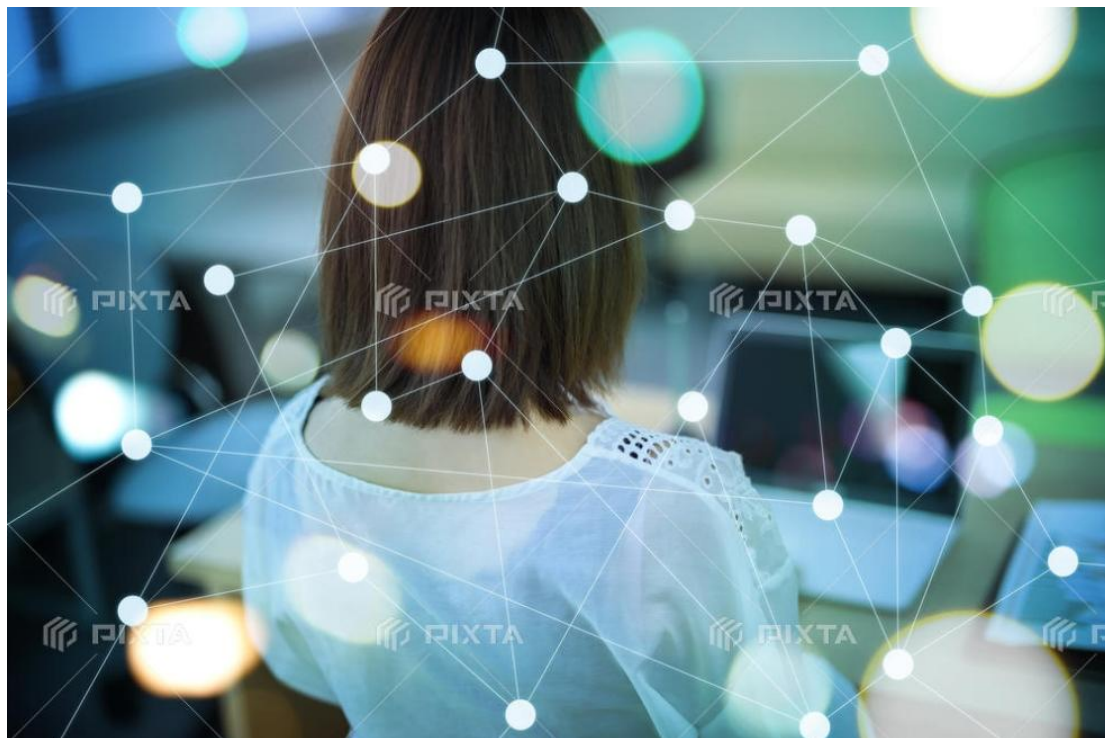


人工智慧與新冠肺炎疫情對測驗與評量發展與實施的影響



圖片來源：Pixa

【測驗及評量研究中心助理研究員 陳繼成】

壹、前言

測驗與評量的發展往往隨著教育一同演進。在中國，最早的測驗可追溯至隋代（西元 581-619 年）的科舉制度。隨著各朝代的不同，考試的科目與方式均有些許不同。若風氣為重文輕武，考試內容偏向詩書五經；反之則亦會加入體能的評測等。在制度上，各朝代的科舉制度隨著時間推移而更加完備。如同讀者們在歷史課本上學的「科舉制度起於隋、確立於唐，自北宋而大備。」顯示考試的制度隨著時間的推移而越發完整。各種制度的演進一方面期待能適切地用人選材，一方面則避免營私舞弊。在國外的測驗發展中，醫療和教育等領域在 19 世紀就有使用測驗的紀錄。然測驗的推廣與快速成長則是在第一次世界大戰美國參戰後對於欲參軍的人員進行的選材。

近一個世紀以來，電腦與網路的發展對於在此之前一千多年均很穩定的測驗理論與測驗形式迎來改變的曙光。此時，電腦化測驗與相關測驗理論與技術也跟著蓬勃發展。然 2019 年底至 2020 年初，全球亦迎來新冠肺炎（COVID 19）的肆虐。為了兼顧

學習與健康，教育部採用停課不停學策略，積極推動線上課程。無疑這便是一個可能對測驗發展造成影響的一個重大事件。2022 年，ChatGPT 引領了全球對人工智慧（Artificial Intelligence，AI）的重視，此時 ChatGPT 與 AI 對教育帶來無法忽視的衝擊，而教育測驗的型態與內容亦收到影響。在面對如此遽變的環境下，各國的測驗與評量受到何種影響，發展趨勢為何？本文聚焦於近年來測驗評量的發展，蒐集不同的國家分別受到人工智慧（AI）與新冠肺炎疫情的影響，在大型測驗上所做的調整以及在測驗的發展上所遇上的情況以及所做的決策，以期能理出一絲脈絡，提供讀者作為參考。

貳、當代考試典範的轉移

隨著電腦與網路的發展，考試的發展方式已逐漸從紙筆測驗(Paper & Pencil testing)轉型至電腦化測驗(Computerized testing)。此外，多元評量(Multiple Assessments)也逐漸發展在不同的領域上(Barton and Donahue, 2009; Chan, 2008)，其包含面試與實作評量等不同評量學生學習成果的方式。此外，在考試內容方面，傳統較重視學生學科的方向也逐漸改變為素養導向的評估。三面九向之核心素養亦成為本國課程綱要中課程發展的主軸(教育部，2014)。而此測驗與評量的發展在不同國家亦呈現一致的趨勢，可歸納出電腦化測驗與學生素養表現之評估為目前發展的趨勢。

一、電腦化測驗

眾所周知，國際大型評比的各項測驗均已電腦化，尤其是 IEA (International Association for the Evaluation of Educational Achievement) 的各項測驗(如，Progress in International Reading Literacy Study 與 Trends in International Mathematics and Science Study 等)均為電腦化施測。其他國家亦發展其電腦化測驗與系統。如韓國於 2022 年全面引進電腦化學習成就評量*(i 成就)測量學生的各項能力(駐韓國代表處教育組，2022)。英國亦於冠狀病毒疫情發展，實施線上教學和線上評量(駐英國代表處教育組，2020a)。此外，比利時布魯塞爾自由大學(Université Libre de Bruxelles，ULB)亦因應疫情發展採線上測驗(駐歐盟兼駐比利時代表處教育組，2020)。

二、考試內容與科目的改變

在考試內容的部分，各國的考試內容也逐漸增加關於非認知能力、學生素養的比

重，尤其關注於創造力、資訊能力、問題解決能力等。以韓國為例，期全面引進的電腦化學習成就評量系統則關注於學生的問題解決能力、信息處理能力等未來社會所需的能力（駐韓國代表處教育組，2022）。此外，日本國立大學的入學測驗亦新增「情報」（資訊）考科（駐大阪辦事處，2021）。

參、受到疫情影響的發展

隨著電腦網路的發展，電腦網路應用在測驗的情境也越來越受到重視。與此同時，測驗與評量的發展受到一全球性事件的影響，新冠肺炎的肆虐。在疫情的影響下，學生的表現是否受到影響？在面對疫情的時候測驗與評量有哪些措施？後續又會朝著哪些方向前進？

首先，無可厚非學生的表現受到學習場域轉變、學習監督者的改變或是學習方式不同的影響而下滑（駐英國代表處教育組，2020b）。相同的學生表現下降的也包含美國地區，調查顯示新澤西公校中學生的英數成績退步（駐紐約辦事處教育組，2022a）、全美四年級與八年級學生的數學、閱讀成績也下滑（駐紐約辦事處教育組，2022b；駐休士頓辦事處教育組，2022）。此外，日本的調查亦顯示受疫情影響，小學生的體能狀態下降（駐日本代表處教育組，2021）。各國均顯示疫情對學生在各方面的成長均有負面的影響。

為了防治疫情擴散，各國在考試上亦會因此而改變考試型態。最直接的就是在疫情嚴重的時候暫停考試，如越南胡志明市的國際外語能力測驗和河內的思維能力測驗考試招生均因疫情而停辦（駐胡志明市辦事處教育組，2021；駐越南代表處教育組，2021）。無獨有偶，美國亦於 2020 年取消六月的 SAT（Scholastic Assessment Test）考試（駐芝加哥辦事處教育組，2020）。顯示在疫情初期，現場考試均受到疫情不小的影響。然暫停考試僅是過渡的階段，各國政府在面對考試的時候亦逐漸有配套措施。如美國大學委員會曾考慮在家施測（駐芝加哥辦事處教育組，2020），雖然受到了不少政府官員和輿論的質疑，但是若在險峻之時，亦為不得之選。此外，測驗評量領域已逐漸成熟的線上測驗此時也派上用場。比利時法語布魯塞爾自由大學（ULB）因應疫情採線上測驗，ULB 在大家關注的公平性上選擇相信學生不安裝監控軟體。但是仍為了避免作弊而期待採用開書考試（駐歐盟兼駐比利時代表處教育組，2020）。英國的大學亦因應冠狀病毒疫情發展實施線上評量（駐英國代表處教育組，2020a）。此外，亦有

學校因為考試種類或是特性，無法避免不正當考試行為而無法採用線上測驗（駐福岡辦事處派駐人員，2021）。

肆、人工智慧（AI）的影響與應對

在電腦化測驗蓬勃發展的後期，除了疫情對電腦化測驗與線上測驗帶來影響之外，人工智慧（AI）對於測驗來說絕對是另外一個測驗發展上的里程碑。然一體兩面的，測驗的管理者可以使用人工智慧（AI）來讓測驗評量更為合理高效，測驗的受測者亦可以此作為答題手段。這件事情的普遍性也愈發受到重視，如日本市場行銷支援公司 Nyle 的一項調查結果顯示，約每三位學生中就會有一位學生使用生成式 AI 來進行暑假作業（駐大阪辦事處，2023）。

然而很明確的，評量上若非評量學生如何使用 AI 的專業知識，使用人工智慧（AI）的結果來作為個人成果的證明是不恰當的。日本文部科學省所公布的「中小學使用 AI 方針」中，就提及使用不當的情況包含：隨意使用生成型 AI 軟體產生的作品作為報告、論文，甚至參加徵文比賽等；在創作詩、音樂、美術等作品時使用該軟體等；在考試或測驗時讓學生使用該軟體（駐日本代表處教育組，2023）。英國教育部提出的「AI 人工智慧科技在學校的應用方針」中亦提到人工智慧（AI）並不能取代學生主動求知、查證的學習歷程（駐英國代表處教育組，2023）。亦即學生使用人工智慧（AI）進行的作業成果並不能表示其真實能力。就筆者初淺的認知，使用人工智慧（AI）確實為學生用來學習的突破性工具，不論在教師教學、學生學習或是評量上均為一可靠方便的工具，但是 AI 的專業知識確實並不能取代學生的專業知識，因此不論是在學習或測驗上，AI 的使用亦須受到適合的規範，方能對學習場域的參與者有更大的助益。

為了應對人工智慧（AI）對測驗與評量帶來的影響，各國的因應方式亦有所不同。避免學生考試時使用 AI 為其中一種應對的方式。在美國，有多數大學嘗試使用口試避免學生使用人工智慧（AI）進行不正當的考試行為（駐美國代表處教育組，2023）。澳洲的大學改變考試方式，恢復紙上測驗；駐法國代表處教育組（2023）亦稱澳洲與美國部分大學基於防止學生作弊的考量，決定禁止使用 ChatGPT。

值得注意的是，一味的禁止並不能取得最大的效益。英國專門提供考試規則並確保測驗品質的資格認證聯合委員會（Joint Council for Qualifications）已自行發布了「人工智能在測驗中的應用（AI Use in Assessments）」，指引教師在具備人工智慧（AI）的

環境下，如何評估學生的學習：首先，此指南開宗明義說明在測驗中學生可以使用人工智慧（AI）進行作答，就會受到此指南的規範。接著說明人工智慧（AI）軟體的種類（Chat GPT 僅是其中之一）、其能做到的事情（回答、分析、總結、寫文章、程式碼、翻譯、產生想法等）以及使用的風險（如，看來可信實則充滿偏見的回答、虛假引用、提供有危險的答案等）。接著說明學生使用人工智慧（AI）來進行作業屬於濫用行為（包含直接貼上部分或完整的段落、以及未正確完整引用等共六項），只有在學生的成果（考試或作業的產出）能被證明是獨立工作並獨立思考的的產物時才被允許使用人工智慧（AI）。此外，該指南亦提出其他可能被誤用的可能性、以及說明不同單位應該教學生正確的人工智慧（AI）使用方法、完整引用（包含使用截圖等）方式。

伍、結語

測驗與評量的發展在電腦發明前的發展較為緩慢，然卻無可否認地其確實會受到不同事件影響。近年來，全球性的大事件，新冠肺炎疫情的發展與 AI 人工智慧的崛起，對測驗與評量發展的影響甚鉅。筆者整理近年國際發展形勢後歸納三大發展趨勢。首先，電子產品做為載具的測驗型態逐漸普及，而新冠肺炎的發生以及 AI 人工智慧的發展卻對其造成了相反的方向。為了防止人與人的接觸造成健康疑慮，使用電子載具（如，電腦、筆記型電腦、平板或手機等）進行施測可以避免人與人的過多接觸，但是 AI 人工智慧的發展卻讓有些考試對電腦施測望之卻步。然此種測驗具備的強大優勢依舊不可忽視，包含測驗的型態極具彈性（支援多媒體試題與適性測驗等）、測驗研發成本回歸考科專業、亦可依需求降低人與人的接觸（因應疫情或是考生隱私）。其次，由於測驗的內容與型態會極大程度受到考試科目的影響。因此 AI 人工智慧的崛起雖然改善了我們的生活，同時學生的學習內容與教師的教學也會因此而必須進行調整。考試科目、考試方式、甚至到閱卷與評分均會隨之而異。舉例來說資訊素養逐年受到重視、多元評量方式的崛起、測驗理論的蓬勃發展、甚至將人工智慧（AI）導入閱卷的發展等均為近年測驗評量發展的目標。第三，政府或是測驗機構亦逐漸完善相關規範與制度，如讓考生在更公平的施測環境下進行測驗以因應 AI 人工智慧的發展，或是使用不同的考試方式來讓學生能在嚴峻的疫情下卻依舊有友善的考試環境而免於健康威脅等。而上述三個趨勢亦可作為臺灣在未來測驗評量發展上的參考。

參考文獻

教育部（2014）。十二年國民基本教育課程綱要總綱。取自

<https://www.naer.edu.tw/upload/1/16/doc/288/%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%B8%BD%E7%B6%B1.pdf>

駐大阪辦事處（2023）。日本市調公司調查 34%學生利用 AI 進行暑假作業。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2064125

駐大阪辦事處（2021）。日本國立大學入學測驗將新增「情報」考科。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2058927

駐日本代表處教育組（2023）。日本文科省公佈中小學生成型 AI 軟體使用方針。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2063777

駐日本代表處教育組（2021）。受新冠肺炎疫情影響，日本學童體能退步。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2058738

駐休士頓辦事處教育組（2021）。全美 4 年級及 8 年級學生成績大退步。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2060693

駐芝加哥辦事處教育組（2020）。美國大學委員會取消 6 月 SAT 考試，並規劃在家測驗方案。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2052499

駐法國代表處教育組（2023）。法國大學面對人工智慧程式 ChatGPT 所帶來的教學挑戰與契機。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2062138

駐胡志明市辦事處教育組（2021）。因 COVID-19 病毒，國際外語能力測驗繼續暫停辦理。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2052505

駐美國代表處教育組（2023）。美國多數大學嘗試採用口試測驗因應大學人工智能的考試作弊。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2063747

駐英國代表處教育組（2023）。英國教育部提出 AI 人工智慧科技在學校的應用方針。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2063611

駐英國代表處教育組（2020a）。英國大學因應冠狀病毒疫情發展，所實施的線上教學和評量。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2052459

駐英國代表處教育組（2020b）。英國教育標準局研究顯示因新冠肺炎疫情關閉學校影響學生學習甚鉅。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2054633

駐紐約辦事處教育組（2022a）。疫情後整體學習表現不佳，新澤西公校生英數成績退步。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2060845

駐紐約辦事處教育組（2022b）。疫情所致？全美各州學生數學、閱讀成績創紀錄退步。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2060560

駐越南代表處教育組（2021）。2022 年河內許多大學將通過思維能力測驗考試招生。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2058627

駐福岡辦事處派駐人員（2021）。大學入學考試採用線上測驗的學校占 2.9%。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2057756

駐歐盟兼駐比利時代表處教育組（2020）。比利時法語布魯塞爾自由大學(ULB)因新冠肺炎疫情採線上測驗相關配套措施。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2052928

駐韓國代表處教育組（2022）。2021 年韓國國家學習成就評量結果分析及提高學習成就水平支援方案（一）。國家教育研究院臺灣教育研究資訊網。

https://teric.naer.edu.tw/wSite/ct?ctNode=647&mp=teric_b&xItem=2060096

Barton, A., & Donahue, C. (2009). Multiple assessments of a first-year seminar pilot. *The Journal of General Education*, 58(4), 259-278.

Chan, Y. C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English language teaching*, 7(1), 37-62.