

# 考試院第 12 屆第 298 次會議考選部重要業務報告

民國 109 年 8 月 20 日

## 壹、考選行政

### 國家考試採行 AI 輔助閱卷之可行性評估

#### 一、前言

本部掌理全國考選行政事宜，依用人機關與專技執業需求，每年舉行約 19 次考試，為我國最主要之掄才管道。考量現行紙筆測驗之申論式試題係延聘閱卷委員進行人工評閱，為期提高閱卷效率，本部前於 107 年透過科技部「AI 創新研究中心專案計畫」與國立清華大學合作，針對自然語言申論題輔助評分系統等 7 大主題進行為期 4 年之研究計畫，惟該團隊未能接續獲得科技部第 2 年之經費補助，爰終止該項專案計畫。

衡酌近年因機器學習(Machine Learning, ML)與巨量資料(Big Data, BD)等分析技術具有突破性進展，使得人工智慧 (Artificial Intelligence, AI)再次廣泛運用於資通訊發展領域。爰本部期借鏡國內外運用 AI 輔助自動閱卷評分之應用現況，探尋國家考試數位化作業未來發展之可能性。

#### 二、國內外發展概況

國內外主要測驗機構大多已關注或長期投入電腦自動評分輔助閱卷之研究及運用，惟均侷限於中英文作文寫作範疇，尚未應用至特定專業領域申論題，茲就其發展現況與相關限制說明如下：

- (一)大學入學考試中心：僅止於關注此項議題，尚未啟動研究規劃。
- (二)國立臺灣師範大學心理與教育測驗研究發展中心(以下稱心測中心)：心測中心負責國中教育會考之命題、組卷、閱卷及計分，業研究中文寫作 10 餘年，惟考量國情針對電腦評分信心、文章立意、是否符合題旨等因素，初步規劃作為品質監控一環，尚未納入正式評分中。

(三)美國 ETS 之 e-rater 自動評分系統：為國際間最早投入英文寫作自動評分研究機構，已發展 30 餘年，技術最為純熟，廣為應用於 GRE、TOEFL 等考試，並採「人機雙閱」共同評閱模式，然從 ETS 官方研究報告中發現，其評分系統之演算法仍存在「作答特徵」(如篇幅較長、使用多量複雜詞彙等)偏誤問題及無法辨識文章是否為「東拼西湊」之嚴重缺陷。

(四)英國 Write & Improve 作文評分與回饋系統：由英國劍橋大學所屬劍橋英語語言測評中心(Cambridge Assessment English)提供之免費英文寫作教學及練習環境，分成初級、中級、高級 3 等級，每一等級均有逾 30 個題目供使用者進行寫作練習，系統除可評量使用者寫作程度外，亦可回饋基本文法及拼字錯誤之處，惟現階段尚未對於單字選擇不正確或句子使用錯誤等問題提供進一步解釋說明。

### 三、可行性分析

國家考試欲採行 AI 輔助閱卷工作，可從技術面、實務面及社會面 3 個層面進行可行性評估，說明如下：

#### (一)技術面

在本部現行紙筆考試及既有線上閱卷架構下，接續發展 AI 自動評分機制，初步評估可行之基本系統流程示意如附圖，惟下列關鍵技術需特別納入考量範疇：

1. **光學字元辨識(Optical Character Recognition, OCR)**：透過試卷掃描與特徵辨識等 OCR 技術，可將手寫文字快速自動轉化為數位化文字，以供後續機器學習與評閱之用。然而因中文字符結構複雜程度遠高於歐美文字，造成特徵取樣與辨識難度提升，且該技術仍存在因字跡潦草等因素致無法辨識或辨識錯誤尚待克服，從心測中心實務運作結果可知，目前中文字辨識正確率僅達 9 成，距 100%辨識正確率仍有待努力。

2. **自然語言處理(Natural Language Processing, NLP)**：該技術主要藉由斷詞及字句等標記，再與事先建立之大量語料庫，進行文章中之句子、詞彙等語法及語義分析，以嘗試讓電腦理解人類語言寫作之能力，惟目前該技術僅止於辨別文章之字詞、語法、結構等語言使用質量，尚無法讓電腦完全了解文章之真正意涵。

## (二)實務面

發展 AI 輔助閱卷過程中，各項資訊技術如同骨骼框架之串接，而實務作業則如同肌肉之生成，其考慮要素如下：

1. **國家考試種類與規模**：國家考試每年舉行約有 1,000 個類科、3,000 個應試科目(108 年各項國家考試類科、科目數量統計如附表)，其科目屬性含括文、法、商、理、工、醫、農、警政等不同領域，與現行國際間 AI 輔助閱卷僅應用於語文寫作範疇差異極大；再者，國家考試部分類科又因應考人數較少，無法取得足夠有效樣卷供機器進行學習，將難以發揮 AI 最佳功效。
2. **跨領域協同合作**：採行 AI 輔助閱卷涉及領域相當廣泛，包含資訊科學專家負責提供 AI、ML、OCR、NLP 等專業知識與實務經驗，以進行系統實作；語言學專家負責建立中文語詞知識庫，以作為機器學習、了解中文語法及語義之重要依據；測驗統計專家運用決策樹、相似度、統計分析等原理，設計可信賴之測驗評量演算法；各學科領域專家針對特殊專業領域，除協助建立學科專有名詞、慣用術語、核心概念等資訊外，並提供各題目之評量項目、參考答案等內容，以作為後續電腦評分之依據。以上皆需經歷大量資料蒐集、反覆調校及長期驗證等過程，短期內難見成效，以 Write & Improve 系統為例，經 20 多年分析方累積 3,000 萬個帶有錯誤標記及解釋之英文語料庫，其工程繁複且時程漫長。

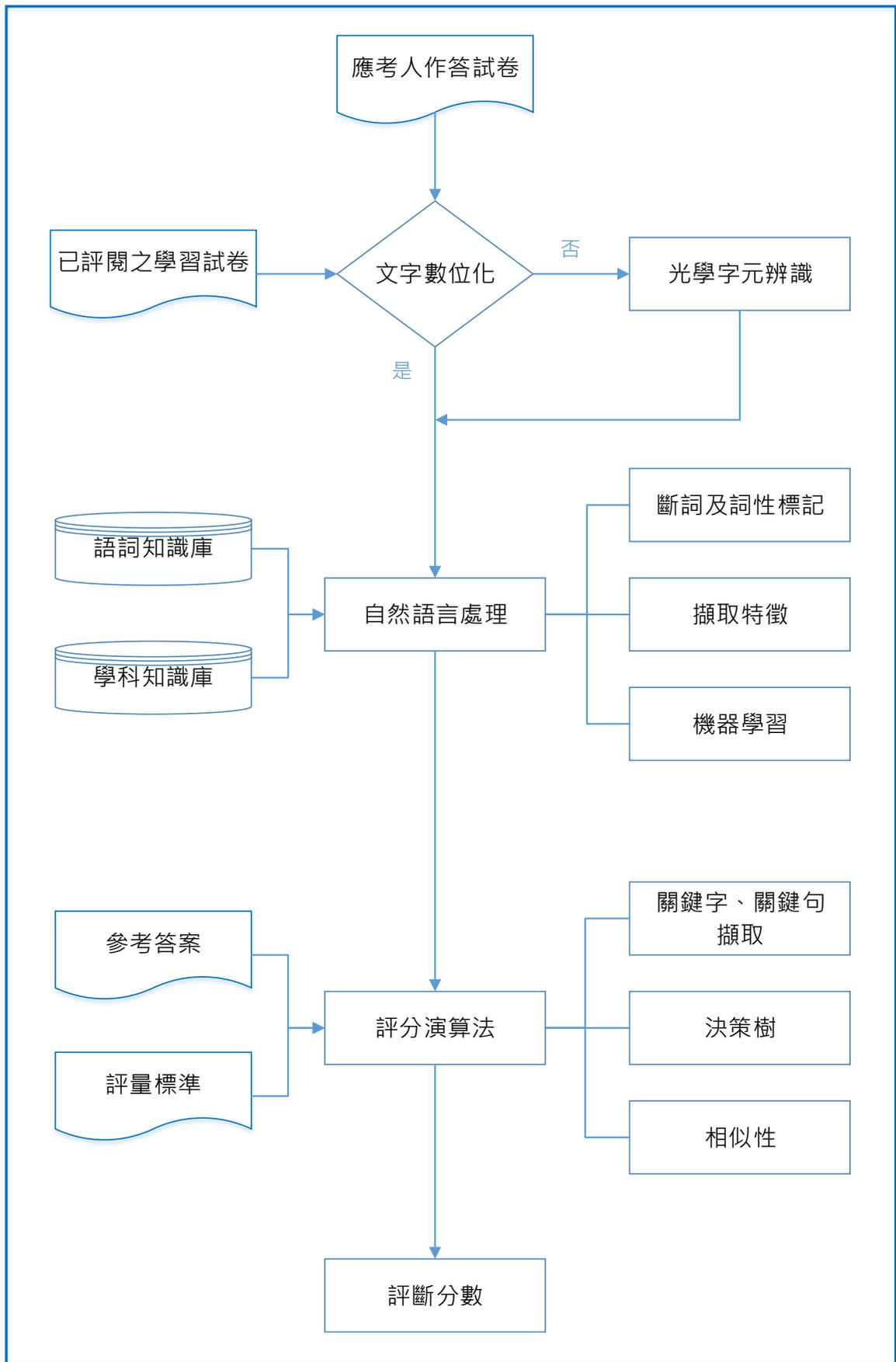
### (三)社會面

國家考試分屬公務人員任用資格與專技人員執業資格考試，就應考人而言，除競爭激烈外，尚攸關其得否執業之重大權益；就本部而言，係為國家拔擢優秀人才，並保障國民生命財產權益，故須以極其嚴謹態度面對評分工作與結果。從國內外案例研究發現，因其評分方式均採級分制(如國中教育會考中文寫作測驗共分 0 至 6 級分)，系統可容許之誤差範圍較大，而國家考試係採百分制(即 0 至 100 分)，能力刻度劃分較為細緻，可容許之誤差相對較小；再者，我國國情對電腦評分的信心因素，也深刻影響應考人或國人對電腦評分之接受程度。

### 四、結語

現行 AI 技術雖具重大進展，國內外亦有 AI 輔助閱卷相關研究與應用案例，惟仍難以精準處理機器評分作業，且國家考試應試科目種類繁多及部分類科應考人數過少，其考試結果又關乎應考人重大權益，爰未來國家考試發展 AI 輔助閱卷，藉助 AI 初步評鑑作答之完熟度，提供專業閱卷委員參考之可行性，尚富有高度挑戰性。

本部將廣續關注國內外資通科技最新動態發展趨向，並善用大專校院在該領域之研究創意與整合能量，適時採委託研究案方式尋求學界技術合作，以探測 AI 運用於國家考試之最適方案。初期可先嘗試結合線上命題之推動，藉以蒐集各領域專業科目核心關鍵詞，經由長期機器學習及資料模型調校，以作為試題相似度比對與試題難易度預測之用；未來視申論式試題線上作答推動情形，再適時導入 AI 閱卷輔助作業，透過電腦初步評鑑試卷作答內容之完整性與周密度，以提供閱卷委員參考，俾提升國家考選制度之信效度。



國家考試採 AI 自動評分系統流程示意圖

附表

108 年各項國家考試類科、科目數量統計表

序號	考試代碼	考試名稱	類科數	申論	混合	測驗
1	108010	108 年公務人員初等考試	15			29
2	108020	108 年第一次專門職業及技術人員高等考試中醫師考試分階段考試、營養師、心理師、護理師、社會工作師考試	7		26	11
3	108030	108 年第一次專門職業及技術人員高等考試醫師牙醫師藥師考試分階段考試、醫事檢驗師、醫事放射師、物理治療師、職能治療師、呼吸治療師、獸醫師考試	12			54
4	108040	108 年專門職業及技術人員普通考試導遊人員、領隊人員考試	31			25
5	108050	108 年公務人員特種考試關務人員考試、108 年公務人員特種考試身心障礙人員考試、108 年國軍上校以上軍官轉任公務人員考試	56	123	16	28
6	108060	108 年專門職業及技術人員高等考試大地工程技師考試分階段考試、驗船師、引水人、第一次食品技師考試、高等暨普通考試消防設備人員考試、普通考試地政士、專責報關人員、保險代理人保險經紀人及保險公證人考試、108 年專門職業及技術人員特種考試驗光人員考試	24	45	27	15
7	108070	108 年公務人員特種考試警察人員考試 108 年公務人員特種考試一般警察人員考試 108 年特種考試交通事業鐵路人員考試 108 年特種考試退除役軍人轉任公務人員考試	62	135	36	50
8	108080	108 年第二次專門職業及技術人員高等考試醫師考試分階段考試(第二階段考試)	1			4
9	108090	108 年公務人員高等考試三級考試暨普通考試	167	618	36	12
10	108100	108 年第二次專門職業及技術人員高等考試醫師考試分階段考試(第一階段考試)、牙醫師藥師考試分階段考試、醫事檢驗師、醫事放射師、物理治療師、職能治療師、呼吸治療師、獸醫師考試、108 年專門職業及技術人員高等考試助產師考試	12			55
11	108110	108 年第二次專門職業及技術人員高等考試中醫師考試分階段考試、營養師、心理師、護理師、社會工作師考試、108 年專門職業及技術人員高等考試法醫師、語言治療師、聽力師、牙體技術師考試、高等暨普通考試驗光人員考試	13	3	28	37
12	108120	108 年公務人員特種考試司法官考試(第一試)、108 年專門職業及技術人員高等考試律師考試(第一試) 108 年公務人員特種考試司法官考試(第二試)、108 年專門職業及技術人員高等考試律師考試(第二試)	3	12		7
13	108130	108 年公務人員特種考試司法人員考試 108 年公務人員特種考試法務部調查局調查人員考試 108 年公務人員特種考試國家安全局國家安全情報人員考試 108 年公務人員特種考試海岸巡防人員考試 108 年公務人員特種考試移民行政人員考試	49	126	21	20
14	108140	108 年專門職業及技術人員高等考試會計師、不動產估價師、專利師考試	14	14	11	1
15	108150	108 年公務人員特種考試外交領事人員及外交行政人員考試 108 年公務人員特種考試民航人員考試 108 年公務人員特種考試稅務人員考試 108 年公務人員特種考試原住民族考試	60	89	53	22
16	108160	108 年公務人員高等考試一級暨二級考試	29	117	1	
17	108170	108 年公務人員升官等考試 108 年關務人員升官等考試 108 年交通事業郵政人員升資考試 108 年交通事業公路人員升資考試 108 年交通事業港務人員升資考試	277	408	13	16
18	108180	108 年專門職業及技術人員高等考試建築師、25 類科技師(含第二次食品技師)考試暨普通考試不動產經紀人、記帳士考試	35	154	7	4
19	108190	108 年特種考試地方政府公務人員考試	113	382	34	36
合 計			980	2,226	309	426