生的人數」而有顯著差異。另外以條件模式進行檢定，以 100 人以下的人數作為截距，係數為 0.55， t 為 6.18 ($p = .00$)，達顯著水準；而 100 人以上到 1000 人以下和 1000 人以上分別低於 100 人以下 0.30 和 0.65 個標準差，均達顯著水準，顯示 100 人以下人數的調整效果最佳，其次為 100 人以上到 1000 人以下（係數為 0.25），最低為 1000 人以上（係數為 -0.10）。這可能原因為在人數少的情況下，考試調整策略的設計較符合學生的需求，也較能有效掌控調整策略的實施。

(五) 身心障礙學生的障礙特性

以身心障礙學生的障礙特性為調節變項，共納入 28 篇研究，28 個效果量，如表十。先進行無條件模式檢定，效果量總平均估計值為 0.33，顯示接受調整的身心障礙學生平均數較

未接受調整者高 0.33 個標準差，屬於小效果量，而變異數估計值為 0.19，呈現各篇研究真正效果量 0.19，相當於標準差 0.44， χ^2 為 760.99 ($p = .00$)，表示考試調整對身心障礙學生的效果會因「身心障礙學生的障礙特性」而有顯著差異。另外以條件模式進行檢定，以全體身心障礙學生作為截距，係數為 0.28， t 為 2.44 ($p = .02$)，達顯著水準；而學障學生高於全體身心障礙學生 0.12 個標準差（係數為 0.40），未達顯著水準。此結果與 Chiu 和 Pearson (1999) 的研究類似，均顯示考試調整對不分障礙類別學生的效果較低；可能原因為身心障礙學生異質性大，不分類別進行調整，無法考量不同障礙的特徵和需求，例如感官和動作障礙與心智障礙者（例如智障、學障）間特徵和需求有差異，調整效果可能存在不一致的現象。

本研究實施過程中發現以下問題，首先，部分文獻如會議論文或其他出版品在搜尋過程中僅有篇名，無法從資料庫中取得，對於全面了解考試調整的成效，有不足之處。其次，本研究考試調整對身心障礙學生成效之調節變項分析中，其他科目和用在 1000 人以上人數的

表十 無條件和條件模式下考試調整在研究對象障礙特性上之成效差異分析 ($n = 28$)

固定效果	係數	標準誤	t	自由度	p
無條件模式下					
研究對象障礙特性總平均值 γ_0	0.33	0.09	3.68**	27	.00
條件模式下					
截距 γ_0 (全體身心障礙)	0.28	0.11	2.44*	26	.02
學障	0.12	0.18	0.70	26	.49
隨機效果	標準差	變異數成分	自由度	χ^2	p
無條件模式下					
研究對象障礙特性真實效果量 U_0	0.44	0.19	27	760.99**	.00
條件模式下					
排除研究對象障礙特性後效果量 U_0	0.44	0.20	26	668.20**	.00

* $p < .05$; ** $p < .01$

研究篇數和效果量數都較少，故在解釋研究結果上宜謹慎，另需要更多研究探討考試調整對身心障礙學生其他科目，以及接受調整人數多之應用成效。再者，研究者發現研究對象的學科能力及智力，可能會影響調整策略的成效；但多數研究未提供這兩方面的資料，也無法確認比較之組間研究對象的學科能力及智力是否相似。因此，調整策略的成效是否會因研究對象的學科能力與智力而有差異，值得探究。最後，部分研究未呈現各障礙類別之考試調整策略成效資料，因此，無法分析調整策略的成效是否因研究對象的障礙類別而有差異，此部分的研究值得探討。

研究結論與建議

一、研究結論

本文旨在後設分析美國 1999 到 2008 年間，38 篇考試調整策略對一般和身心障礙學生的成效；接著分析考試調整的效果是否會因學科、策略類型，以及身心障礙學生的障礙特性、教育階段和人數而有差異。結論如下：

1. 在考試調整對一般和身心障礙學生的成效上，對一般學生為些微效果；對身心障礙學生為小效果量，符合「差別增長」和「互動」假設。未接受考試調整之一般學生的測驗成績，高於接受調整之身心障礙學生，二者間呈現「中」效果量的差異，顯示考試調整能維持「公平與權益的平衡」，身心障礙學生接受考試調整並不會超過一般學生的表現。

2. 在考試調整對身心障礙學生之成效因調節變項造成的差異上，調整的科目以其他學科最佳，其次分別為閱讀和數學。於調整策略上，以延長時間的效果較佳，報讀較差。在考試調整實施的教育階段上，以國中最佳，國小次之，而高中職效果最差。關於身心障礙學生的人數，考試調整對 100 人以下人數的效果最

佳，其次為 100 以上到 1000 人以下，最低為 1000 人以上。最後，考試調整對學障學生的效果，優於全體身心障礙學生。

二、研究建議

(一) 對實施考試調整策略之建議

1. 本研究發現考試調整可提升身心障礙學生的測驗成績，而且對國中的效果最好，國小次之，因此，建議對學生有困難的學科，平時評量即提供相同的調整策略，以協助學生更有效運用考試調整策略，提升考試調整的效果。

2. 從研究結果來看調整可提升障礙學生考試成績，因此為了讓學生能更有效運用調整，應該在平常教學過程即提供相同調整策略，除可增進教學成效外，亦可使考試調整效果提升，透過考試前的適當訓練，使學生得以有效運用調整策略，達成事半功倍之效。

3. 以考試調整策略在各教育階段的結果來看，年級越小效果較佳，就如同馬太效應 (Matthew's effect) 「貧者愈貧，富者愈富」所言，學習上的困難使得他們在自尊和動機上產生低落，同時讓他們在閱讀能力、獲得字彙和閱讀理解等能力上漸趨落後，因此在平日應提供補救教學和培養閱讀能力使學生獲得應具備之學科能力來面對考試。

4. 調整的提供應視學生的需求與考試的目的，如同時提供二個以上的調整，有時反造成學生負擔，僅提供學生真正需要的評量調整策略，太多的調整將造成學生本身的負擔，因此當調整策略超過一種以上時，必須確認這些調整都是適合學生需求的。

4. 早期發現，早期介入

在教育階段上的結果發現，愈早提供調整，對身心障礙學生的助益也愈大，如何早期發現其考試調整需求，並加以判斷提供服務，使學生在考試過程提升動機，獲得成就感，可使學生學習成果與努力成正比。

(二) 對未來研究之建議

1. 建議未來考試調整研究的設計，能納入調節變項（例如：學科、策略類型，以及身心障礙學生的障礙類別、教育階段、學科能力和人數），形成多因子設計，以了解調節變項與考試調整的交互作用。而後由調節變項的分析為基礎，進一步建構身心障礙學生考試調整的決策流程。

2. 建議未來考試調整的研究，可以加入除了報讀和延長時間以外調整策略的探討。

3. 建議在國內實施考試調整對身心障礙學生成效之研究，以累積考試調整效果之實徵研究基礎。

4. 建議未來細分身心障礙類別，以了解考試調整對不同身心障礙類別的成效。另外，學障學生異質性大，建議細分特定障礙類型，例如閱讀障礙、書寫障礙、數學障礙等，以了解考試調整對不同學障類型的成效。

5. 建議未來可針對兩個以上的調節變項，運用階層線性模式了解變項間的關聯性，例如考試調整策略類型和科目等變項間如何影響效果量。

參考文獻

（文獻前加註星號*者為本研究納入後設分析之論文）

吳裕益（2009a）：**HLM 在整合分析之應用——變異數已知的分析**。未出版的手稿，國立高雄師範大學特殊教育學系。

吳裕益（2009b）：**後設分析**。未出版的手稿，國立高雄師範大學特殊教育學系。

李佩蓉（2007）：**南部三縣市國中小普通班教師對身心障礙學生考試調整方式的看法與實施之調查研究**。國立臺南大學特殊教育研究所碩士論文（未出版）。

林邦傑（1987）：**整合分析的理論及其在國內**

的應用。教育與心理研究，10，1-38。

林筱汶、陳明聰（2006）：**語音調整策略對不同能力學生數學文字題解題表現成效之研究**。載於國立嘉義大學特殊教育系主編：**2006 年特殊教育國際學術研討會論文集**（113-131 頁）。嘉義：編者。

洪靜怡（2008）：**國小普通班教師對學習障礙學生實施評量調整之調查研究**。國立臺中教育大學特殊教育學系碩士論文（未出版）。

教育部（2009）：**特殊教育法**。臺北：教育部。

莊虹姿（2008）：**國民小學資源班教師對身心障礙學生評量調整意見之研究**。國立臺中教育大學特殊教育學系碩士論文（未出版）。

陳明聰、張靖卿（2004）：**特殊教育工作者對身心障礙學生測驗調整意見之調查研究**。**特殊教育與復健學報**，12，55-80。

陳蓓蓉（2003）：**探討全盲生處理國中基測數學題表現的外在因素及內在因素**。國立臺灣師範大學科學教育研究所碩士論文（未出版）。

鈕文英（2009）：**如何運用後設分析方法檢視證據本位實務**。未出版的手稿，國立高雄師範大學特殊教育學系。

黃巧雲、陳明聰、陳政見（2007）：**語音合成與真人錄音對國小高年級低閱讀能力數學文字題解題表現差異之研究**。載於國立臺東教育大學特殊教育系主編：**特殊教育學術研討會論文集**（81-106 頁）。臺東：編者。

黃鳳慈（2006）：**台南地區國民中學學習障礙學生評量調整措施之探討**。國立彰化師範大學特殊教育學系碩士論文（未出版）。

黃寶園、林世華（2002）：**合作學習對學習效果影響之研究——統合分析**。**教育心理學報**，34(1)，21-42。

- 劉婷妤、陳明聰、陳政見 (2008)：不同電腦化試題呈現方式對國小五年級學生在自然科成就測驗表現影響之研究。載於國立臺東教育大學特殊教育系主編：九十七年度東台灣特殊教育學術研討會論文集 (23-40 頁)。臺東：編者。
- 鍾素香 (2004)：學習障礙學生家長對評量調整的看法。載於國立臺灣師範大學特殊教育系主編：2004 年「追求研究卓越」特殊教育學術研討會論文集 (100-105 頁)。臺北：編者。
- *Baker, J. S. (2006). Effect of extended time testing accommodations on grade point averages of college students with learning disabilities. *Dissertation Abstracts International*, 67(1), 574 B.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bielinski, J., Ysseldyke, J. E., Bolt, S., Friedebach, M., & Friedebach, J. (2001). Prevalence of accommodations for students with disabilities participating in a statewide testing program. *Assessment for Effective Intervention*, 26, 21-28.
- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education*, 25(3), 141-152.
- *Bolt, S. E., & Ysseldyke, J. E. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education*, 19(4), 329-355.
- Bolt, S. E., & Ysseldyke, J. E. (2008). Accommodating students with disabilities in large-scale testing: A comparison of differential item functioning (DIF) identified across disability types. *Journal of Psychoeducational Assessment*, 26(2), 232-238.
- *Booth, D. A. (2003). The effects of tape recorded testing for college students with and without learning disabilities in reading and processing speed. *Dissertation Abstracts International*, 54(01), 534B.
- Byrnes, M. (2005). *Taking sides: Clashing views on controversial issues in special education* (2nd ed.). Chicago, IL: Irwin Press.
- *Camara, W. J., & Schneider, D. (2000). *Testing with extended time on the SAT I: Effects for students with learning disabilities*. New York: College Entrance.
- *Cahalan, C., Mandinach E., & Camara, W. J. (2002). *Predictive validity of SAT I: Reasoning test for test-takers with learning disabilities and extended time accommodations*. New York: The College Reporting Board.
- *Cahalan-Laitusis, C., King, T., Cline, F., & Bridgeman, B. (2006). *Observational timing study on the SAT reasoning test for test takers with learning disabilities and/or ADD/ADHD*. New York: The College Entrance Examination Board.
- *Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly*, 23, 271-282.
- Chiu, C. W. T., & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommoda-*

- tions for special education and limited English proficiency students. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.
- *Cohen, A. S., Gregg, N., & Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research & Practice, 20*(4), 225-233.
- Cook, T. D., Cooper, H., Cordary, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1994). *Meta-analysis for explanation: A casebook*. New York: Russell Sage Foundation.
- *Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time? *Journal of Learning Disabilities, 37*(2), 132-142.
- DeCoster, J. (2004). *Meta-analysis notes*. Retrieved May 21, 2009, from <http://www.stat-help.com>.
- *Dempsey, K. M. (2004). *The impact of additional time on SAT scores: Does time really matter? The efficacy of making decisions on a case-by-case basis*. Unpublished doctoral dissertation, La Salle University, Philadelphia, PA.
- *Dolan, R. P., Hall, T. E., Bannerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test design: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *The Journal of Technology, Learning, and Assessment, 3*(7). Retrieved May 23, 2009, from <http://escholarship.bc.edu/jtla/>.
- *Elbaum, B. (2007). Effect of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *The Journal of Special Education, 40*(4), 218-229.
- *Elbaum, B., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality, 12*(2), 71-87.
- Elliott, J. L., Thurlow, M. L., & Ysseldyke, J. E. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments* (Synthesis Report 25). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- *Elliott, S. N., Kratochwill, T. R., & McKeivitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39*, 3-24.
- Elliott, S. N., Kratochwill, T. R., & McKeivitt, B. C. (2002). Testing accommodation and decision making: The case of good scores being highly valued but difficult to achieve for all students. *Measurement and Evaluation in Counseling and Development, 35*, 153-166.
- *Elliott, S. N., & Marquart, A. M. (2003). *Extended time as an accommodation on a standardized mathematics test: An investigation of its effects on scores and perceived consequences for students with varying mathematical skills*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Finch, H., Barton, K., & Meyer, P. (2009). Differ-

- ential item functioning analysis for accommodated versus nonaccommodated students. *Educational Assessment, 14*(1), 38-56.
- *Fletcher, J. M., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children, 72*(2), 136-150.
- Fuchs, L. S., & Fuchs, D. (1999). Fair and unfair testing accommodations. *School Administrator, 56*(10), 24-29.
- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice, 16*, 174-181.
- Fuchs, L. S., Fuchs, D., & Capizzi, A. M. (2006). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children, 37*, 1-8.
- Fuchs, L. S., Fuchs, D., Eaton, S., Hamlett, C. L., & Karns, K. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29*, 65-85.
- *Fuchs, L. S., Fuchs, D., Hamlett, C., Eaton, S. B., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*, 67-81.
- Gregg, N. (2008). *Adolescents and adults with learning disabilities and ADHD: Assessment and accommodation*. New York: The Guilford Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- *Helwig, R., Rozek-Tedesco, M. A., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *The Journal of Special Education, 36*(1), 39-47.
- *Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children, 69*(2), 211-225.
- Hollenbeck, K. (2005). Validity issues and decisions about testing accommodations. *Assessment for Effective Intervention, 31*, 7-17.
- *Huesman, R., & Frisbie, D. (2000, April). *The validity of ITBS reading comprehension test scores for learning disabled and non learning disabled students under extended-time conditions*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- *Huynh, H., & Barton, K. E. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education, 19*(1), 21-39.
- *Huynh, H., Meyer, J. P., & Gallant-Taylor, D. (2002, April). *Comparability of scores of accommodated and non-accommodated testing for a high school exit examination of mathematics*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- *Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21*, 261-267.
- Johnstone, C. J., Altman, J., Thurlow, M. L., & Thompson, S. J. (2006). *A summary of research on the effects of test accommodations:*

- 2002 through 2004 (Technical Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved June 10, 2009, from <http://www.education.umn.edu/NCEO/OnlinePubs/Tech45/>
- Kavale, K. A., & Glass, G. V. (1981). Meta-analysis and the integration of research in special education. *Journal of Learning Disabilities, 14*, 531-538.
- Ketterlin-Geller, L. R., Alonzo, J., Braun-Monegan, J., & Tindal, G. (2007). Recommendations for accommodations: Implications of (in)consistency. *Remedial and Special Education, 28*, 194-206.
- Ketterlin-Geller, L. R., Yovanoff, P., & Tindal, G. (2007). Developing a new paradigm for accommodations research. *Exceptional Children, 73*(3), 331-347.
- Kim, D. H., Schneider, C., & Siskind, T. (2009). Examining the underlying factor test under oral and standard administrations. *Journal of Psychoeducational Assessment, 27*(4), 323-333.
- Lewandowski, L. J., Lovett, B. J., & Rogers, C. (2008). Extended time as a testing accommodation for students with reading disabilities: Does a rising tide lift all ships? *Journal of Psychoeducational Assessment, 26*(4), 315-324.
- Lindstrom, J. H. (2007). Determining appropriate accommodations for postsecondary students with reading and written expression disorders. *Learning Disabilities Research & Practice, 22*(4), 229-236.
- *Lindstrom, J. H., & Gregg, N. (2007). The role of extended time on the SAT for students with learning disabilities and/or attention-deficit hyperactivity disorder. *Learning Disabilities Research & Practice, 22*(2), 85-95.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, England: Oxford University Press.
- *Mandinach, E. B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). *The impact of extended time on SAT test performance* (Research Report No 2005-8). New York: The College Board.
- *Marquart, A. (2000, August). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels*. Paper presented at the Annual Meeting of the Council of Chief State School Officers, Snowbird, UT.
- *McKevitt, B. C. (2000, August). *The use and effects of testing accommodations on math and science performance assessments*. Paper presented at the Annual Meeting of the Council of Chief State School Officers, Snowbird, UT.
- *McKevitt, B. C., & Elliot, S. N. (2003). Effects and perceived consequences of using read aloud and teacher-recommended testing accommodations on a reading achievement test. *The School Psychology Review, 32*(4), 583-600.
- *Melo, L. L., Deville, C., & Frisbie, D. (2000, April). *The effect of a reading accommodation on standardized test scores of learning disabled and non learning disabled students*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

- Ofiesh, N. S., Hughes, C., & Scott, S. S. (2004). Extended test time and postsecondary students with learning disabilities: A model for decision making. *Learning Disabilities Research and Practice, 19*(1), 57-70.
- *Ofiesh, N. S., Mather, N., & Russell, A. (2005). Using speeded cognitive, reading, and academic measures to determine the need for extended test time among university students with learning disabilities. *Journal of Psychoeducational Assessment, 23*, 35-52.
- Phillips, S. E. (1994). High stakes testing accommodations: Validity vs. disabled rights. *Applied Measurement in Education, 7*(2), 93-120.
- *Pomplun, M., & Omar, M. H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *The Journal of Applied Psychology, 85*(1), 21-29.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Beverly Hills, CA: Sage.
- Schnirman, R. K. (2005). *The effect of audiocassette presentation on the performance of students with and without learning disabilities on a group standardized math test*. Unpublished doctoral dissertation, University of Florida, Gainesville, FL.
- *Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. *School Psychology Review, 30*(4), 527-547.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Centre for Educational Assessment Research Report No. 485.). Amherst, MA: School of Education, University of Massachusetts.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457-490.
- Smith, D. D. (2007). *Introduction to special education: Making a difference* (6th ed.). New York: Pearson Education Inc.
- Stretch, L. S., & Osborne, J. W. (2005). Extended test time accommodations: Directions for future research and practice. *Practical Assessment, Research, and Evaluation, 10*(8). Retrieved June 16, 2009, from <http://pareonline.net/pdf/v10/n8.pdf>.
- *Tavani, C. M. (2004). The impact of testing accommodations on students with learning disabilities: An investigation of the 2000 NAEP mathematics assessment. Unpublished doctoral dissertation, The Florida State University, Tallahassee, FL.
- Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved June 10, 2009, from <http://www.education.umn.edu/NCEO/OnlinePubs/Technical34.htm>.
- *Thornton, A. E., Reese, L. M., Pashley, P. J., & Dalessandro, S. P. (2002). *Predictive validity of accommodated SAT scores*. Pennsylvania, PA: Law School Admission Council.
- Thurlow, M. L., & Bolt, S. (2001). *Empirical*

- support for accommodations most often allowed in state policy* (Synthesis Report 41). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., House, A., Boys, C., Scott, D., & Ysseldyke, J. E. (2000). *State participation and accommodation policies for students with disabilities: 1999 update* (Synthesis Report 33). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., McGrew, K. S., Tindal, G., Thompson, S. L., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (Technical Report 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., Moen, R., & Altman, J. (2006). *Annual performance reports: 2003-2004 state assessment data*. Retrieved June 10, 2009, from <http://education.umn.edu/nceo/OnlinePubs/APR2003-04.pdf>
- *Tindal, G. (2002). *Accommodating mathematics testing using a videotaped, read-aloud administration*. Washington, DC: Council of Chief State School Officers.
- Tindal, G., & Fuchs, L. (2000). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: University of Kentucky, Mid-South Regional Resource Center.
- Tindal, G., Helwig, R., & Hollenbeck, K. (1999). An update on test accommodations: Perspectives of practice and policy. *Journal of Special Education Leadership, 12*(2), 11-20.
- Wasburn-Moses, L. (2003). What every special educator should know about high-stakes testing. *Teaching Exceptional Children, 35*(4), 12-15.
- *Weston, T. J. (2003). *NAEP validity studies: The validity of oral accommodation in testing*, (NCES 2003-06). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Zenisky, A. L., & Sireci, S. G. (2007). *A summary of the research on the effects of test accommodations: 2005-2006* (Technical Report 47). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- *Zurcher, R., & Bryant, D. P. (2001). The validity and comparability of entrance examination scores after accommodations are made for students with LD. *Journal of learning disabilities, 34*(5), 462-471.
- Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13*(1), 99-117.

收稿日期：2010.05.28

接受日期：2010.09.29

The Effects of Test Accommodation on Students with Disabilities: A Meta-analysis

Wan-Feng Chang

Teacher,

Kaohsiung Municipal Ming-Hua Junior High School

Wern-Ing Niew

Professor,

Dept. of Special Education,
National Kaohsiung Normal University

ABSTRACT

The purposes of this research were to use meta-analysis to examine the overall effects of 38 test accommodation studies conducted in 1999-2008 on the performance of students with and without disabilities, and to analyze whether there were differences on moderator variables of subject area, accommodation strategy, educational level, sample size, and disability status. The findings were as follows: first, test accommodation had a minimal effect on students without disabilities, and a small effect on students with disabilities. The results validated the interaction effect and differential boost hypotheses, whereby students with disabilities benefited more than students without disabilities. Second, students without disabilities tested on standard formats scored higher than students with disabilities accepting test accommodation, and the differences amounted to a medium effect. The results validated the hypothesis that test accommodation could create a level playing field for all test-takers. Third, the effect of test accommodation on other subject areas was higher than on math and reading. There was a higher effect from “extended time” strategy than “reading test items aloud”. Regarding educational level, the effects of test accommodation were ranked as follows, from high to low: junior high school, elementary, and senior high school levels. There was a higher effect on a smaller number of students with disabilities (i.e., below 100), and the effect decreased as the number of students increased. The effect of test accommodation on students with learning disabilities was higher than on students with other disabilities.

Keywords: test accommodation, meta-analysis, students with disabilities