國立臺灣師範大學教育心理與輔導學系
教育心理學報，2014，46 卷，2 期，271-288 頁
DOI：10.6251/BEP.20140516

電腦化適性職涯性向測驗編製研究*

鄭育文　　陳柏熹　　宋曜廷
國立臺灣師範大學
教育心理與輔導學系

陳信豪　　蕭孟莛
國立臺灣師範大學
心理與教育測驗研究發展中心

一個好的職涯測驗可讓學生瞭解自身之心理狀態，亦可作爲生涯規劃與抉擇的工具，而國內現有之性向測驗大多爲紙筆測驗，其題數較多且題目內容與真實職業情境差異較大，使其在施測及應用與解釋時有較大的限制。有鑑於此，本研究整合測驗、心理計量之適性化技術與多媒體技術，開創測驗內容、向度與題型，編製「電腦化適性職涯性向測驗—國中版」（CACAT-J），共包括八個分測驗，分別爲語文、數學、空間、邏輯推理、科學推理、觀察、美感和創意。此測驗可同時測量學生多面向之潛能，以作爲學生與教師在進行生涯輔導與科系媒合時之有效工具。本研究預試樣本人數爲 3163 人，正式樣本爲 5820 人。在 IRT 之樣本層次信度介於 .69~ .81；在效度方面，與國中基本學力測驗之相關係數介於 .28~ .67。綜合上述，本測驗具有良好之信度及效度。在常模建置方面，因屬電腦化測驗，可定期更新常模資料。透過測驗參照團體定期更新，以確保測驗分數的可比較性。CACAT-J 所提供之測驗結果，學生能瞭解其自身之優勢能力，將有助於學生探索與掌握自我，進而培養生涯決策與規劃能力。

關鍵詞：生涯發展、性向測驗、適性化測驗、職涯測驗

青少年的分流教育一直以來都是各國關注的教育政策，隨著教育理念與國家政策的不同，各國的分流教育時間亦有所差異。例如：德國可視爲採用早期分流的代表，該國在學生完成初等教育課程（修業年限多爲 4 年）後，便進入不同類型的中等教育學校接受教育，不同類型的學校各有不同的課程內涵、教學目標以及未來的教育方向（梁福鎮，2009）。採用晚期分流者，則以美國爲代表，至高等教育第一年時仍不設定就讀科系，可至大學二或三年級時，再進行決定主修的科目。而臺灣現行的教育體系，則是學童在完成九年的國民義務教育之後，透過基本學力測驗並

* 1. 本研究本文部分研究承蒙科技部補助研究經費，計畫編號(NSC 103-2911-I-003-301)、(NSC 102-2511-S-003-001-MY3)、(NSC101-2511-S-003-047-MY3)、（NSC 100-2631-S-003-001-）。
2. 本篇論文通訊作者：宋曜廷，通訊方式 sungtc@ntnu.edu.tw。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

藉由申請入學或考試分發的管道，進入普通高中、綜合高中或是高職（五專）就讀，屬於中期分流。根據教育部 100 學年度之統計，國中畢業學生進入高中（137084 人）及高職（131943 人）的學生比例爲 1.04：1（教育部，2011）。這意味著國中畢業學生有超過 13 萬的學生選擇就讀高職，然而現行高職學制爲 15 群、80 餘個類科，學生如何在眾多類科中進行選擇，一直是值得關注的議題。在早期和中期分流的國家，學生尚處於生涯發展的探索階段，而教育分流（streaming）制度卻可能迫使個體必須提前進行關鍵的生涯抉擇。國內外許多研究發現，青少年最常遭遇的困擾之一就是升學與就業的選擇（胡海國，1991；Friedman, 1991; Taveria, Silva, Rodriguez, & Maia, 1998）。生涯決策（career decision making）是一個複雜的任務，思考如何決策的過程往往導致個體產生焦慮和困惑的感覺，而青少年通常在生涯決策歷程的初步階段產生最多的困難（Gati, Saka, & Krausz, 2001）。

根據研究發現，歐美國家多透過生涯資訊系統來協助學生進行生涯決定（Gati & Amir, 2010）。一個完整的生涯資訊系統中，應具備三個不可或缺的成分：對學生相關（心理）狀態之訊息、外在職業環境之訊息與如何將前兩者訊息加以適配並進行選擇（Gati, Gadassi, & Shemesh, 2006）。在瞭解自我狀態的部分，包含潛能、興趣及工作價值觀等面向，其中又多以測驗作爲進行自我探索及了解自我之工具。Whitfield、Feller 與 Wood（2009）更指出生涯測驗可作爲在進行生涯規劃與抉擇的工具之一。性向測驗（aptitude test）被視爲測量個體潛在能力之工具，透過性向測驗的實施來推測或預測個體未來的學習、行爲或表現之結果。但因性向測驗所測量到的能力是個體目前的能力，而這樣的能力是受到遺傳與外在環境交互作用而成的，因此性向會隨著經驗的累積與環境有所變化（Linn & Gronlund, 2000）。由上述可知，性向測驗的內涵應隨著時代以及社會文化而有所修正或更新。

一、現有之職涯性向測驗及其使用限制

性向測驗在生涯諮商、學校輔導及人事心理上使用頻率頗高。因爲測驗結果可協助個體了解其優勢及劣勢，並可參考個人之興趣、動機以及其他條件，做成有關升學或就業之決策參考資料。在教育場域之中，更採用多元性向測驗，協助學生進行自我探索及瞭解。美國在進行生涯輔導工作時，常透過生涯資訊系統來協助學生或成人進行生涯探索或是生涯決擇。美國勞工局所建構的職業探索工具系統（O*NET, http://online.onetcenter.org/），提供了能力剖析量表（Ability Profiler, AP），採用紙筆方式評量，以測量九種主要的能力，作答時間需約兩個小時，試題量爲 229 題。在測驗結果產出，可透過紙筆記分或是電腦化讀卡。此外，美國大學測驗中心（American College Test, ACT）的 Discover 系統（http://www.act.org/discover/）則有別於過去採用性向測驗來瞭解能力，改採用自評（自我效能）的方式，以能力評估量表（Inventory of work-relevant abilities, IWRA）來取代常見之性向測驗，提供紙本及電腦化測驗，採用李克特點量表的作答方式，由受試者自行評估是否具備該項能力並與同年齡之學生進行比較，此量表共有十五個職業能力描述，如：社交、助人與銷售等能力。除了上述兩大系統，另一個常見的則爲互動性指引及資訊系統（System of Integrated Guidance and Information-Plus, SIGI-Plus），結合輔導和教育職業資料的電腦軟體，提供學生及成人進行教育發展計畫或職業發展規劃。此系統在了解能力方面開發了技能自評量表，將工作技能分爲六大類別，如：與人工作、用手操作和溝通等，每個類別有四到七句技能的描述，受試者衡量自己的能力選擇「擅長」或「不擅長」。

表 1 中爲美國常見之三大生涯資訊系統所發展與能力相關之測驗，其適用的對象多爲高中至成人階段，這或許跟美國爲晚期分流的教育制度有所關聯。此外，所有測驗的題目編排方式採固定題目，而 AP 則爲傳統的性向測驗，以測量潛在能力爲主，爲紙筆式測驗，且施測時間較長。另外兩個系統在進行能力評估時，其測量的內涵雖涵蓋不同的職業能力，但在作答形式上改採用自陳量表，受試者透過對自我的評估來進行作答，其作答時間較短。但是，自我效能測驗則是對個人信心程度的評估（Betz & Hackeet, 1981），並非反應個人真實之能力，因此採用自陳量表來評估學生能力是否恰當，則有待商榷。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

表 1　美國生涯資訊系統之能力測驗分析

系統名稱	測驗名稱	使用對象	測驗方式	題目	測量方式	作答時間（分鐘）	題量	測驗內涵
Discover	IWRA	5-9 年級 9 年級至成人	電腦化/紙本	固定	李克特量表	8-10	72/90	15 種職業能力
O*NET	AP	16 歲以上成人	紙本/實作	固定	選擇題	120	229	9 種能力
SIGI-Plus	技能自評量表	高中以上學生、成人	電腦化	固定	李克特量表	—	—	6 項工作技能

註：－表示未提供相關訊息。

反觀國內各種性向測驗，如表 2 所示。現有之性向測驗皆採紙本測驗、題型皆爲選擇題，測驗向度方面多相似。在編製年代方面，除了國中新編多元性向測驗爲 2011 年出版，多數測驗出版年代較爲久遠，且平均所需之測驗時間約爲一個半小時，其題量介於 296~496 題，測驗結果提供各個分測驗之分數以及側面圖。透過上述的國內外常用之性向測驗之介紹，可知目前國內使用的性向測驗雖種類不少，但不論是測驗本身或是實務運用上皆有其待改進之處，因此導致性向測驗在實務上運作之頻率及其效益不如預期，無法爲學生提供適切的協助（宋曜廷、田秀蘭、鄭育文，2012）。以下就現行性向測驗之問題及限制做一整理。

表 2　國內常用之性向測驗

測驗名稱（出版年代）	施測方式	作答方式	測驗向度	題量	作答時間（分鐘）	測驗結果
國中新編多元性向測驗（2011）	紙本	選擇題	語文推理、數學推理、圖形推理、機械推理、空間關係、中文詞語、英文詞語、知覺速度與確度	296	71	測驗分數、側面圖、性向組合分數
中學多元性向測驗—國中版（2003）	紙本	選擇題	語文推理、數學、科學推理、知覺速度與確度、空間關係、抽象推理、字詞釋義、文法修辭	496	80	測驗分數、側面圖
多向度性向測驗組合（2003）	紙本	選擇題	語文推理、機械推理、電腦能力、數學推理、抽象推理、電學知識、文句重組、資料核對	383	94	測驗分數、側面圖
區分性向測驗第五版（1999）	紙本	選擇題	語文推理、數學推理、抽象推理、知覺速度與確度、機械推理、空間關係、中文字詞、中文語法	420	90	測驗分數、側面圖
多因素性向測驗（1994）	紙本	選擇題	語文推理、數學推理、機械推理、空間關係、抽象推理、錯別字、文法與修辭、知覺速度與確度	439	60	測驗分數、側面圖

（一）職涯測驗包含之能力向度不符需求，測驗結果無法對應職業學群

國內高職職科可分爲 15 大類，近年來更新增不同科別，如：家政群的時尚模特兒科、藝術群的多媒體設計科等。現有的性向測驗所涵蓋之分測驗內涵並無法符合於目前教育環境之需求，而其測驗結果更難以與職業學群對應，造成教師在進行職涯輔導工作的不便。現有的性向測驗所含括之能力向度不足以對應目前的高職類群，以藝術類群爲例，此類群必須具備創造力、美感與觀

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

察等能力，但現有的測驗並未有相關能力（宋曜廷、田秀蘭、鄭育文，2012）。因此必須納入新的概念並加以整合，根據現代教育環境增加新要素。

（二）受限於紙筆式測驗導致測驗題型無法反應真實能力

由於過去皆採用紙筆測驗，測驗題型較受限制。例如：在空間測驗方面，主要希望能透過測驗測量個體之空間能力。以高職土木建築群爲例，其所需要之空間能力多展現於俯視圖或是多角度轉換空間的能力。但目前現行使用之性向測驗其空間題型多採用平面的方式呈現題目，卻乏立體圖、多角度轉換等不同的題型。然現今測驗技術及資訊科技已經優於從前，所以應利用新技術發展新題型或改善題目呈現方式及內容，以增進性向測驗之豐富性及效益。

（三）題數過多，施測時間過長

目前性向測驗的題數都相當多，施測時間都至少要花費 1 個半小時的時間來完成。受試者常因作答時間過長而感到疲憊，而在測驗過程後段容易出現精神不濟、隨意作答的情形，所以無法達到標準化施測標準，而影響測驗的結果，則測驗結果的可信度就會受到影響。若能透過電腦化適性測驗（computerized adaptive testing, CAT）的技術，讓受試只需較少的題數就能達到與傳統測驗相同的測量精準度（Sand, Water, & McBride, 1997）。

（四）常模建置年代久遠導致測驗結果解釋力下降

多數常使用之性向測驗，多爲一、二十年前所編製，使用測驗出版前所建置之常模來解釋現在學生的測驗結果，其效度是否恰當，則有待商榷。若能有一定的機制能在固定的時間更新常模，才能確保測驗結果的解釋力。

（五）測驗結果之取得需耗費較多的時間

多數性向測驗採用紙筆測驗，無論是施測或成績計算，皆由人工進行處理，需耗費較多時間等待測驗結果。

（六）測驗結果應用受限，群科建議指標性不足

多元性向測驗實施的目的之一，即是提供學校進路輔導，但是測驗結果往往缺乏相關資訊加以解釋與應用，亦即測驗結果往往無法和實際的情況相互符應。現有之性向測驗指導手冊，於解釋的部分多僅呈現各分測驗的百分比或折線圖，學生僅能從中得知各個分測驗之分數表現，無法獲得其他的訊息。因此若是要透過性向測驗結果作爲未來生涯抉擇的輔助工具其功效有限。

相對於美國相關測驗的發展，在國內現有之性向測驗則面臨了六大問題及限制。本研究將針對這些限制，試圖透過開發新型態及新技術的性向測驗解決相關問題，並根據國內高中及高中所需之能力作爲編製之架構。

二、電腦化適性測驗

電腦化適性測驗（CAT）的技術就是利用測驗計量技術根據受試者的答題反應，即時估計受試者當下的能力，並選出適合該受試者能力的試題作爲下一題進行施測，直到達到能力估計精準度或是測驗題數上限。受試者所接受的試題之難度皆相當符合其能力，所以受試者僅需施測較少的題目即能達到與傳統測驗相同的測量精準度（Sand, Water, & McBride, 1997）。CAT 的主要理論基礎是試題反應理論（item response theory, IRT），由於 IRT 的單向度假定，其可利用概似函數（likelihood function）來估計施測不同試題之受試者的潛在能力（Hambleton & Swaminathan, 1985）。在 IRT 架構下，接受不同試題的受試者能力可被放在同一量尺上相互比較，故受試者能力具備跨測驗的可比較性，此特點爲 CAT 的重要基石之一。

以下介紹幾個適性測驗施行的重要程序：測驗起始選題、能力估計、選題及測驗結束條件。

（一）測驗起始選題

於測驗開始之初，因並不確定受試者的能力，故通常選擇中等難度試題作爲初始選題。

NATIONAL CENTRAL LIBRARY ROC 國家圖書館

（二）能力估計

目前常用的能力估計方法有三種：最大概似法（maximum likelihood, ML）、貝氏最大後驗法（maximum a posteriori, MAP）及貝氏期望後驗法（expected a posteriori, EAP）（Embretson & Reise, 2000）。ML 雖具一致性、漸進的常態性與量尺不變性的特性，但需在題數夠多時才成立，且當受試者答題反應爲全對或全錯時，ML 無法進行估計（Hambleton & Swaminathan, 1985）。ML 雖較無迴歸性偏誤，但估計誤差較大，而 MAP 及 EAP 估計誤差較小，且不受答題反應全對或全錯之限制，但因受到能力先驗分佈之影響，會有迴歸性的誤差（Hambleton & Swaminathan, 1985）。Bock 與 Mislevy（1982）指出 EAP 因無須迭代，計算上相對簡單有效率。此外，雖然 EAP 與 MAP 皆有迴歸性偏誤，但 EAP 的迴歸性偏誤較小，故本研究採 EAP 進行能力估計。

（三）選題

依上述能力估計方法，根據受試者所有施測過之試題的答題狀況來估計受試者暫時的能力估計值，並選擇在此能力值下，能提供最大訊息量的試題作爲下道試題。在最大訊息法中，最常使用的訊息函數爲費雪訊息函數（Fisher information function, FI）。在 Rasch 模式下，訊息函數恰爲對某道試題的答對機率與答錯機率相乘，計算上相當便利。若在其他 IRT 模式下，訊息函數則略有不同。若以能力區間估計值代替點估計，即爲費雪區間訊息函數（Fisher interval information function, FII），此舉是因爲測驗初期能力估計的不確定性，因而考量 $\hat{\theta}$ 附近能力值來計算訊息函數，隨著測驗進行，區間縮小後，FII 就會變成 FI。除了上述的局部訊息函數外，KL 訊息函數（Kullback-Leibler information, KLI）則是一全域訊息函數，也可被用來做爲選題策略。由於過去模擬研究指出，在測驗初期時，受試者能力接近 0 時則 FI 的表現較佳，隨著題長增加，各種選題方式結果相似。且當測驗長度大於 10 題時，平均偏誤（bias）與均方根誤（root mean square of error, RMSE）在前述選題策略上的表現並沒有太大差異（Chen, Ankenmann, & Chang, 2000）。因此考量計算便利性與系統選題效能下，故本研究採用最簡單的 FI 作爲選題的訊息函數。

（四）測驗結束條件

常見的測驗終止方式有「固定題長」與「固定精準度」。固定題長，是施測前就已經決定受試者的作答題數。固定精準度，是當估計值達到特定標準後，停止施測。考慮到在學校場域中有課程安排之時間限制，因此僅能在有限題數進行施測，並能對於所有題目皆答對或是答錯之人進行能力估計，因此本研究採用貝氏期望後驗法（EAP）進行能力估計，以及採用費雪訊息函數（FI）作爲選題之依據，最終測驗結束條件採用固定題長。

有鑑於上述現有職涯性向測驗的限制，本研究的目的爲利用電腦化技術提升職涯性向測驗實施的效益，以協助解決上述問題。隨著時代進步，電腦化技術被廣泛應用在許多測驗有關的活動，如題目編寫、建立題庫、測驗編製、施測和評分、分析和印製成績報告等。比起傳統紙筆測驗，本研究所開發之電腦化適性職涯性向測驗之主要優勢在於：

1. 利用資訊和多媒體科技技術擴充性向測驗新架構與新題型的開發

有別於過去測驗向度，爲融入高職類群與高中所需之能力，在新開發之測驗架構方面，增加了觀察、創意以及美感等分測驗。透過多樣化之題型，如：空間測驗使用模型題型來測量空間定位能力、美感測驗使用拖曳題型來測量學生對於美的感受，透過電腦做爲媒介，在題型設計上可以更接近真實生活情境以及能更貼近所欲測量之能力。

2. 利用電腦化適性測驗技術提升施測效率

如果增加題型，學生在作答時的時間勢必增加。爲避免產生施測時間過長，反而影響作答動機的問題，本研究將透過 CAT 的技術，讓受試只需較少的題數就能達到與傳統測驗相同的測量精準度（Sand, Water, & McBride, 1997）。

3. 常模更新快速

將有效的結果資料納入常模的計算，即時更新常模，以確保測驗對個人分數的解釋力。

4. 快速獲得測驗結果

增進測驗實施與結果產生的便利性，節省施測人力與紙張成本，提升學校與學生使用測驗的意願。

5. 與現有高中職類群結合，提供學生類群適合度建議

NATIONAL CENTRAL LIBRARY ROC 國家圖書館

提供學生自身之個人性向組型與高職類群、高中建議，增加兩者媒合的效度，作爲生涯抉擇之參考工具。

綜上所述，本研究將結合電腦化、多媒體和計量技術編製一套電腦化適性職涯性向測驗，透過較少的題量以及較能反應學生各方面的題型測量學生多面向之潛能。透過本研究所開發之測驗，提供學生瞭解其自身之優勢能力，將有助於探索與掌握自我、培養生涯抉擇與規劃能力。

電腦化適性職涯性向測驗編製過程

本研究以不同於傳統測驗之題型、內容與架構之設計，建構一套完整且適用於國中學生之電腦化適性職涯性向測驗（Computerized Adaptive Career Aptitude Test-Junior, CACAT-J）。以下將CACAT-J 發展步驟分別敘述之。

一、測驗架構

研究者蒐集國內外相關文獻與多種性向測驗，並進行焦點團體與問卷調查後，確立本測驗架構、題型及命題原則。首先，本研究參酌過去相關研究和數種多元性向測驗進行內容分析，其分析面向含括測驗向度、題型（含內容）、信度、效度、施測方式等面向進行細項分析，探究相關測驗其內涵與現況，並且嘗試提出新觀點作爲後續發展測驗之參考與指標。第二階段，邀請高職 15 群科中心學校代表教師（爲高職各類科之任教超過十年之教師，在此定位爲專家教師）以及一般高中教師共計 17 位教師，填寫高職群科所需之能力問卷，填寫完畢後寄回本研究單位。第三階段，邀請上述之專家教師進行兩次的焦點團體，針對能力問卷內容進行討論，並取得各類能力定義，以及高中和高職各群科能力與各能力間之對應分析。爲透過內容分析、專家問卷以及焦點團體後所獲得之能力分析，並以此做爲本研究之測驗架構，如表 3 所示。

表 3　高中及高職各群科能力分析

高中及高職群科	能力							
	語文	數學	空間	邏輯推理	科學推理	觀察	美感	創意
機械群		V	V		V			
動力機械群	V		V	V	V			
電機電子群		V	V	V				
化工群		V		V		V		
土木建築群		V	V	V			V	
商業管理群	V	V		V				
外語群	V			V				
設計群			V			V	V	V
農業群				V		V		V
食品群		V		V				V
家政群	V		V				V	V
餐旅群	V					V	V	V
水產群				V		V		
海事群	V		V		V			
藝術群			V	V		V	V	V
普通高中	V	V		V				

註：V 爲該群科所需具備之能力。

本研究所開發之測驗共計八個分測驗，如表 4 所示，包括語文、數學、空間、邏輯推理、觀察、科學推理、美感與創意測驗。此外，本研究利用資訊和多媒體科技擴充測驗題型，提供更豐富化的施測內容以及測驗題型，讓題目呈現與真實生活情境更加貼近，以提升測驗效度。

表 4　電腦化適性職涯性向測驗之定義與題型

測驗名稱	定義	題型
語文	在不同情境下與人之間的溝通、互動之語言表達以及理解文本內容的能力。	選擇題
數學	在數字與數量概念的運用能力，以及與數量方面有關的推理能力，又稱「數感」能力。	選擇題、填充題
空間	對於物體在腦海中方向轉換如：翻轉、旋轉與折疊，以及對於空間、圖形中的線索和小空間距離感判斷的能力。	選擇題
邏輯推理	對於找出事物共同原理原則，及可將既有的原則應用到相似的情境，來協助問題解決的能力。	選擇題
觀察	對於覺察和辨別實物或現象在圖形與外形上的改變或細小差別，並快速擷取資訊的能力。	選擇題、熱點題
科學推理	在運用線索解決日常生活相關之情境所遭遇到之科學問題的能力。	選擇題
美感	在視覺感知的表達能力和判斷視覺形象的敏感度。透過色彩、線條與畫面等方式，並運用不同的作答方式展現出美感。	選擇題、拖曳題、配對題
創意	測在有限時間內針對某個主題，能夠提出不同的構想並且能夠改變思考方式，突破成規的能力。	多媒體

二、預試設計與預試資料分析

在測驗向度以及題型確認後，隨即召開命題人員說明會編製試題，在試題設計上，力求描述清晰以及符合所欲測量之構念。經過多次的各領域專家之內、外審查作業，確立試題品質。最後將審查通過之題目匯入題庫系統之中，以利進行後續正式施測。

（一）樣本

本測驗施測對象爲臺灣各公私立國民中學 8 年級之男女學生，抽樣方式係採分層隨機抽樣，先根據各縣市學校班級數，計算各縣市在全國所佔班級比率，以此比例制定各縣市的抽樣班級數目，爲達有效樣本數量，預計抽樣 51 間學校，共 102 個班級。將母群學校依據縣市區分後，則以亂數值隨機排列該縣市之學校，然後配合目標抽樣數取得抽樣學校，各地區抽樣數分別爲北部 19 間、中部 16 間、南部 15 間以及東部 1 間。施測時間爲 2010 年 4 月中旬開始，至 2010 年 6 月下旬結束，資料收集完畢後，即進行分析。其中北部 19 間、中部 16 間及南部 10 間，而東部學校因施測期間學校業務繁忙無法配合，所以未能進行施測，共計施測 45 間學校。總有效施測人數爲 3163 人，由表 5 可知，各地區學校分布表可知其分布情形爲北部地區樣本數爲 53.05%、中部地區樣本數爲 28.64%、南部地區樣本數爲 18.31%。由上述可知，預試樣本與母體分配情況大致符合。在性別部分的分布是男生 1635 人（51.69%）、女生 1528 人（48.31%）。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

表 5　各地區國中八年級學生受試者人數

地區	北部	中部	南部	總計
學校數	19	16	10	45
人數	1678	906	579	3163
人數比例（%）	53.05	28.64	18.31	100

（二）題本設計

在大型測驗中，常見的資料收集設計有平衡不完全區塊（balanced incomplete block design, BIB）與定錨不等組（non-equivalent groups anchor test, NEAT）。BIB 設計須考量題庫題本數、題庫區塊數、每一題本的試題區塊數、每一試題區塊在題本出現的次數等等基本條件限制，因此在實務上 BIB 題本配置的相關參數並不易取得，所以組卷的困難度也較高。在 NEAT 設計下，所有受試者都會作答同一組定錨試題，然而非定錨試題則僅有少部分的受試者作答，因此，定錨試題的參數有較高的估計精準度，但非定錨試題的估計精準度則較差。因此，在考量不同題本編排方式之優點及限制，以及受試者無法在有限時間內作答所有試題的情況下，本研究採用改良自 NEAT 的 NEVAT（non-equivalent groups with variable anchor test, Chen, Kuo, & Sung, 2011）的等化設計。NEVAT 有多組定錨試題，採用不同的試題區塊作爲定錨試題，能夠使各試題區塊內試題難度的估計精準度較爲一致。在 Chen、Kuo 與 Sung（2011）的模擬研究指出，雖然 NEVAT 的 RMSE 比 BIB 高，但 NEVAT 的 RMSE 比 NEAT 低。再者，NEVAT 的題本配置比 BIB 設計更爲簡單，因此該研究建議可採用 NEVAT，兼顧題本配置的便利性與試題參數估計精準度。

本研究將各分測驗之試題分成 34 個試題區塊，總計 34 個題本，各題本下皆含 3 個試題區塊，而成對區塊出現在每個題本中的次數爲 1 次，設計方式如下圖 1 所示：題本 1 包含區塊 1、區塊 2 與區塊 5，則其他題本中若有出現區塊 1，將不會再配對區塊 2 或區塊 5。各題本之題目均與其他題本爲相互之共同題。在題本與學生的配對上，本研究亦使各題本在男生與女生的施測比例上盡量相近。

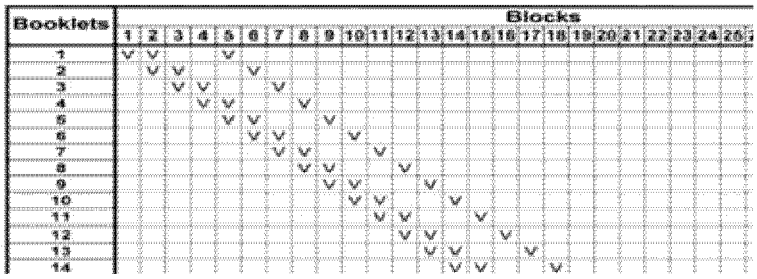


圖 1　預試題本試題排序之 NEVAT 設計示意圖

（三）計分方式

在計分方面，採二元計分及多元計分，如表 6 所示。二元計分方面，答對正確解答者即可得一分，反之零分。部分測驗因爲受到測驗內涵、題型等影響而採用多元計分法。以美感測驗爲例，由於美感是一種感知能力的強弱，並無唯一客觀的標準，因此美感測驗的拖曳題型中關於測量美感表達方面，採用專家（大學美術及藝術相關科系教授、高中美術教師及國中美術教師）認定之最佳答案以及完全錯誤答案之位置爲標準，若受試者之作答位置與專家認定最佳範圍相同者給予 2 分，反之則 0 分，若是範圍有所落差，則依受試者拖曳座標的範圍和專家認定的範圍進行比較給予部分計分。在觀察測驗的熱點題型方面，則紀錄原始作答反應數量（0~5 個），根據後續作答反應進行 PCM 分析取得階難度後進行轉換，最後取得該題目之分數，例如：反應數量 0~1 個爲一分、2~3 個爲兩分、4~5 個爲三分。創意測驗方面，其計分準則乃依據則採流暢力、變通力以及獨創力三個構念。在流暢力方面，流暢力旨在計算個體回答的總反應數，答案寫得愈多則得分愈高；變

NATIONAL CENTRAL LIBRARY ROC

通力旨在計算個體回答的不同類別變化數（本測驗共可區分爲六類別，分別爲含原始結構-橫豎、單部件-橫豎、單部件-撇點捺挑勾折、多部件上下結構、多部件左右結構、多部件內外結構），寫出的答案分屬於愈多類別則得分愈高；獨創力旨在計算個體回答於常模中的分配比率，分配比例與獨創力分數的計算，即個體每寫出一個常模當中反應頻率比例低於 2%的合法答案得 2 分，每寫出一個常模當中反應頻率比例介於 2%~10%的合法答案得 1 分，每寫出一個常模當中反應頻率比例大於 10%的合法答案則得 0 分，整體而言，寫出的答案於常模分配中出現的比例愈低則得分愈高。

（四）施測方式

各題本包含 8 個分測驗，測驗出現順序爲固定設計，需兩堂課時間的作答時間，第一堂課依序爲語文、美感、數學、科學推理；第二堂課依序爲空間、創意、觀察、邏輯推理，題量分配及計分方式，如表 6 所示。

表 6　各分測驗之時間限制、題量及計分方式

時間安排	分測驗	時間限制	題量	計分方式
第一部分	語文	無	9	二元
	美感	無	9	多元、二元
	數學	單題限時	9	二元
	科學推理	無	9	二元
第二部分	空間	無	9	二元
	創意	單題限時	1	多元
	觀察	部分單題限時	9	多元、二元
	邏輯推理	無	13	二元

（五）試題分析

本研究採用 Con Quest 電腦軟體進行試題分析。根據 Masters（1982）的部分計分模式（PCM）進行分析，而二元計分可以視爲部分計分之特例，公式（1）如下：

$$P_{ix}(\theta)=\frac{\exp[\sum_{j=0}^{x}(\theta-\delta_{ij})]}{\sum_{r=0}^{m_i}[\exp\sum_{j=0}^{r}(\theta-\delta_{ij})]} \quad (1)$$

其中，m_i 表示第 i 題的最高得分，δ_{ij} 表示第 i 題中第 j 個得分的階難度，則表示受試者在第 i 題得到 x 分的機率。判斷的試題配適度有兩種指標，分別爲非加權均方誤（unweighted mean square error, Outfit MNSQ）以及加權均方誤（weighted mean square error, Infit MNSQ）。兩者的相同處在於皆爲自由度爲 1 的卡方統計量，數值介於 0 與無限大之間，期望值爲 1.0。相異處在於非加權均方誤較容易受到能力極端者之不合理作答反應，導致較劇烈的數值變化；相反的，加權均方誤則較不受極端值影響（Linacre & Wright, 1994）。本研究將以加權均方誤爲主要指標，當某試題之配適度超出 0.8～1.2 區間之外，則視爲不適配試題（misfit items）（Linacre & Wright, 1994）。此外，除了適配度檢定之外，亦參考古典測驗理論（classical test theory, CTT）之鑑別度，挑出鑑別度不佳之試題（小於 .3），進行試題內容分析。本研究將以 IRT 適配度爲主，而 CTT 鑑別度爲輔，並透過領域專家進行試題是否合保留或刪除之綜合性評估。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

三、正式 CAT 測驗與資料分析

（一）樣本

從表 7 可得知，不同地區及性別之人數分佈情形。因爲採各校自主登記，導致有部分地區之樣本數較少。總人數爲 5820 人，在性別部分，男生爲 2949 人（50.67%）、女生爲 2871 人（49.33%）。

表 7　正式施測樣本之描述性統計

年級	地區	人數	比例（%）	性別	人數	比例（%）
八年級＋九年級	北部	4399	75.58			
	中部	274	4.71	男	2949	50.67
	南部	313	5.38			
	東部	834	14.33	女	2871	49.33
	總計	5820	100		5820	100

（二）模式選取及題本設計

根據預試資料分析後所得之試題參數，進行 CACAT-J 測驗設計。每一個分測驗（除創意測驗以外）之參數估計與選題均採用單向度模式，計分方式採用部分計分模式（parial credit model, PCM）。此外，試題訊息函數公式（2）爲：

$$Ii(\theta)=\frac{P'^{2}_{i}}{PiQi} \quad (2)$$

P_i' 爲 P_i 對 θ 的一階微分，Q_i 則爲 $1\text{-}P_i$。在終止測驗方面，採用固定題長作爲終止測驗的條件，除了邏輯推理測驗作答題數爲 13 題和創意測驗爲 1 題，其餘各分測驗的作答題數皆爲 9 題，共計作答數量爲 68 題。

（三）信度檢驗

在古典測驗理論中，觀察分數變異（σ_x^2）由真實分數變異（σ_t^2）與誤差變異所組成（σ_e^2）。而信度的定義爲真實分數變異除以觀察分數變異數之比率，如公式（3）所示。在 IRT 中，也有類似傳統信度的概念。本研究採 bilog-mg 軟體（Du Toit, 2003）計算信度的方式進行信度的估算。其概念如同傳統信度計算真實分數變異數佔觀察分數變異數之比率。然而當採用不同能力估計時，能力估計值的變異數則代表觀察分數的變異數或是真實分數的變異數。對 EAP 而言，EAP 估計值變異數即爲真實分數變異數，所以其信度計算方式爲 EAP 估計值變異數佔 EAP 估計值變異數加上誤差變異數之比率，見公式（4）。

$$\rho_{xx}=\frac{\sigma_t^2}{\sigma_x^2}=\frac{\sigma_x^2-\sigma_e^2}{\sigma_x^2} \quad (3)$$

$$\rho_{\hat{\theta}\hat{\theta}}=\frac{s_{\hat{\theta}}^2}{s_{\hat{\theta}}^2+\bar{s}_e^2} \quad (4)$$

（四）效度檢驗

因爲現有之性向測驗因爲編製年代久遠、採用固定題目以及測驗架構上有別於本研究所發展之電腦化適性職涯性向測驗，有鑑於此，並未採用與現有之性向測驗進行效標關聯效度之檢驗，而採用基本學力測驗進行效度檢驗。另以美感測驗及創意測驗分數與學業成績，進行效度研究。此外，在創意測驗方面，採用陳學志等人（2009）所編製之《有限解答之擴散性思考測驗─「里的刪減」作業》及吳靜吉等人（1998）《新編創造思考測驗─「人的圖形」作業》爲效標，求本創意測驗分數與上述之測驗分數之相關。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

研究結果

一、CACAT-J 試題難度及受試者能力分析

各分測驗試題根據 IRT 分析以後，採用邏輯思（logit）爲量尺單位，其平均難度皆爲 0，標準差介於 0.61~1.53，試題難度介於-2.98~3.23。在人的能力分佈上，平均能力介於 0.16~1.27，標準差介於 0.84~1.31，如表 8 所示。

表 8　各分測驗 IRT 難度範圍以及受試者能力分佈

試題難度	語文	數學	空間	邏輯推理	觀察	科學推理	美感
平均值	0.00	0.00	0.00	0.00	0.00	0.00	0.00
標準差	1.08	1.53	0.88	0.99	1.24	.61	1.09
最大值	2.42	2.71	1.47	3.23	2.04	1.42	2.51
最小值	-2.05	-2.58	-2.98	-2.38	-2.25	-1.42	-2.10
IRT 能力值							
平均值	0.81	0.75	0.16	0.37	1.27	0.31	0.93
標準差	1.20	1.31	1.18	0.84	1.08	0.71	0.92

本測驗所編製的創意測驗則根據流暢力、變通力及獨創力等三個構念分別計分後再進行加總。三個構念其平均值介於 2.64~10.82、標準差介於 0.78~4.21 以及最大值之數量介於 3~25。在總分方面，平均值爲 16.57、標準差爲 5.20 及最大值之數量爲 32，如表 9 所示。

表 9　創意測驗之試題分析

	流暢力	變通力	獨創力	總分
平均值	10.82	3.11	2.64	16.57
標準差	4.21	0.85	0.78	5.20
最大值	25	6	3	32

由表 10 中可知，創意測驗的三個構念間之相關介於 .43~ .56，達顯著之中度相關。三個構念和總分之間的相關介於 .63~ .98，達顯著之中高相關。

表 10　創意測驗之相關分析

	流暢力	變通力	獨創性	總分
流暢力	—			
變通力	.56**			
獨創力	.51**	.43**	—	
總分	.98**	.68**	.63**	—

** $p < .01$

二、CACAT-J 信度檢驗

本測驗正式樣本爲進行適性測驗之八年級學生 5158 人，透過各分測驗之所有受試者的期望估計標準誤與能力總變異，求得 IRT 信度指標。在各分測驗試題數爲 9~13 題的情況下，信度介於 .69~ .81 之間，創意測驗因並非適性選題，因此未能提供相關信度係數，詳如表 11 所示。在平均測量標準誤方面，美感、觀察和科學測驗爲 0.15，語文、數學和空間測驗爲 0.36，邏輯測驗爲 0.35。

表 11　各分測驗之 IRT 樣本層次信度係數（$N = 5158$）

各分測驗	實徵資料		模擬資料
	適性選題施測題數	信度	隨機選題之題數
美感	9	.80	25
語文	9	.75	15
空間	9	.74	15
邏輯推理	13	.69	20
數學	9	.78	15
科學推理	9	.69	15
觀察	9	.81	20
總題量	67		125

此外，透過模擬資料分析，若透過隨機選題達到相同之信度，則各分測驗需要作答的題數則爲 15~25 題，總題量達 125 題，約爲適性選題的兩倍題量。由上述可知，CAT 選題約僅需隨機選題數目的一半即可達到類似的信度，換句話說，適性測驗的進行可以有效率地減少傳統測驗的題目量。

三、CACAT-J 效度檢驗

（一）試題分析

透過 IRT 與 CTT 兩種方式進行試題分析，建立正式試題題庫之難度參數。本研究以加權均方誤（Infit MNSQ）爲主要指標，挑出不適配之試題，以及參考 CTT 之鑑別度，挑出鑑別度不佳之試題，進行試題內容分析。在計算 CTT 鑑別度時，會採用每位受試者的原始總分與每道試題的得分進行相關係數的計算，然而由於採用 NEVAT 題本配置，各題本的難度可能不一，因而導致 CTT 的鑑別度可能失真。本研究以 IRT 適配度爲主，而 CTT 鑑別度爲輔，並透過領域專家進行試題是否合保留或刪除之綜合性評估。各分測驗之試題納入正式題庫之數量，如表 12 所示。創意測驗因爲在正式施測時每位學生僅做一題，並不會透過適性選題機制來進行挑選題目，因此未進行 IRT 相關分析。

NATIONAL CENTRAL LIBRARY ROC 國家圖書館

表 12　各分測驗之題量分配

分測驗	總預試題數	試題數		正式題庫試題數
		IRT 適配度不佳	CTT 鑑別度不佳	
語文	102	1	0	101
數學	102	5	6	97
空間	102	3	2	99
邏輯推理	146	0	13	133
科學推理	93	0	12	81
美感	102	4	2	96
觀察	129	13	20	110

（二）與學業成績之相關

1. 國中基本學力測驗成績

本測驗又以國中基本學力測驗（以下簡稱基測）成績爲效標。在效標樣本方面，因多元入學管道，採用正式樣本中之九年級學生且參加基測之 483 人爲樣本進行資料分析。由表 13 可知，其相關係數介於 .28~ .67，介於低度至中度相關範圍。若從名稱相近之基測成績（包含國文科、數學科、自然科等）來看與本分測驗之相關，兩者間所得之相關係數大多普遍高於性質不同之分測驗所得之相關，如科學推理測驗和基本學力測驗中的自然科之相關（$r = .42$）高於其他分測驗。

表 13　本測驗與基測成績之相關（$N = 483$）

分測驗	99 年度基本學力測驗				
	國文科	英文科	數學科	社會科	自然科
語文	*.57***	.44***	.46***	.53***	.50***
數學	.56***	.58***	*.67***	.61***	.63***
空間	.42***	.37***	.48***	.44***	.47***
邏輯推理	.35***	.33***	.32***	.36***	.35***
科學推理	.38***	.32***	.40***	.38***	*.42***
觀察	.34***	.35***	.33***	.31***	.28***
美感	.50***	.43***	.41***	.46***	.44***
創意	.41***	.37***	.39***	.36***	.42***

註：斜體之數字表示基本學力測驗和本測驗其名稱或內容有相關性，** $p < .01$

2. 美感、創意測驗與在校成績

在美感和創意測驗方面，透過立意抽樣選取新北市及北市四間學校，採八年級共 114 名學生作爲效標樣本。從相關係數中可得知，美感和創意測驗和美術在校成績之相關分別爲 .27 和 .25，達顯著水準。

（三）創意測驗與其他性向測驗之效度分析

採前述創意測驗之效標樣本，以《新編創思考測驗─「人的圖形」作業》（吳靜吉等人，1998）與《有限解答之擴散性思考測驗─「里的刪減」作業》（陳學志等人，2009）等兩份測驗爲效標，進行效標關聯效度分析。如表 14 所示，整體而言，本測驗與現有之創意測驗其相關介於 .39~ .57。

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

表 14　創意測驗與效標測驗相關係數（$N = 114$）

	1	2	3
1. 本測驗之創意測驗	—		
2. 新編創造思考測驗	.39**	—	
3. 有限解答之擴散性思考測驗	.57**	.60**	—

** $p < .01$

四、常模建置

根據試題反應理論（IRT）所獲得之能力，轉換爲平均數 80、標準差 14、總分 125 的量尺分數。依據量尺分數轉換爲百分等級，並將正式施測之樣本按北部、中部、南部及東部各地區學生比例抽取，建置本測驗之全國常模及性別常模。

五、測驗結果之高中及高職群科適合度建議

透過專家教師挑選其所任教之群科所需具備之重要能力，並依據教育部 98 年課程綱要，以各類群之學科學分之比重，決定重要能力之加權比例。以設計群爲例，美感、創意、空間和觀察能力爲專家教師所挑選之該群科所需具備之能力，並根據教育部 98 年課程綱要學分之比重，依序之權重爲 37.78%、26.66%、20.00%、15.56%。根據八個分測驗之結果進行標準化後取得 Z 值，根據上述之方式，建置出高職 15 群科以及高中適合度算則，如公式（5）採用設計群爲例。

設計群 =（空間 Z 值* 0.2+觀察 Z 值* 0.1556 +美感 Z 值* 0.3778 +創意 Z 值* 0.2666）　（5）

本研究透過立意抽樣，選取高級職業學校中 15 群科中技能專業具代表之科別的高二及高三學生，進行本研究所編製之測驗（CACAT-J）。研究結果發現，就讀該群科學生之群科其適合度推薦皆高於非就讀該群科學生。以設計群爲例，就讀設計群之高職學生中有 31.08%被推薦就讀設計群，而非設計群之學生爲 18.99%被推薦就讀設計群。在經過檢定後發現，有超過一半之群科在就讀該群科學生之適合度與非該群科學生之比例有顯著性差異，由此可知，透過本測驗之測驗結果，在推薦高職群科適合度有不錯的效果。

討論及建議

一、CACAT-J 測驗編製

本研究之目的在於編製電腦化適性職涯性向測驗（CACAT-J），以期往後供學校輔導單位使用。本研究所發展之 CACAT-J 測驗共有八個分測驗，分別爲語文、數學、空間、邏輯推理、觀察、科學推理、美感與創意測驗。該測驗適用對象爲國中二至三年級學生，作答時間爲兩節課。從試題品質來看，本測驗所涵蓋的試題範圍相當廣泛，難度介於-2.98~3.23，表示不同能力之受試者，皆有適合其能力之試題可以進行作答。本測驗在 IRT 之樣本層次信度介於 .69~ .81；在效度方面，與國中基本學力測驗之相關系數介於 .28~ .67；美感、創意測驗與美術成績之相關介於中低相關；而創意測驗與其他效標測驗間之關係介於 .39~ .57，表示兩者之間具有中低相關。大致上，該性向測驗具有不錯的信度及效度。

整體而言，本測驗有以下幾點特色。第一，此測驗架構不僅參酌過去相關研究和數種多因素性向測驗內容外，並針對現今所需之能力，納入新的測驗向度，擴充測驗內涵，以便能真實測量

出學生之能力。其次，透過電腦化適性測驗的計量技術，學生作答之題量可以減少 50%，則可降低學生施測之疲憊感，藉此提升作答動機。第三，全國常模定期更新，可讓受試者有最適當之參照團體可以進行比較，以了解自身之能力。第四，透過電腦化測驗可以快速取得測驗結果，不需進行等待。最後，根據測驗結果，提供學生高職及高中類群之建議，作爲生涯抉擇之參考工具。

二、CACAT-J 測驗之功能及實務建議

首先，對學校輔導教師而言，透過測驗工具爲生涯輔導工作增加其專業性之展現。本測驗提供豐富之測驗結果，包括測驗成績以及提供高中職群科適合度，有助於教師在進行生涯輔導時，擁有更具體及明確之內容，以作爲課堂或諮商時之應用及參考。其次，對學生而言，透過測驗結果，能夠瞭解自我之潛能，有助於自我探索，進而提升對於自身優勢能力的瞭解。在進行生涯抉擇時，除了既有之學業成績以外，能有更多面向之訊息進行決策， 期許能降低生涯決定之困難。最後，對家長而言，除了平日對於孩子的觀察，亦可透過測驗結果，多面向瞭解其孩子的潛能。家長透過測驗結果與其他相關之資料，與孩子一同討論未來生涯方向，協助孩子適性發展。

參考文獻

吳靜吉、陳甫彥、郭俊賢、林偉文、劉士豪、陳玉樺（1998）：**新編創造思考測驗指導及研究手冊**。台北：教育部訓委會。[Wu, C. C., Chen, F. Y., Kuo, C. H., Lin, W. W., Liu, H. H., & Chen, Y. H. (1998). *Guidance manual of the new tests of creative thinking*. Taipei, Taiwan: Student Affairs Committee of Ministry of Education.]

宋曜廷、田秀蘭、鄭育文（2012）：國中與高中職階段生涯測驗使用現況之分析研究。**教育心理學報，43**（4），875-898。[Sung, Y. T., Tian, H. L., & Cheng, Y. W. (2012). Analysis of career development tests in secondary, high, and vocational high school. *Bulletin of Educational Psychology, 43*(4), 875-898.]

胡海國(1991)：**發展心理學**。台北：桂冠。[Hu, H. K. (1991). *Developmental Psychology*. Taipei, Taiwan: Gui-Guan.]

教育部(2011)：**教育統計 100 年版**。台北：教育部。[Ministry of Education (2011). *Education Statistics* (2011 ed). Taipei, Taiwan: Ministry of Education.]

梁福鎮(2009)：德國中等教育的現況與改革措施。**教育資料集刊，42**，273-296。 [Liang, F. C. (2009). The current situation and reformative measures of German secondary education. *Bulletin of National Institute of Education Resource and Research, 42*, 273-296.]

陳學志、洪素蘋、許禕芳、邱皓政、關秉寅、詹志禹（2009）：擴散性思考與聚斂性思考的交會：有限解答之擴散性思考測驗之編製與信效度檢驗。**教育科學研究期刊，54**（4），29-61。[Chen, H. C., Hung, S. P., Hsu, Y. F., Chiou, H. C., Kuan, P. Y., & Chan, C. Y. (2009). The connection of divergent thinking and convergent thinking: Developing of constrained

國家圖書館 NATIONAL CENTRAL LIBRARY ROC

divergent thinking test and examining reliability and validity. *Journal of Research in Education Sciences, 54*(4), 29-61.]

Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology, 28*, 399-410.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444.

Chen, S. Y., Ankenmann, R. D., & Chang, H. H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement, 24*, 241-255.

Chen, P. H., Kuo, J. W., & Sung, Y. T. (2011). *Influence of pre-test design on the precision of the parameters estimation in the multidimensional items bank*. Paper presented at the IMPS 2011, the 76rd Annual Meeting of the Psychometric Society, Hong Kong.

Du Toit, M. (2003). *IRT from SSI: Bilog-mg, multilog, parscale, testfact.* Lincolnwood, IL：Scientific software international.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Friedmen, A. I. (1991). Areas of concern and sources of advice for Israeli adolescents. *Adolescents, 26*, 967-976.

Gati, I., & Amir, T. (2010). Applying a systemic procedure to locate career decision-making difficulties. *Career Development Quarterly, 58*(4), 301-320.

Gati, I., Gadassi, R., & Shemesh, N. (2006). The predictive validity of a computer-assisted career decision-making system: A six-year follow-up. *Journal of Vocational Behavior, 68*, 205-219.

Gati, I., Saka, N., & Krausz, M. (2001). "Should I use a computer-assisted career guidance system?" It depends on where your career decision-making difficulties lie. *British Journal of Guidance and Counseling, 29*, 301-321.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.

Linacre, J. M., & Wright, B. D. (1994). Dichotomous Mean-square Infit and Outfit Chi-square fit statistics. In B. D. Wright & J. M. Linacre (Eds.), *Rasch measurement transactions* (Vol. 8, Part 2, p. 360). Chicago, IL: MEAS Press.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.).Upper Saddle River, NJ: Prentice-Hall.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

NATIONAL CENTRAL LIBRARY ROC

Sand, W. A., Water, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: from inquiry to operation.* Washington, DC: American Psychological Association.

Tavereira, M. D. C., Silva, M. C., Rodriguez, M. L., & Maia, J. (1998). Individual characteristics and career exploration in adolescents. *British Journal of Counseling and Guidance, 26*, 89-104.

Whitfield, E. A., Feller, R. W., & Wood, C. (2009). *A counselor's guide to career assessment instruments* (5th ed.). Broken Arrow, OK: National Career Development Association.

收 稿 日 期：2013 年 07 月 11 日
一稿修訂日期：2014 年 03 月 24 日
二稿修訂日期：2014 年 05 月 16 日
接受刊登日期：2014 年 05 月 16 日

NATIONAL CENTRAL LIBRARY ROC 國家圖書館

Bulletin of Educational Psychology, 2014, 46(2), 271-288
National Taiwan Normal University, Taipei, Taiwan, R.O.C.

Construction of the Computerized Adaptive Career Aptitude Test for Junior High School Students

Yu-Wen Cheng Po-Hsi Chen Yao-Ting Sung
Department of Educational Psychology and Counseling
National Taiwan Normal University

Shin-Hau Chen Meng-Ting Hsiao
Research Center for Psychology and Educational Testing
National Taiwan Normal University

A qualified career test could play an important role in providing information for students' self-understanding and career decision-making. However, there are limitations and problems within the existing career aptitude tests in Taiwan. This study integrates psychological testing, adaptive techniques of psychometrics, and multimedia technology to establish a computerized adaptive career aptitude test (CACAT - J). The CACAT - J consists of brand new subject matters, dimensions and test items. Not only is the test more relevant with increased prediction power, but it also provides beneficial career counseling for students and teachers. The CACAT - J consisted of eight sub-tests: Verbal, Numerical, Spatial, Logical Reasoning, Scientific Reasoning, Observation, Aesthetics, and Creativity. There were 5820 participants in the present study. Reliability for the findings in this study ranged from .69 to .81. In terms of validity, the aptitude test scores were significantly correlated with junior high school students' basic competence tests and ranged from .28 to .67. Through this test, students were able to understand their aptitudes on different skills that would enhance their career exploration and their capacity for effective career decision-making and planning.

KEY WORDS: adaptive test, aptitude test, career development, career test

NATIONAL CENTRAL LIBRARY ROC