

電腦化測驗在我國大學的應用現況及在國家考試的發展潛力

趙秀怡*

摘要

隨著數位科技發展與高等教育治理模式轉變，評量方式逐步由傳統紙筆測驗邁向電腦化與資料化。本文以我國大學電腦化測驗實務為基礎，整理其應用現況並分析其在國家考試中的發展潛力。研究發現，大學電腦化評量大致可區分為三種類型：電腦化問卷與能力評估測驗、題庫導向與能力檢測型測驗，以及電腦適性測驗。其發展歷程呈現由施測媒介數位化，逐步邁向題庫管理與測量模型導向之演進趨勢。進一步分析顯示，大學場域在施測流程標準化、題庫管理與使用者支持等面向已累積豐富經驗，為電腦化測驗制度化運作奠定基礎。在國家考試方面，我國已進入電腦化施測導入與擴展階段，惟其發展脈絡與大學場域不盡相同，未來發展重點在於制度深化與風險管理，包括固定題本電腦化施測之穩定運作、題庫管理能力之強化，以及多梯次施測與適性測驗之審慎評估。本文建議採取循序漸進之推動策略，未來之推動或可同時包含題庫導向與臨時命題等不同模式，由低風險之媒介轉換逐步邁向高精準度之測量模式，並在測驗透明性與題庫管理之間取得平衡，以兼顧測量效率、品質與公平性。

關鍵詞：國家考試、測驗等化、電腦化測驗、電腦適性測驗、題庫管理

*東吳大學心理學系副教授

Current Applications of Computer-Based Testing in Taiwanese Universities and Its Development Potential in National Examinations

Hsiu-Yi Chao*

Abstract

With the advancement of digital technology and transformations in higher education governance, assessment practices have gradually shifted from traditional paper-and-pencil testing toward computer-based and data-driven approaches. Drawing on practical experiences of computer-based testing in Taiwanese universities, this study reviews current applications and examines their developmental potential within the national examination system. The findings indicate that computer-based assessments in higher education can be broadly categorized into three types: computerized questionnaires and assessment instruments, item bank-based competency tests, and computerized adaptive testing. The developmental trajectory reflects a progression from digitalization of test administration to more advanced approaches emphasizing item bank management and measurement model-based design. Further analysis suggests that universities have accumulated substantial experience in areas such as standardized administration procedures, item bank management, and user support, thereby establishing a foundation for the institutionalization of computer-based testing. In the context of national examinations, Taiwan has entered a stage of implementation and expansion of computer-based testing. Although its developmental trajectory differs from that of higher education institutions, future development should focus on institutional refinement and risk management, including the stable operation of fixed-form computer-based testing, enhancement of item bank management capabilities, and careful evaluation of multi-session testing and adaptive testing. This study recommends a gradual implementation strategy. Future development may incorporate both item bank-based and ad hoc test development approaches, progressing from low-risk modality transformation to high-precision measurement approaches, while maintaining a balance between test transparency and item bank management, in order to ensure efficiency, quality, and fairness in assessment.

Keywords: computer-based testing, computerized adaptive testing, item bank management, national examinations, test equating

* Associate Professor, Department of Psychology, Soochow University

壹、前言

近年來，隨著數位科技快速發展與高等教育治理模式的轉變，評量方式亦逐步由傳統紙筆測驗邁向電腦化與資料化。電腦化測驗不僅改變了測驗施測的媒介，也重新形塑評量資料的蒐集、分析與回饋機制，成為高等教育與大型考試制度中不可忽視的重要發展趨勢（American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; International Test Commission, 2006）。在國際間，許多標準化測驗與能力評量已逐步導入電腦化施測，強調即時資料處理、測驗安全性與施測效率（International Test Commission, 2006）；在我國，隨著學習成效評估機制與校務研究（institutional research, IR）的推動，大學端亦逐漸累積以電腦化方式進行評量與資料分析的實務經驗，逐漸呈現由單次測驗走向長期數據治理之發展趨勢（林靜慧等，2017b）。

具體而言，電腦化測驗乃指以電腦為施測媒介，並結合數位系統進行試題呈現、作答紀錄、計分與資料管理之評量形式。其範圍並不限於單一測驗型態，而是涵蓋不同技術成熟度與測量策略之多元實踐。從以電腦施測的學習成效問卷與能力評估系統、題庫導向線上測驗，到結合心理計量模型的適性測驗，皆可視為電腦化評量在不同發展階段的實踐（International Test Commission, 2006; van der Linden & Glas, 2010; Wainer, 2000）。以我國大學為例，部分學校已透過學習成效評估系統與校務研究資料庫，長期蒐集學生學習經驗與能力發展資料，作為課程改善與校務決策的重要依據（林靜慧、陳俊宏，2017；林靜慧等，2017a, 2017b）；另一方面，也有研究團隊投入電腦適性測驗（computerized adaptive testing, CAT）之建置與實證研究，顯示大學端在電腦化測驗之技術與制度面均已具備一定基礎（Huang et al., 2022）。這些多元實務經驗，提供了觀察我國電腦化測驗發展脈絡的重要切入點。

然而，相較於高等教育場域較具彈性的評量環境，國家考試制度在轉型過程中面臨更為複雜的制度條件與社會期待。國家考試以高風險、大規模與高度標準化為主要特性，在導入新型評量方式時需更加審慎。因此，隨著考試數位轉型議題逐漸受到重視，如何在維持測驗公平與信度的前提下，適度引入電腦化測驗，以提升施測效率、優化考生體驗並強化評量功能，已成為值得探討的議題。

(AERA, APA, & NCME, 2014; International Test Commission, 2006)。而大學場域所累積的電腦化測驗實務，可作為觀察與思考國家考試未來發展的重要參考。

基於上述背景，本文旨在整理電腦化測驗於我國大學的應用現況，從不同型態的電腦化評量實務出發，分析其在測驗設計、資料運用與制度推動上的經驗與挑戰，並進一步探討其在國家考試中的發展潛力與可能路徑。期望透過結合高等教育實務與考選制度需求的觀點，提供對我國電腦化測驗發展之整體性觀察與建議。

貳、我國大學電腦化測驗的應用型態

隨著資訊科技與數位學習環境的成熟，我國高等教育場域之評量實務亦逐步朝向電腦化發展。相較於傳統紙筆施測模式，電腦化測驗在施測流程管理、題庫維護、測驗結果即時回饋與跨年度資料整合等面向展現更高之彈性與效率 (International Test Commission, 2006)。整體而言，大學端電腦化評量並非一次性技術導入，而是在教學需求、行政管理與學習成效分析等多重因素交互作用下逐步形成，其發展路徑大致由電腦化問卷與固定題本測驗，進展至題庫導向測驗與結合心理計量模型之適性測驗 (van der Linden & Glas, 2010; Wainer, 2000)，呈現由基礎應用邁向進階測量的演進趨勢。

在實務層面，我國部分大學已透過電腦化系統長期蒐集與分析學生學習成效與能力發展資料，並將其納入校務治理與課程改進機制。以國立臺灣師範大學推動之大學生基本素養電腦化測驗為例，其結合情境題組設計與試題反應理論 (item response theory, IRT) 分析，建立標準化施測與測驗等化機制 (陳柏熹, 2014; 陳柏熹等, 2015)。相關研究亦指出，透過校務研究資料庫與學習成效評估系統之整合，可系統性蒐集學生能力指標與學習歷程資料，作為課程品質檢核與教育決策參考之依據 (林靜慧、陳俊宏, 2017; 林靜慧等, 2017a, 2017b)。在學生發展輔導面向，部分大學亦建置結合個體資訊與環境資訊之生涯資訊系統，以資訊科技支援學生生涯規劃與決策歷程 (王思峰等, 2014)。上述多元實務型態顯示，大學場域已逐步將電腦化評量視為支持學習品質提升與制度運作之重要工具。

綜觀上述發展，我國大學電腦化測驗之應用可概略區分為三種型態：其一為電腦化問卷與能力評估測驗，其二為題庫導向與能力檢測型電腦化測驗，其三則

為結合心理計量模型之電腦適性測驗與進階電腦化測量。不同型態在技術複雜度與測量精緻度上雖有差異，但皆反映高等教育評量逐步朝向數位化與資料導向發展之趨勢。以下分別就各類型之發展脈絡與應用實例加以說明。

一、電腦化問卷與能力評估測驗

在我國大學推動電腦化測驗的歷程中，以電腦化問卷與能力評估測驗為代表的基礎型評量工具，扮演重要的起點角色。此類測驗多以線上平台進行施測，內容涵蓋學生學習經驗、通識能力、自我評估量表及基礎能力檢測等面向，具有施測彈性高、行政成本低與資料蒐集效率佳等特性（International Test Commission, 2006）。相較於傳統紙本問卷，電腦化施測不僅縮短資料整理時間，也使評量結果得以即時回饋，逐步改變大學評量活動的運作模式。

在學習成效評估面向，國立臺灣師範大學建置之「電腦化大學生基本素養測驗」即為代表性實例。該測驗結合情境題組設計與試題反應理論分析，透過電腦化施測與等化機制建立跨年度比較基準（陳柏熹，2014；陳柏熹等，2015）。此外，部分大學亦透過「學習成效評估系統」或「學生學習成效平台」進行全校性的能力指標施測。例如，國立中山大學建置之 SELF-COLA 系統整合學習成效評量與資料庫機制，建立可長期追蹤之評量資料結構（林靜慧等，2017a, 2017b）。此類系統有助於累積跨年度能力資料，並作為教學改善與學生輔導之參考依據。

在教學評量層面，多數大學亦建置線上課程教學意見調查機制。例如，國立臺灣大學教務處透過線上問卷系統定期實施期中與期末教學意見調查，學生於指定期間登入校務系統填答，調查結果提供教師與教務單位作為課程改進與教學調整之參考依據。另一方面，國立中央大學亦建置線上教學評量系統，學生透過平台完成課程教學意見回饋，並由教學發展與行政單位彙整分析。此類系統通常採固定試題與標準化流程設計，強調施測機制之穩定性與資料之可比較性。雖未涉及適性選題或即時能力估計等進階測量技術，但在大規模施測、行政效率與資料管理層面，已顯示電腦化施測相較傳統紙筆模式之制度優勢。

整體而言，此一階段之電腦化應用，主要著重於施測媒介與資料流程之數位轉換，而非測量模型之革新。然而，其在建立施測規範、培養使用者熟悉度及累積題庫與資料管理經驗方面，仍扮演關鍵過渡角色，為後續題庫導向測驗與電腦適性測驗之發展奠定基礎。

二、題庫導向與能力檢測型電腦化測驗

在電腦化問卷與固定題本測驗逐漸普及後，我國大學亦於能力檢測與分級測驗情境中導入題庫導向之電腦化測驗。相較於前述之基礎型評量，此類測驗更著重於試題資源之整合與組卷（test assembly）彈性，使施測單位得以依能力層級或課程需求建構不同版本之測驗題本。此類測驗在實務上可再區分為分級測驗與題庫導向能力檢定兩種類型，其在題庫管理與測量設計上之成熟度亦有所差異。

在語言能力分級測驗方面，部分大學語言中心於新生入學時實施電腦化分級測驗，以作為課程分班依據。例如，靜宜大學外語教學中心設置「大一英文線上分級測驗（Freshman English Online Placement Test）」，透過線上施測方式進行能力分級與課程安排。此類分級測驗多以電腦化平台施測，並依測驗結果進行分班配置。在部分情境中，分級測驗亦可能結合題庫與難度配置機制，以提升不同施測梯次間之測驗一致性並降低試題重複與外洩風險。

除語言能力外，大學資訊能力檢定亦為題庫導向電腦化測驗的重要應用場域。例如，國立高雄大學設有資訊能力檢定制度，測驗試題由題庫隨機選題，並由教學發展中心規劃施測與成績認定機制。國立虎尾科技大學管理學院亦實施資訊能力檢定測驗，作為學生能力檢核與畢業門檻之一，其測驗由命題委員建立試題來源，並於每次檢定中依類型抽題施測。此類檢定通常透過線上施測系統進行，並結合題目抽選與自動計分機制，使能力檢測具備制度化與常態化特徵，並初步呈現題庫管理與組卷邏輯之運作模式。

整體而言，題庫導向與能力檢測型電腦化測驗顯示，我國大學電腦化評量之發展，已由單純的施測媒介轉換，進一步邁向測驗資源整合與組卷策略優化階段。其制度意義在於建立可持續維護之試題資源體系，透過版本控管與題庫管理提升測驗安全性與施測彈性，並為後續結合心理計量模型之電腦適性測驗奠定制度與技術基礎。此類測驗雖尚未導入即時能力估計機制，但已具備初步之題庫標記與組卷邏輯，顯示其已由固定題本模式逐步朝向測驗結構精緻化方向發展。

三、電腦適性測驗與進階電腦化測量

在電腦化問卷與題庫導向測驗逐步成熟的基礎上，部分研究團隊開始導入結合心理計量模型之電腦適性測驗（CAT），使電腦化評量由固定試題配置進一

步發展為依受測者能力即時調整測驗內容之進階模式。相較於固定題本或隨機選題測驗，CAT 透過即時能力估計與動態選題機制，為每位受測者提供最具訊息量之試題，在維持測量精準度之同時縮短測驗長度與作答時間（Chen & Chao, 2024, 2025; van der Linden & Glas, 2010; Wainer, 2000）。

CAT 的核心技術建立於試題反應理論（IRT）之共同量尺（common scale）架構，使不同受測者即使作答不同試題，仍能在同一能力尺度上進行比較。此一機制突破傳統測驗必須「所有受測者作答相同試題」之限制，使測驗設計由試卷導向轉為能力估計導向。相較於題庫導向測驗僅著重於版本管理與隨機抽題，CAT 更進一步將試題訊息函數（item information function）與能力估計模型納入選題算則，提升測量效率。

近年來，我國大學亦開始出現電腦適性測驗的實務嘗試。例如，國立中山大學研究團隊曾以大學生語言與資訊能力評量為目標，建置在地化的電腦適性測驗系統，從題庫發展、試題參數校正（item parameter calibration）、到施測流程，均依循完整的測驗建置架構進行（Huang et al., 2022）。透過大規模學生資料的蒐集與分析，不僅能建立穩定的試題參數估計值，也使適性測驗在本土教育情境中逐步累積可行性經驗。相關研究結果顯示，在相同或較短的測驗長度下，適性測驗仍能維持穩定的測量精準度，顯示其在提升施測效率與改善受測體驗上的潛在優勢（Huang et al., 2022）。這類大學場域的實證成果，象徵電腦化測驗在我國已由基礎應用逐漸邁向精準測量的階段。

然而，電腦適性測驗之推動亦涉及較高之制度與技術門檻，包括大量預試資料之蒐集、試題參數校正、題庫維護與系統開發成本等。因此，目前多數大學之 CAT 應用仍以研究、試辦或特定能力測驗為主，尚未全面取代既有固定題本或題庫導向測驗模式。儘管如此，其所代表之「能力估計導向」測驗思維，已為我國大學電腦化評量發展開啟新的可能方向。

整體而言，電腦適性測驗與進階電腦化測量之出現，標誌著我國大學電腦化測驗由媒介數位化與題庫管理階段，邁向測量模型導向與測驗精緻化階段。此一發展不僅提升測驗效率與精準度，亦反映測量設計逐漸精緻化的趨勢。

四、大學電腦化測驗經驗之啟示

綜合前述三種型態可見，我國大學電腦化測驗實務呈現由「低技術門檻、重

施測流程與回饋」逐步走向「以題庫管理與測量模型支撐的精準評量」之演進趨勢。就功能定位而言，電腦化問卷與能力評估測驗著重於施測流程之建立與資料回饋機制；題庫導向電腦化測驗則強調題庫資源之維護與版本控管，以提升施測彈性與測驗安全性；電腦適性測驗則進一步由「試卷導向」轉向「能力估計導向」，對試題參數校正、等化與系統運算等提出更高要求（Wainer, 2000）。

從制度推動角度觀察，大學場域的經驗顯示，電腦化測驗之穩定運作仰賴多項要素之整合：其一為施測流程之標準化與可備援性（含身分驗證、作答紀錄保存、異常處置與資料回收）；其二為題庫與測驗資料之管理機制（含試題編碼、版本控管、安全性監控與試題品質維護）；其三為使用者支持與可及性設計（含操作引導、示範測驗與無障礙需求之配套）。此三者共同構成電腦化測驗制度化運作之基礎條件（AERA, APA, & NCME, 2014; International Test Commission, 2006）。

整體而言，大學場域之實務經驗顯示，電腦化測驗之發展涉及施測流程、題庫管理與測量模型等多層面之整合，並呈現由施測媒介轉換逐步邁向測量導向之發展趨勢。

參、我國國家考試電腦化測驗之發展現況與關鍵議題

相較於高等教育場域之電腦化評量多由教學與校務需求逐步發展，我國國家考試制度之電腦化轉型則建立於試務效率、測驗安全與行政管理優化之考量。近年來，考選部已逐步推動電腦化測驗制度，包括建置電腦試場與發展電腦化作答系統，並於部分類科推動電腦化施測（如醫師、心理師考試），使國家考試正式邁入數位化施測與資訊化管理階段。

在基礎設施方面，考選部透過電腦試場認證機制與大專校院合作推動試場建置，使電腦化施測資源逐步擴充。相關資料顯示，全國電腦化考試座位數由 2020 年約 5,900 席增加至 2024 年約 16,000 席，顯示國家考試電腦化施測之基礎設施正快速成長（余慶杉，2024）。此一發展亦反映出我國國家考試電腦化已由政策規劃階段邁入制度實作與擴展階段。

值得注意的是，國家考試之電腦化發展脈絡，與大學場域並不完全相同。相較於大學端多由教學評量需求逐步帶動，國家考試中的題庫建置、電腦化施測與試務數位化等措施，可能因不同考試類科特性與制度需求而分別推動，未必呈現

單一線性發展關係。此一差異亦顯示，國家考試之電腦化發展除測量技術因素外，亦涉及試務治理與制度推動等多重考量。

在此背景下，國家考試電腦化發展的討論焦點，已由「是否導入電腦化」轉向「如何深化與精緻化電腦化測驗制度」，包括固定題本電腦化施測之穩定運作、題庫管理能力之強化、多梯次施測之可行性，以及未來是否朝向更進階測量模式發展等議題。以下即依不同發展層次，分析其制度內涵與關鍵挑戰。

一、固定題本電腦化施測之制度深化

目前國家考試之電腦化施測，多採固定題本模式，在維持既有測驗內容與評分機制的前提下，將紙筆作答轉換為電腦介面進行施測。此種模式可在不改變測驗架構與評量邏輯的情況下完成媒介轉換，因此成為多數大型考試推動電腦化初期常見之發展策略。此外，電腦化施測亦有助於導入多媒體與互動式試題素材，例如影像、音訊、動畫或模擬操作情境等，使部分專業能力之評量能更貼近實務情境，提升測驗之生態效度與評量精準度。

在制度運作層面，固定題本電腦化施測已產生多項試務管理效益。首先，自動計分與資料整合機制可減少人工閱卷與資料輸入流程，提升行政效率並降低人為錯誤風險。其次，電腦化作答系統可完整保存作答紀錄與操作歷程，例如登入驗證、作答時間追蹤與異常操作監測等，有助於強化測驗安全與試務監控能力。此外，電腦化施測亦可減少試卷印製與運送成本，對長期辦理的大型考試而言，具有顯著的資源管理優勢。

然而，固定題本電腦化施測之穩定運作，仍仰賴完善之試場環境與備援機制。現行電腦化國家考試多於經認證之電腦試場辦理，而這些試場多設置於大專校院資訊教室或電腦教學環境，使高等教育機構在國家考試電腦化基礎設施中扮演重要角色（余慶杉，2024）。未來制度深化可進一步著重於設備規格、網路穩定度與系統備援標準之整合。

二、題庫導向電腦化測驗與試題管理能力

在固定題本電腦化施測逐步穩定後，題庫導向電腦化測驗可視為國家考試電腦化發展的重要制度方向之一。題庫式測驗不僅涉及試題組卷方式的改變，更象徵試務管理由傳統「試卷管理」逐步邁向「題庫管理」的制度轉型。

在題庫式架構下，試題不再以單一試卷形式存在，而是以可分析、可組合與

可監控的題庫資源形式管理。此一制度轉變要求試務單位建立更完善的試題分析與品質控管機制，包括試題難度與鑑別度監測、試題曝光控管（item exposure control）、組卷策略、試題參數校正與共同量尺建立等（Chao & Chen, 2023; Simpson & Hetter, 1985; van der Linden, 2005）。透過長期累積作答資料與試題統計資訊，試務單位得以持續監控題庫品質，確保不同試題組合間仍具公平性與可比較性。

此外，題庫式測驗亦可為多梯次施測提供制度彈性，使考試單位得以分散施測壓力並提升試場利用效率。然而，多梯次施測同時涉及測驗安全與等化問題，因此在制度設計上需建立嚴謹的試題曝光控制與測驗難度平衡機制。此類發展亦反映試題管理方式由傳統試卷導向，逐步轉向資料導向之測驗管理模式。

值得注意的是，現行國家考試制度多於考後公開試題與答案，以確保考試透明度並維持社會信任。然而，題庫式電腦化測驗的運作通常需仰賴試題的長期使用與曝光控制，以維持題庫穩定性與測驗安全。試題全面公開雖有助於提升制度透明度，卻可能增加題庫消耗速度，進而限制題庫式測驗與多梯次施測之發展。因此，未來若欲推動題庫化或適性測驗，如何在測驗透明性與題庫管理之間取得平衡，將是制度設計上需持續面對的課題。

然而，電腦化測驗之推動未必皆需採取題庫導向模式。就國家考試實務而言，不同考試類科在試題型態、專業領域特性與命題需求上存在相當差異，因此除題庫導向測驗外，部分科目亦可能採取臨時命題或較具彈性的命題模式。在電腦化施測環境下，此類模式雖未建立長期題庫架構，惟在考試類科眾多、試題更新需求差異較大的情況下，仍可作為部分類科推動電腦化施測之可行方式。若能配合封閉安全之命題與審查機制，並透過專業領域專家進行試題品質檢核，則其在內容效度與測驗品質上，仍可維持與傳統紙筆測驗相當之水準。此外，臨時命題亦有助於即時反映法規修正、專業知識更新或社會議題變化，使測驗內容能更貼近當前專業實務需求。整體而言，未來國家考試之電腦化推動，或可同時包含題庫導向與臨時命題等不同模式，以保留制度運作與命題設計之彈性。

三、電腦適性測驗與長期測量策略

在固定題本與題庫式測驗制度逐步成熟後，電腦適性測驗（CAT）可被視為國家考試電腦化發展的長期策略之一。CAT 建立於試題反應理論（IRT）的共同

量尺架構，使不同試題組合仍能在同一能力尺度上進行比較，並透過即時能力估計與動態選題機制，使測驗能依受測者能力水準選取試題，在維持測量精準度的同時縮短測驗長度與作答時間（Wainer, 2000; van der Linden & Glas, 2010）。

然而，相較於固定題本或題庫式測驗，CAT 對題庫品質、試題參數穩定性與制度透明度的要求更高。由於不同考生可能面對不同試題組合，制度推動必須清楚說明共同量尺與等化原理，以維持社會大眾對測驗公平性的理解與信任。因此，在高風險的國家考試情境中，CAT 較適合作為中長期發展方向，而非短期全面推動的制度選項。

在技術層面上，電腦適性測驗之推動亦涉及試題參數估計與能力估計精準度之關鍵議題。於 IRT 架構下，受測者能力估計之標準誤（standard error, SE）主要由試題訊息函數所決定，並受試題參數穩定性影響，若試題參數校正偏誤或樣本數不足，將可能導致試題訊息量估計偏誤，進而影響試題選取與能力估計之準確性（Embretson & Reise, 2000; van der Linden & Glas, 2000）。在高風險測驗情境中，此類估計誤差可能進一步影響考生排名與決策結果之公平性。因此，於導入適性測驗前，須確保題庫具備充分之試題數量與穩定參數估計基礎，並透過持續性的參數校正與統計監測機制，以維持測量品質並確保測驗信、效度。

從測驗發展脈絡來看，固定題本電腦化施測與題庫管理能力的建立，為發展適性測驗的重要前提。當題庫累積足夠試題數量與穩定的作答資料後，考試單位方能透過試題校正與統計分析，逐步探索更精緻的測量策略。因此，CAT 可被視為在制度與題庫基礎逐步累積後所發展之測量策略，其推動時機與應用範圍仍需審慎評估。

值得進一步關注的是，隨著電腦化測驗技術之發展，評量焦點亦逐漸由傳統結果導向（product-oriented）轉向歷程導向（process-oriented）的測量觀點。在電腦化施測環境中，系統可自動記錄受測者之作答歷程資料（如作答時間、操作程序與互動行為等），使評量不再僅依據最終作答結果，而得以進一步分析其問題解決過程與決策策略。

此類歷程導向評量已廣泛應用於醫療專業證照考試中。例如，美國醫師執照考試第三階段（USMLE Step 3）透過電腦化臨床案例模擬（computer-based case simulations），記錄考生之診斷決策程序與時間歷程，並依據完整行為軌跡進行評分（Clauser et al., 2002; Dillon et al., 2004）；放射科專科醫師考試（ABR Core

Exam) 則要求考生於影像系統中進行影像瀏覽與參數調整等操作，以評估其臨床判讀與操作歷程 (Becker & Dunnick, 2008; Hollingsworth et al., 2010)。

相關發展顯示，電腦化測驗已能透過歷程資料有效捕捉受測者之認知策略與實務能力 (He et al., 2021; Parshall et al., 2002)。對於高風險之國家考試而言，未來若能適度導入此類歷程導向評量，不僅有助於提升測驗之生態效度，亦可更全面評估考生之專業判斷能力。

四、推動關鍵議題與風險管理

在國家考試逐步推動電腦化施測的背景下，制度深化之核心課題在於風險管理與制度信任的建立。電腦化測驗的推動不僅涉及技術導入，亦需同步處理法規制度、試務管理、資訊安全與公平性等多面向議題。

首先，在法規與作業規範面向，需建立明確的電腦化施測作業標準與異常處理機制，包括系統故障、網路中斷或設備異常時之應變流程、補考規則與證據保存方式，以維持制度運作之可預期性。其次，在試務運作層面，需確保考場設備與環境的一致性，並建立備援機制與演練制度，使試務人員能在異常情境下迅速處置。第三，在資訊安全與測驗安全面向，應建立多層次資安防護機制，例如身分驗證、權限控管、資料加密與作答紀錄稽核等，以確保試題安全與系統穩定。此外，隨著生成式人工智慧 (artificial intelligence, AI) 工具之普及，考生於電腦化測驗過程中可能透過外部工具取得不當協助之風險亦逐漸浮現，因此在測驗安全設計上，亦需納入 AI 介入風險之防範機制，例如強化作答環境控管、瀏覽行為監測、異常作答歷程分析，以及結合人機協作之監考模式，以降低非能力因素對測驗結果之干擾。第四，在公平性與可及性面向，應提供示範測驗與操作說明，並透過無障礙設計與合理調整措施，降低非能力因素對作答表現的影響。最後，在制度溝通層面，需建立透明且可理解的說明機制，特別是在題庫式或適性測驗情境下，清楚說明測驗運作原理與品質監控機制，以利制度理解與接受性。

整體而言，國家考試電腦化之推動涉及多面向制度條件之整合，其發展重點在於風險管理機制與制度運作之持續強化。

肆、結論與建議

綜合前述討論可見，電腦化測驗在我國大學場域的發展，已呈現由基礎應用逐步邁向進階測量的明確趨勢。從電腦化問卷與能力評估測驗的普及，到題庫導

向電腦化測驗的制度化應用，再到電腦適性測驗的研究與試辦，大學端已累積相當程度的實務經驗與技術基礎（Huang et al., 2022; 林靜慧、陳俊宏，2017）。這些發展顯示，電腦化測驗並非單一技術的導入，而是一個隨著測驗需求、題庫管理與測量模型逐步深化的制度演進過程，也為思考國家考試的數位轉型提供了重要參考脈絡。惟考量國家考試具備高風險與大規模之制度特性，其發展脈絡與大學場域仍存在本質差異，難以直接全面沿用，亟需建立專屬於考選制度的深化路徑。

對國家考試而言，電腦化測驗的發展重點，已不再只是導入電腦化施測與否，而是如何在既有制度基礎上進一步深化與精緻化。換言之，現階段的核心課題，在於如何於維持測驗公平性、制度穩定性與社會信任的前提下，持續優化固定題本電腦化施測的作業品質，逐步強化題庫治理能力，並在條件成熟時審慎評估題庫式測驗與電腦適性測驗的發展可能性。而為保留制度運作與命題設計之彈性，未來國家考試之電腦化推動，或可同時包含題庫導向與臨時命題等不同模式，以因應不同類科之專業特性與現實需求。此一由低風險至高精準度的推進路徑，不僅有助於降低制度轉型的不確定性，也能兼顧測驗效率、品質與公平性。

具體而言，未來可從短、中、長程三個層面規劃。短期可聚焦於固定題本電腦化施測之制度常態化，持續完備考場設備標準、作業流程、備援機制與考生熟悉措施；中期則可著力於題庫管理能力之建置，包括試題編碼、版本控管、曝光監測、試題校正與資料分析，使制度由「試卷管理」逐步邁向「題庫管理」；長期則在題庫規模、測驗資料累積與品質監測機制成熟後，再審慎評估適性測驗或更精緻電腦化測量模式之試辦可行性。在此過程中，可進一步結合具體之檢核機制。例如，在統計層面，可於題庫發展階段進行模擬研究（simulation studies），在考量實務測驗限制（如內容平衡、曝光控制與測驗長度限制等），評估能力估計之偏誤與測量精確度，以作為題庫品質與測量模式可行性之重要依據；在行政配套層面，則需同步檢視制度推動所需之支持條件，包括考場硬體設備與網路穩定性、試務流程與異常處理機制之完備性，以及考生對電腦化測驗之熟悉程度與制度信任基礎。尤其在現行國家考試仍多於考後公開試題的制度安排下，如何在測驗透明性與題庫管理之間取得平衡，亦需納入制度設計之整體考量，作為後續制度深化的重要依據。

整體而言，我國大學在電腦化測驗上的多元實務，不僅提供了觀察高等教育

評量數位化發展的重要窗口，也為國家考試電腦化制度的深化治理提供了可參照的制度經驗。此外，隨著資訊技術快速發展，資訊素養已逐漸成為多數專業領域之核心能力之一，未來國家考試在測驗內容與評量方式上亦可能面臨調整需求。例如，除傳統知識測驗外，如何評量考生運用 AI 工具進行問題解決之能力，以及在適當規範下導入自動化評分機制，均可能成為後續制度發展的重要議題。在此基礎上，未來若能在試場基礎設施、題庫管理、測量技術與社會溝通等層面持續累積成果，並建立跨單位合作與試辦回饋機制，電腦化測驗將有機會在我國考選制度中發揮更具前瞻性與制度化的角色。

伍、參考文獻

- 王思峰、吳濟聰、夏侯欣鵬、劉淑慧、鄔榮霖、周志明、蔡秦倫（2014）。大學生涯資訊系統 CVHS-M1 評估研究：生涯行為、輔導意涵與影響。**中華輔導與諮商學報**，**40**，95-126。
- 余慶杉（2024）。起風與揚帆！國考數位轉型。**國家人力資源論壇**，**32**。
https://www.exam.gov.tw/NHRF/News_EpaperContent.aspx?n=3778&s=48945&type=5DE406C89FFD8BCF
- 林靜慧、陳俊宏（2017）。校務研究與內部品質保證。**評鑑雙月刊**，**65**，9-12。
- 林靜慧、陳俊宏、施慶麟（2017a）。使用學習成效系統（SELF-COLA）評估大學生學習經驗。**評鑑雙月刊**，**66**，29-32。
- 林靜慧、陳俊宏、施慶麟（2017b）。建立以數據驅動教育決策的學生學習成效平台。**教育研究月刊**，**280**，17-29。
<https://doi.org/10.3966/168063602017080280002>
- 陳柏熹（2014）。電腦化大學生基本素養測驗簡介。**評鑑雙月刊**，**47**，24-29。
- 陳柏熹、黃馨瑩、陳郁欣、葉泰廷、蘇少祖（2015）。大學生基本素養測驗的發展及信度效度分析。**教育科學研究期刊**，**60**（3），95-126。
[https://doi.org/10.6209/JORIES.2015.60\(3\).04](https://doi.org/10.6209/JORIES.2015.60(3).04)
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

- Becker, G. J., & Dunnick, N. R. (2008). Intended consequences of computer-based core and certifying examinations in diagnostic radiology. *American Journal of Roentgenology*, 191(5), 1302–1305. <https://doi.org/10.2214/AJR.08.1232>
- Chao, H. Y., & Chen, J. H. (2023). Controlling the minimum item exposure rate in computerized adaptive testing: A two-stage Symptom–Hetter procedure. *Applied Psychological Measurement*, 47(7-8), 460-477. <https://doi.org/10.1177/01466216231209756>
- Chen, J. H., & Chao, H. Y. (2024). Utilizing real-time test data to solve attenuation paradox in computerized adaptive testing to enhance optimal design. *Journal of Educational and Behavioral Statistics*, 49(4), 630-657. <https://doi.org/10.3102/10769986231197666>
- Chen, J. H., & Chao, H. Y. (2025). Maximin criterion for item selection in computerized adaptive testing. *Behavior Research Methods*, 57(7), 180. <https://doi.org/10.3758/s13428-025-02673-8>
- Clauser, B. E., Margolis, M. J., & Swanson, D. B. (2002). An examination of the contribution of computer-based case simulations to the USMLE Step 3 examination. *Academic Medicine*, 77(Suppl. 1), S80-S82. <https://doi.org/10.1097/00001888-200210001-00026>
- Dillon, G. F., Boulet, J. R., Hawkins, R. E., & Swanson, D. B. (2004). Simulations in the United States Medical Licensing Examination (USMLE). *Quality and Safety in Health Care*, 13(Suppl. 1), i41–i45. <https://doi.org/10.1136/qshc.2004.010025>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, 166, Article 104170. <https://doi.org/10.1016/j.compedu.2021.104170>
- Hollingsworth, C. L., Wriston, C. C., Bisset, G. S., Strife, J. L., Bosma, J. L., Gerdeman, A. M., & Becker, G. J. (2010). American Board of Radiology certifying examination: Oral versus computer-based format. *American Journal*

- of Roentgenology*, 195(4), 820-824. <https://doi.org/10.2214/AJR.09.3618>
- Huang, H.-T. D., Hung, S.-T. A., Chao, H.-Y., Chen, J.-H., Lin, T.-P., & Shih, C.-L. (2022). Developing and validating a computerized adaptive testing system for measuring the English proficiency of Taiwanese EFL university students. *Language Assessment Quarterly*, 19(2), 162-188. <https://doi.org/10.1080/15434303.2021.1984490>
- International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143-171. https://doi.org/10.1207/s15327574ijt0602_4
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. Springer.
- Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. In *Proceedings of the 27th annual meeting of the Military Testing Association* (pp. 973-977). Navy Personnel Research and Development Center.
- van der Linden, W. J. (2005). *Linear models for optimal test design*. Springer. <https://doi.org/10.1007/0-387-29054-0>
- van der Linden, W. J., & Glas, C. A. W. (2000). Capitalization on item calibration error in adaptive testing. *Applied Measurement in Education*, 13(1), 35-53. https://doi.org/10.1207/s15324818ame1301_2
- van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Elements of adaptive testing*. Springer. <https://doi.org/10.1007/978-0-387-85461-8>
- Wainer, H. (Ed.). (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum Associates.