

出國報告（出國類別：考察）

日本電腦化測驗與性向測驗 發展現況與趨勢考察報告

服務機關： 考選部

姓名職稱： 鄭中平政務次長、楊淑如科長、
歐陽安琪科長

派赴國家： 日本

出國期間： 113年9月16日至113年9月20日

報告日期： 113年12月2日

摘要

本次考察行程共計拜會5個機關，包含1個中央政府機關（人事院）、4個民間公司（CBT-Solutions 公司、e-communications 公司、Prometric Japan 公司、Recruit Management Solutions 公司），考察目的設定為了解所參訪機構使用之考試方法與經驗，以供考選部後續政策研議參考。

本報告包含電腦化測驗與性向測驗兩大主題，電腦化測驗就日本當地運用資訊科技推動電腦化測驗之各類應試方式加以介紹，且就運用人工智慧（Artificial Intelligence, AI）提供監考功能之趨勢略以描述；性向測驗則是針對日本公私部門對於融入性向測驗內涵之基本能力測驗，就其評量設計之思考脈絡、融入何種性向測驗向度、命題者的產生、試題開發和維護，以及對於新興科技運用的情形進行介紹；最終針對本次考察設定參訪的重點與問題，參酌受訪機關網站所蒐集之資料，輔以考察過程獲得之資訊加以論述，進而提出考察心得與建議。

關鍵詞:電腦化測驗、性向測驗、基本能力測驗

目錄

第一章	問題意識與考察緣起.....	6
壹、	電腦化測驗.....	6
貳、	融入性向測驗內涵之基本能力測驗.....	6
第二章	受訪機關與行程	8
第三章	日本電腦化測驗推動現況與發展趨勢	9
壹、	用詞定義.....	9
一、	施測地點.....	9
二、	施測時間.....	9
三、	監試方式.....	9
四、	國家考試電腦化測驗對照日本之用詞.....	10
貳、	CBT-Solutions 公司考試服務.....	11
一、	受訪機關概述.....	11
二、	電腦化測驗考試類型.....	11
三、	應試系統與試題類型說明.....	15
四、	電腦化測驗發展趨勢.....	16
參、	e-communications 公司資訊服務.....	16
一、	受訪機關概述.....	16
二、	考試系統軟體與功能說明.....	16
三、	考試類型概述.....	22
四、	其他電子化服務與趨勢.....	24
肆、	Prometric Japan 公司考試服務	25
一、	受訪機關概述.....	25
二、	電腦化測驗考試類型與演進.....	25
三、	考試會場與報到流程概述.....	26
四、	人工智慧在考試服務之應用.....	28
五、	進階考試應用之建議.....	29
伍、	小結.....	29
一、	電腦化測驗應用於多階段徵才的第一階段考試.....	29
二、	應用電腦化測驗的考試多為資格考試，競爭性較低.....	29
三、	尚無辦理一次式同時於固定會場之大型考試案例.....	30
四、	推動 BYOD 須因應公平性議題.....	30
五、	人工智慧輔助命題仍需專人驗證正確性.....	31
第四章	融入性向測驗內涵之基本能力測驗	32
壹、	日本國家公務人員考試基礎能力測驗.....	32
一、	基礎能力測驗概述.....	32

二、施測內容.....	33
三、命題作業及相關問題.....	35
四、關於施測結果的討論.....	35
五、性格測驗的運用.....	37
六、新興科技的導入.....	37
貳、私部門的基本能力測驗—SPI3的介紹.....	37
一、SPI3能力測驗概述.....	37
二、施測內容.....	38
三、施測的方式與題庫管理.....	39
四、命題開發及維運.....	40
五、關於施測結果的討論.....	41
六、地方自治體招募人才採用 SPI3的施測結果.....	42
七、性格測驗的運用.....	42
八、過去建立 SPI3的契機與未來發展及新興科技的導入.....	43
參、小結.....	44
一、日本公私部門招募員工的特色.....	44
二、對我國國家考試的啟發.....	46
三、未來國際合作的可能性.....	46
第五章 考察心得與建議	47
壹、電腦化測驗.....	47
一、宜針對現行電腦化測驗的需求變動擬定對策.....	47
(一) 考選部一次式考試與日本的連續型考試之對比.....	47
(二) 日本推動策略與經驗無法直接擷取或全數採用.....	47
(三) 應試系統與架構相容之前題下得著手研議與他考試機構聯盟..	48
(四) 具質優量多試題與國人觀念改變之前提下得推動連續型考試..	48
二、人工智慧用於國家考試.....	48
(一) 人工智慧監試無立即引用之需求.....	48
(二) 若需人工智慧輔助命題，允宜建立自有或共有大型語言模型..	48
三、日本律師、檢察官等公務人員考試採電腦化測驗值得關注.....	49
貳、融入性向測驗內涵之基本能力測驗.....	49
一、心理測驗之使用.....	49
(一) 多元評量是世界趨勢.....	49
(二) 應辨識公務人員職場的基本能力向度.....	49
(三) 性格測驗不計分，係作為面試的參考資料.....	50
二、心理測驗編制團隊.....	50
(一) 基本能力測驗命題技術須要長期培訓.....	50
(二) 專業團隊維運確保試題品質與考後不公開的必要性.....	50

附錄 1【考察計畫】	52
附錄 2【考察議題訪綱】	54
附錄 3【考察行程表】	68
附錄 4【考察座談紀錄】	71
附錄 5【參訪行程剪影】	96

表次

表 1：考選部電腦化測驗與3家公司考試類型之對應表.....	10
表 2：CBT-Solutions 公司4種考試比較表.....	15
表 3：e-communications 公司技術服務之4種考試比較表.....	24
表 4：日本國家公務人員考試各階段測驗種類、題數時間、配點比率及 內容一覽表-以2024年綜合職大學畢業程度考試為例	32
表 5：日本2024年起國家公務人員考試基礎能力測驗變動情形.....	34
表 6：SPI3能力測驗分類說明表.....	39

圖次

圖 1：CBT 考試服務運作架構	11
圖 2：OLTC 考試服務運作架構.....	12
圖 3：Remoty AI+考試服務運作架構	13
圖 4：IBT 考試服務運作架構.....	14
圖 5：MASTER CBT PLUS 運作示意圖	17
圖 6：MASTER CBT PLUS 軟硬體需求	17
圖 7：RemoteTesting 功能示意圖	18
圖 8：RemoteTesting 軟硬體需求.....	19
圖 9：ULTRA GUARDIAN 功能示意圖.....	20
圖 10：ULTRA GUARDIAN 軟硬體需求.....	21
圖 11：AI 監試+操作控制示意圖	22
圖 12：Prometric 會場配置圖	27
圖 13：Prometric 考試流程圖	28
圖 14：人事院命題作業流程圖.....	35
圖 15：SPI3試題開發及維護示意圖.....	40

第一章 問題意識與考察緣起

壹、電腦化測驗

考選部自93年起推動電腦化測驗，於國家考場自建電腦試場推動航海人員電腦化測驗考試，96年開始辦理分區電腦化測驗考試，於臺北、臺中、高雄等區域辦理牙醫師、助產師、職能治療師、呼吸治療師、獸醫師5個類科考試，試題類型皆為測驗式試題，並逐年擴增採行電腦化測驗之考試類科；此外，為擴大電腦化測驗應試服務地區及規模，99年推動國家考試電腦試場認證機制，與大專校院合作建置國家考試電腦試場，並於111年推動申論式試題純文字作答介面，迄113年10月，電腦化測驗業於臺北、新竹、臺中、臺南、高雄、花蓮區域辦理醫師、中醫師、牙醫師、藥師、獸醫師、醫事放射師、醫事檢驗師、呼吸治療師、物理治療師、職能治療師、助產師、諮商心理師、臨床心理師、營養師、護理師15個類科考試，服務應考約30,000人次/年。

為深化應考人服務，提供方便應試工具，考選部規劃逐步納入考試類科，並增加其辦理之考試次數。惟實務推動上應試座位擴建漸趨飽和，且考試期間電腦試場洽借若與大專校院教學期程相衝突，或將導致考試座位難以借用之窘境，爰規劃赴日本考察推動電腦化測驗之測驗機構，藉由學習日本電腦化測驗施測對象、規模、辦理考試方式及不同考試方式間之選擇，其系統架構、考試安全性與穩定性之規劃，引用新興科技施測技術（如 AI）在電腦化測驗之應用，以及電腦化測驗在人才招募及能力評鑑之發展現況與趨勢，俾因應民意對數位化考試之期望，且精準為國家揀選公務人力及專技人才。

貳、融入性向測驗內涵之基本能力測驗

鑑於性向測驗¹為有效之測驗技術且普遍為先進國家使用，藉以評量受試者未來發展的潛能，考選部擬將性向測驗納入公務人員考試制度規劃已久，除多次邀請學者專家表達專業見解外，邀集測驗評量領域專長之學者專家組成專案小組，適時提供考選部在發展公務人員性向測驗及其他測驗評量疑義之專業建議。更於108年委託辦理「公務人員性向測驗研發計畫」專案研究，歷時15個月，惟因樣本數的問題，該次研究僅具有建構效度，未來仍須有更多的實證資料以強化論述及作為決策基礎。

茲因歷史社會因素，我國公務人員考試考科設計時，基於公平的考量，著重專業科目為評量的主軸，基於此，在推行性向測驗之前，必須對於常見的問題，先予進行思考：

- （一）施行性向測驗是否有助於測得未來發展的潛力，提升甄選的效度；
- （二）透過何種方式取得外界理解與信任；

¹ 目前通用性的普通性向測驗，包括語文推理、數量推理、抽象/視覺推理、短期記憶等，測量的主要功能著重於廣泛的知識、判斷力、專心、抽象思考、立即的回憶、字詞理解的水準、細節的機敏力、非語文空間的察覺力與整體關係的分析力。

（三）不同制度架構下的考試，考試元素如何援引？

作為國家人力資源管理部門的一環，因應多元評量的發展，考選部近年來曾分赴世界各國考察公務人員考試制度，前分別於106年及112年赴日本考察，了解到無論中央人事院或地方自治體舉辦的公務人員考試，第一試多進行基礎能力測驗或教養測驗（統稱為基本能力測驗）。通過第一試後，在中央，要成為綜合職與一般職的國家公務人員仍需申請「官廳訪問」，在地方，要成為公務人員須經多次面試。而106年的考察重點在於制度規劃面的了解，112年赴日考察的重點在測驗編製組織設置與程序、面試的「執行面」，以及其面臨少子女化及高齡化問題，政府部門招募新人有多樣化的招募活動及其效益。

本次考察的重點，係針對日本公私部門對於求職者施測所謂之融入性向測驗內涵之「基本能力測驗」，基於職場的工作發展需要，討論其評量設計的思考脈絡、選擇融入何種性向測驗向度、命題者的產生、試題開發和維護、施測結果的討論，以及對於新興科技運用的情形和命題編製面臨的問題。這些議題在在對應國內對於性向測驗的疑慮，期透過本次考察的探訪過程，對於性向測驗應用於公部門初任人員招募考試階段所能發揮的功能，有更深入的了解及論理，以作為後續相關政策決定及試題開發技術的參考。

第二章 受訪機關與行程

本次赴日考察包含電腦化測驗及性向測驗兩大重點（如附錄1），擬考察機關與行程，原規劃參訪4間機關（構）、各安排1天參訪行程，其中電腦化測驗部分規劃自2間考試管理服務公司（CBT-Solutions 公司及 Prometric Japan 公司）及1間電腦化測驗技術服務公司（e-communications 公司）選擇2間；性向測驗依據擬考察目的之需求，依序自人事測驗研究中心、東京都廳、Recruit Management Solutions 公司（以下簡稱 Recruit-MS 公司）及人事院選擇2間，並預先擬定參訪議題，翻譯成日文後（如附錄2），請駐日本代表處遞交給擬參訪機關說明拜會目的與內容，並協請提供受訪機會。

經多方聯繫，人事測驗研究中心說明9月份皆有舉辦大型考試，東京都廳則有其他公務無法接受參訪，且受訪機關擬提供拜會之時間多僅1.5小時至2小時，爰本次考察行程前後5日，共計拜會5個機關（構），包含1個中央政府機關（人事院）、4個民間公司（CBT-Solutions 公司、e-communications 公司、Prometric Japan 公司、Recruit-MS 公司），考察行程表如附錄3、各參訪機關與公司的訪談紀錄如附錄4。

本次考察，全程幸獲駐日本代表處鼎力協助，費心聯絡參訪機構，百忙中全程支援本次考察行程，本團所有團員深刻地感受外交人員的辛勞，並衷心感謝在此參訪過程中代表處李逸洋代表、蔡明耀政務副代表、何仲民組長、黃冠超組長、蕭瑤秘書及戴如君小姐所提供之協助，再次表達謝意；亦感謝受訪機構誠摯接待，不吝分享相關經驗。

第三章 日本電腦化測驗推動現況與發展趨勢

壹、用詞定義

鑑於此次考察日本之3個公司對於電腦化測驗考試類型與監試表述不盡相同，為統一表達之用詞，先就各用詞加以說明。

一、施測地點

- (一) 固定會場：日本稱為会場型（英文對應為 test center）、考選部稱為電腦試場。會場設置有一定要求與規範，且有專人管理，考試嚴謹度較高。採此種考試方式受限會場總座位數，會場可能包含直營與借用外部電腦教室。
- (二) 臨時會場：指在公司（日本稱為社內考試）、學校或租用臨時場地辦理考試。
- (三) 在家考試：指在家考試（英文對應為 in house）。

二、施測時間

- (一) 一次式同時考試：指考試區間內僅辦理一次考試，有應考人皆須同時間參與之考試。
- (二) 連續型考試：同一考試分散於區間（全年，或一星期或數星期間），該區間內，應考人可以自由選擇考試時間。每位應考人試題不完全相同。試題選題方式或採難易度相同之隨機抽題，或採電腦適性化測驗（Computerized Adaptive Test,CAT）方式。

三、監試方式

- (一) 現場專人監試：應用於固定會場或臨時會場的考試，監試人員受有專業訓練及驗證，並須定期接受再訓練或驗證。會場也可能配置攝影機錄製考試現況，亦可搭配即時現場監試或考後調閱監視畫面。此種監試方式考試嚴謹度及安全性較高。
- (二) 遠端專人監試：應試電腦須配置攝影機，考試期間透過網際網路由專人在遠端監試，此種監試方式考試嚴謹度及安全性居中。
- (三) 人工智慧輔助監試：應試電腦須配置攝影機，考試期間全程錄影，且透過網際網路將錄影結果傳送到雲端，考後透過人工智慧（Artificial Intelligence, AI）技術分析判斷作弊之案例。
- (四) 無監試需求：整個過程沒有監試，多用於模擬考、數位學習之成果考試。

四、國家考試電腦化測驗對照日本之用詞

迄113年10月，考選部一年辦理5次國家考試電腦化測驗，每一次考試所有應考人皆在考選部認證合格電腦試場同時參加考試，考試當下有監場人員現場監試；為比較臺灣與日本電腦化測驗之異同，除定義上開統一表達之用詞外，並統整臺灣國家考試電腦化測驗之考試類別為：**固定會場、一次式同時、現場專人監試**之考試。

表 1：考選部電腦化測驗與3家公司考試類型之對應表

公司	考試類型			
CBT-Solutions	IBT*	Remoty AI+	OLTC Online Test Center	CBT**
e-communications	IBT	IBT+AI 監試 +操作控制	特殊會場型 電腦化測驗	會場型** 電腦化測驗
Prometric Japan	IBT	IBT+AI 監試 +操作控制		CBT**

* IBT 為 Internet Based Test、CBT 為 Computer Based Test 之縮寫。

** 與考選部主辦的電腦化測驗同屬於會場型，惟在日本的會場型考試屬連續型考試，但考選部為一次式同時之考試。

貳、CBT-Solutions 公司考試服務

一、受訪機關概述

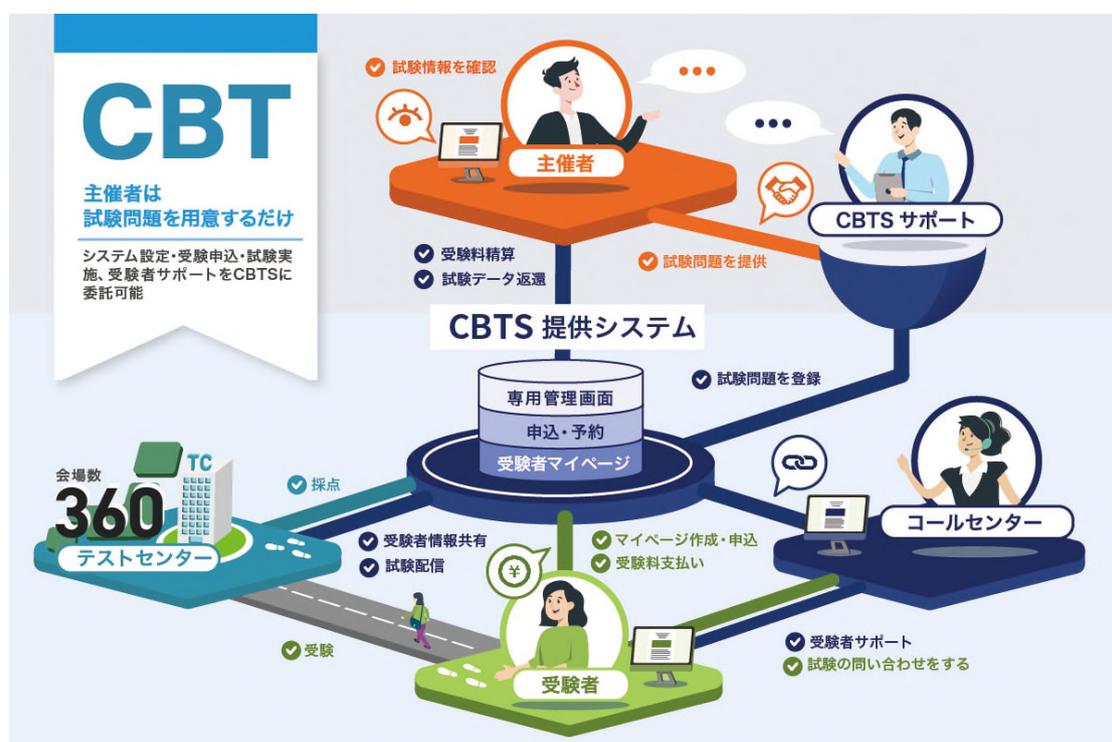
成立於2009年，員工數約120人，屬於考試管理服務公司，支援各階段試務工作，業務內容包括應考人報名（含繳費核銷）、個資管理、成績評閱、電腦化測驗考試（系統與監考）、試題分析、印發證書等；在日本有340個考試會場，約6,000席座位數，一年辦理150萬人次考試

二、電腦化測驗考試類型

（一）4種考試服務說明

1.CBT（Computer Based Test 縮寫）

圖 1：CBT 考試服務運作架構



資料來源：引自受訪機關網站資料（2024）

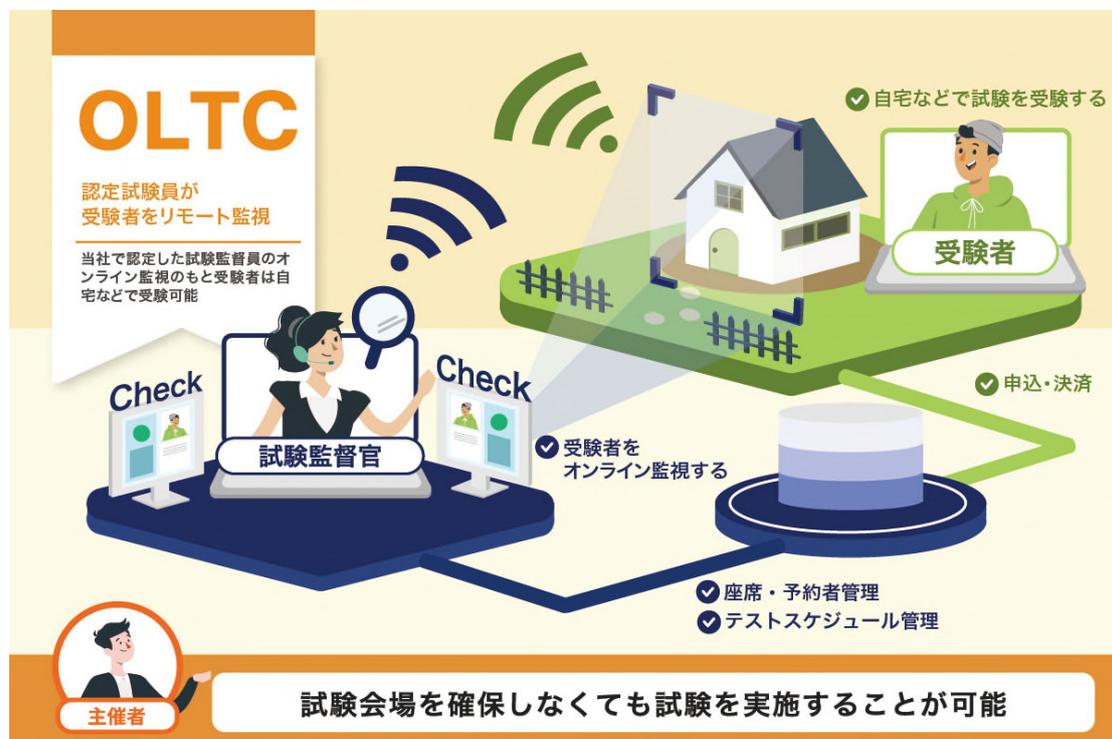
固定會場型，可為一次式同時或連續型考試，應考人須於報名時選擇應試會場（若屬連續型考試，尚須選擇考試日期），應試會場有監試人員；應考人須提前抵達會場並辦理報到手續，報到時檢查身分證明文件、確認應考人身分後，應考人須將非考試用品擺放於置物櫃中，不能攜帶入場，應考人使用該會場的應試電腦應試。適用於嚴謹度、安全性要求較高之國家等級認證考試。

此種考試應用於資格考試（含檢定考試）、企業內部考試（社內應用）及就職進用（適性考試）3大類，其應用比率依序為70%、20%、10%。若以考試時間區分，全年皆可考試約50%、期間限定約30%、一次式同時考試約20%（多為

醫學學會委託之專科資格認定考試，同時間最多1,000位應考人，平均約幾百位應考人)。

2.OLTC (OnLine Test Center 縮寫)

圖 2：OLTC 考試服務運作架構



資料來源：引自受訪機關網站資料 (2024)

非固定會場型，可於臨時會場或在家考試，可為一次式同時或連續型考試，若屬連續型考試，應考人於報名時須選擇考試日期；應試前應考人透過自備電腦攝影機讓遠端監試人員檢查身分證明文件及環境週遭沒有可疑物件後，即可開始考試，並利用該電腦攝影機及應試畫面由監試人員進行遠端監試。此考試方式結合方便性及嚴謹性，整合 IBT 考試技術及便利性，兼備 CBT 考試環境及遠端專人監試之嚴峻度。

線上監試1位監試官負責8位應考人，可放大特定應考人的監試影像，並且監試應考人電腦螢幕操作情況 (如:使用 google 搜尋引擎查閱資料)，此種考試運用較少，多數屬於公司內部升遷升職等考試。

4. IBT (Internet Based Test)

圖 4：IBT 考試服務運作架構



資料來源：引自受訪機關網站資料（2024）

非固定會場型，可於臨時會場或在家通過網際網路考試，可為一次式同時或連續型考試，沒有身分驗證及監試作業，安全性最低，便利性最高，在考試區間內可隨時透過網際網路，提供無時空限制的便利考試，並支援個人電腦、智慧型手機及平板電腦，適用於數位學習及模擬考試。

(二) 4 種考試綜整比較

表 2：CBT-Solutions 公司4種考試比較表

項目	IBT	Remoty AI+	OLTC Online Test Center	CBT
嚴謹度	30分△	60分△	80分○	100分○
監試	無	運用 AI 監試	運用攝影機及應試畫面監試	現場監試
考試環境	不須事先確認	不須事先確認，若 AI 技術偵測到作弊，考後將會有環境檢查程序	檢查可疑物件	須將物品擺放於置物櫃中，不能攜入場
身分驗證	考試當日無驗證	透過相片註冊，並採 AI 技術驗證	採用攝影機驗證身分證明文件	現場檢查身分證明文件
便利性	100分○	80分○	70分△	50分△
地點	居家、學校、辦公室等			特定試場
應試工具	電腦、手機、平板	電腦	電腦	設置於試場之電腦
日期	考試期間內無設限		須先預約考試時間	須預約地點(時間)

資料來源：引自受訪機關網站資料及本報告綜合整理（2024）

三、應試系統與試題類型說明

就連續型電腦化測驗考試，應考人試題不盡相同，其試題公平性係運用試題反映理論（Item Response Theory,IRT）來實作，須具備大量優質試題資料庫及識別度。部分考試採用電腦適性化測驗（Computerized Adaptive Test,CAT）方式來考試。

受訪機關說明試題類型中，系統可以支援圖片與動畫（如醫事考試心臟跳動），若試題需搭配語音時，則應試環境須特別處理聲音撥放及不干擾其他應考人之設計（如逐台啟用聲音撥放，配置及檢查耳機撥放效果等）。

試題類型主要為測驗式試題（單選題及複選題），並有申論式試題，採純文字輸入方式，日文僅一種輸入法，透過發音輸入，轉換成英文、平假名、片假名，並沒有臺灣多元輸入、付費輸入法等問題，推動上相對應阻力較少。

應試系統在考試當下區分為連網與不連網兩種，不連網應試系統於應考人連線登入、並下載試題後，即可離線考試，考試結束後，仍需連上網路上傳作答結果，俟上傳完竣後即刪除硬碟中之應試作答結果；若遇硬碟損毀無法修復，該次應考人作答資料確實無法處理，此時，則提供應考人再次參加考試的機會。至於考試結束後無法連網，因作答結果儲存在硬碟中，可俟網路穩定後再上傳。

四、電腦化測驗發展趨勢

日本電腦化測驗的推動以2020年新冠疫情為分水嶺，2020前應用者多為民間公司，2020後國家考試也開始採用，主要是因為疫情的影響，比較難採用一次式同時考試（紙筆測驗或電腦化測驗），因此日本國家考試也慢慢開始採用連續型電腦化測驗。

日本地方政府公務人員開始採行電腦化測驗，如東京都廳採一週期間限定方式辦理，應考人數約數千人，每次考試2小時，至於中央機關國家公務人員考試仍採用紙筆測驗。近期日本法務省也開始規劃律師、檢察官、司法官公務人員考試於2026年採實驗性質方式辦理電腦化測驗，規劃與大專校院合作運用其既有電腦資源，辦理約10,000-14,000人試行電腦化測驗。

受訪機關取締役（兼人事部部長、兼日本の資格・検定事業部部長）佐々木先生提及：連續型電腦化測驗考試應考人試題不盡相同，應考人難以作弊；紙筆考試屬於形式上公平性，監場人員仍可以選擇性監試，未必達到真正公平。至於電腦化測驗運用資訊科技防弊功效是客觀、公平的。此外，連續型電腦化測驗考試可提供隨時考試的機會，以人才培育角度觀之，將為國家社會培育更多人才；佐々木先生在回應我們的信件中，特別提及：「不應該因為形式上的問題而剝奪年輕人挑戰的機會。隨時可參加的 CBT 能減輕考試的負擔，並創造多次挑戰的機會」。

參、e-communications 公司資訊服務

一、受訪機關概述

成立於2000年，屬電腦化測驗與數位學習技術服務公司，電腦化測驗之業務始於開發多益應試系統之委託，該多益考試電腦化測驗於2004年上線；公司目前負責電腦化測驗之技術人員約6人；公司提供各類資格/認證考試解決方案與系統設計，以及電子學習服務；在日本有190個考試會場，每個會場座位數約5-20席。

二、考試系統軟體與功能說明

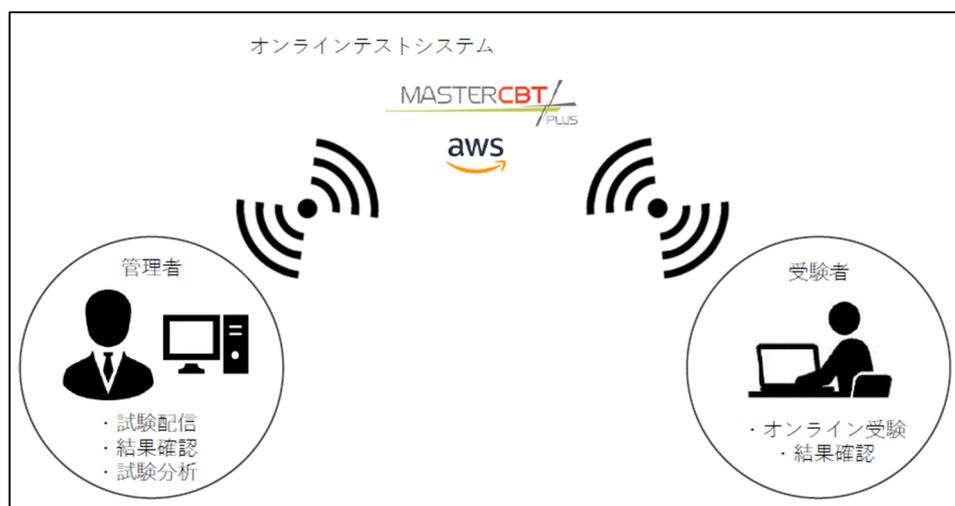
受訪機關提供之電腦化測驗技術服務如 MASTER CBT PLUS（網頁版線上應試系統）、RemoteTesting（遠端監控系統，可進行較嚴格的線上考試）、ULTRA GUARDIAN（控制操作/行為之應用程式）等，可提供其他考試服務管理者依據需求搭配組合成客製化考試。

（一）MASTER CBT PLUS

MASTER CBT PLUS（網頁版線上應試系統）透過網際網路及瀏覽器辦理線上考試，藉由備份及復原機制達成考試正確性與安全性。自2004年推出後，迄今每年約已施測（delivery）100萬次，且已被1,000個主辦考試機關（委託者）採

用。其特色如下：

圖 5：MASTER CBT PLUS 運作示意圖



資料來源：引自受訪機關網站資料（2024）

圖 6：MASTER CBT PLUS 軟硬體需求

MASTER CBT PLUS 動作環境		MASTER CBT PLUS
	受験者サイト	管理者サイト
OS	Windows : Windows10/Windows11 macOS : macOS 10.12以上 Android : Android9.0以上 iOS : iOS12以上 iPadOS : iPadOS13以上	Windows : Windows10/Windows11 macOS : macOS 10.12以上
ブラウザ ※1	Windows : Microsoft Edge、Google Chrome、Firefox ※2 macOS : Apple Safari、Google Chrome Android : Google Chrome iOS : Apple Safari iPadOS : Apple Safari	Windows : Microsoft Edge、Google Chrome macOS : Apple Safari、Google Chrome
その他の条件	<ul style="list-style-type: none"> JavaScriptが動作する設定となっていること Cookieが使用できる設定となっていること 認証証、スコアレポートなどのオプションを利用している場合は、Adobe® Acrobat Reader DC が端末にインストールされていること ブラウザのローカルストレージが使用可能になっていること 利用端末の時計が正確な日時に調整されていること 通信経路上の各機器設定に関して、以下を許可している <ul style="list-style-type: none"> ① 当社サーバーとの通信・非同期通信を許可している ② HTTPプロトコルのGet/Postメソッドを許可している ③ JSON形式によるデータの送受信を許可している 	

※1 ブラウザは最新バージョンをご利用ください。最新バージョンのご利用が難しい場合においては、最大6ヶ月前までにリリースされたバージョンをご利用ください。
 ※2 動作はしますが非推奨ブラウザとなります。

資料來源：引自受訪機關網站資料（2024）

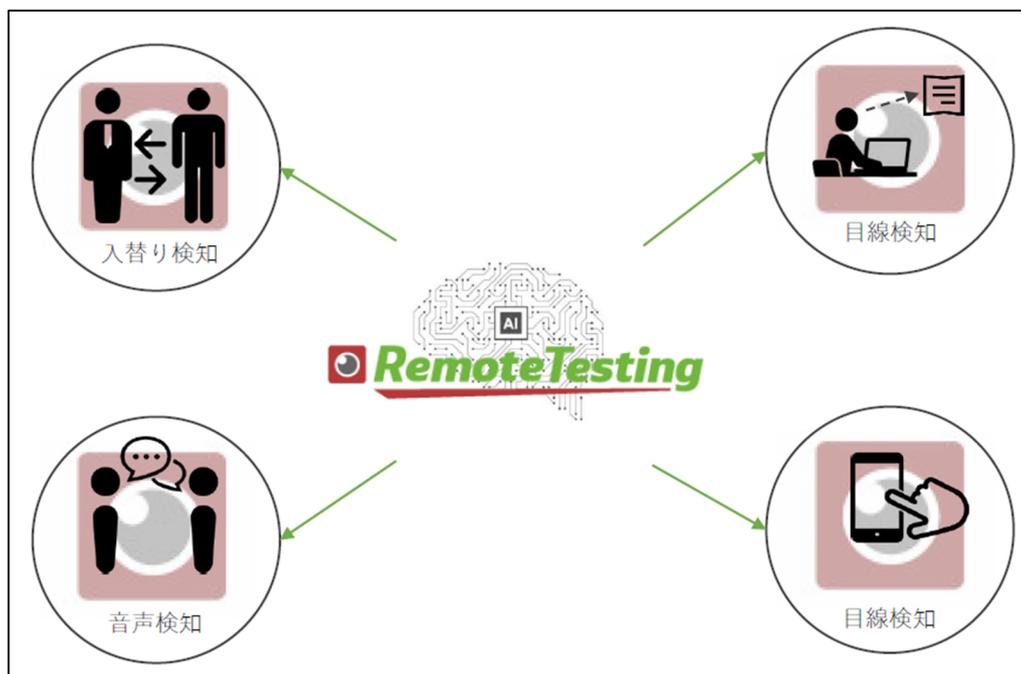
1. IBT（Internet Based Testing）考試架構：屬網際網路型網頁版（browser-based）應試介面，可支援手機、平板電腦及個人電腦，透過網際網路進行，提供無時空限制考試。
2. 理論上同時間可提供 10,000 人考試：應試系統設計架構係當開始考試

時，會將試題下載至應試工具，作答內容最後上傳至主辦單位。系統採離線考試，理論上並不會有同時間應考人數之限制，但基於系統資源瓶頸，在開始考試下載試題與結束考試上傳作答內容時，最大極限為同時間支援 10,000 人，惟目前並無實際驗測或推動之數據來佐證。

3. 試題安全機制與類型：試題及作答結果皆加密，防止非法存取；考後自動刪除應試資料，防止資料外洩。試題支援單選、多選題，下拉式選單、填空、文字輸入（單字或文句）、影音試題。技術上，一次考試之試題大小應限縮於 1G Bytes 為佳。題目選取未採 CAT 方式辦理，而是題庫隨機抽取試題（運用 IRT 理論，同考試應考人試題不盡相同，各應考人試卷之試題難易度及題數都相同），考後未公布試題。
4. 因應偶發的復原機制：離線型考試作答結果儲存於應試電腦之硬碟，當電腦當機時，得以重啟應試電腦後繼續應試，但若硬碟損毀，當次考試則無法完成，需另案處理；惟若應考人應試電腦係屬隨時連線儲存作答資料狀態，則硬碟損毀亦能透過還原機制繼續考試，但前提是網路狀態須隨時連線。
5. 自動評閱：測驗式試題（單選、複選）可以自動評閱，簡單簡答題可以判別文字者也可自動評閱，但申論題打字之內容則無法即時評閱。

（二）RemoteTesting

圖 7：RemoteTesting 功能示意圖



資料來源：引自受訪機關網站資料（2024）

圖 8：RemoteTesting 軟硬體需求

Remote Testing 動作環境 RemoteTesting	
OS	Windows Windows10/Windows11 macOS Chrome利用：10.13.6以上 Safari利用10.15.7以上
ブラウザ	Google Chrome：(PC：80以上) Microsoft Edge：(PC：80以上)[macOS対応版は対象外] MacOS (macOS)：Safari (14.0.3以上)
その他の条件	<ul style="list-style-type: none"> ・Webカメラ：それぞれの最大フレームレート30fps以上 (AI認証を利用するにはオートフォーカス付きカメラが必要) ・Webカメラ：それぞれのOSに対応したカメラ ・マイク：ノートPCやWebカメラ付属のマイク、外付けマイク等 ・ブラウザ設定：JavaScriptが動作する設定となっていること ・ブラウザ設定：ブラウザのローカルストレージが使用可能となっていること ・ネットワーク：受講者一人当たり：4 Mbps以上のスループットを備えるインターネット回線 ・ネットワーク：通信経路上の各機器設定に関して、以下を許容していること <ul style="list-style-type: none"> ① 当社サーバー・ストレージサーバーとの通信・非同期通を許可している ② HTTPプロトコルのGet/Postメソッドを許可している ③ JSON形式によるデータの送受信を許可している ・その他：背景画像設定ソフトは使用不可 ・シンクライアント端末の場合、正常に受験できない場合がありますので、一般的なPCでの受験をお願いします。 ・企業内ネットワーク等、セキュリティ制限がかかっている場合には正常に受験できない場合がありますので、一般的なインターネット回線をご利用ください。

資料來源：引自受訪機關網站資料（2024）

鑑於網際網路改變工作及學習地點限制，近年因為新冠疫情致遠距作業蔚為趨勢，僅須透過網際網路及資訊設備，可無時空限制工作、學習、考試；然而遠距學習與考試在教考訓領域上卻引發防止非本人學習或應試、作弊、試題外洩等討論議題。有鑑於此，RemoteTesting（AI 監試）因應而生，於遠距考試及數位學習當下，經由遠端監控服務，使用攝影機記錄並保存應考人考試或學生參訓實況，再佐以 AI 分析所錄製的紀錄，據以研判是否有代考、作弊、非本人學習等影響考試與學習公平性作為。功能特色如下：

1. 記錄考試與學習之過程與狀況：攝影機記錄考試狀況與參加考試或學習情形（同時錄音與錄影），同時在應考人/學習者的螢幕上也會提供錄製的預覽畫面，有效嚇阻詐欺舞弊等行為。
2. 採用 AI 分析應考人及學生行為：使用 AI 引擎，分析所錄製的考試狀況及學生參加考試或學習情形（如應考人或學習者視線方向或臉部傾斜度、考試時各階段所錄製的應考人或學習者音訊等），判別是否有詐欺或舞弊行為（如冒名頂替、第三人參與考試或學習、應考人參閱考試相關文書、使用手機作弊）。此外也會分析所錄製的聲音，看看是否有人告知答案。
3. 整合於線上考試系統：與 MASTER CBT PLUS（網頁版線上應試系統）相容，能整合該考試系統達成安全性遠端考試；提供易於操作的網頁版介面，透過網際網路、不須額外安裝，使用瀏覽器即可操作應試系統；經登入、設備規格檢查及身分驗證過程，即開始考試與學習。

(三) ULTRA GUARDIAN

RemoteTesting (AI 監試) 得以適度防止冒名頂替、第三人參與考試或學習、應考人參閱考試相關文書、使用手機作弊，但就試題外洩、使用應試工具查閱考試相關資訊或利用資訊技術與工具遠端遙控應試電腦等舞弊行為，卻無法加以偵測與預防，因此，ULTRA GUARDIAN (超級守護者，Windows 安全控管) 因應而生，於應試電腦上啟用該系統，得更深化考試安全性，其安控功能如下：

1. 應試電腦軟體控制：應試畫面以全螢幕方式於前景呈現，應試電腦無法啟動其他視窗或應用程式。
2. 按鍵操作控制：取消特殊功能按鍵之操作，如取消錄製或列印螢幕、取消啟動 Windows 選單、取消啟動瀏覽器、取消鍵盤操作等，防止舞弊或試題外洩。
3. 防止作弊：禁止使用螢幕分享軟體或多個顯示器連結，防堵分享考試螢幕。
4. 控制螢幕擷取軟體，防止應考人擷取螢幕內容致試題外洩。

圖 9：ULTRA GUARDIAN 功能示意圖



資料來源：引自受訪機關網站資料 (2024)

圖 10：ULTRA GUARDIAN 軟硬體需求

Ultra Guardian 動作環境

OS	Windows 10 Version 22H2 以降 または Windows 11 Version 22H2 以降 ※Enterprise LTSC（長期サービスチャネル）はご利用頂けません ※Versionの確認方法は こちら をご参照ください
アーキテクチャ	64ビット ※ビット数の確認方法は こちら をご参照ください
プロセッサ	インテル® Intel Core i5 / i7 / i9 または AMD® Ryzen 5 / 7 / 9 4コア8スレッド以上，基本クロック 3.0GHz以上
メモリ	8GB以上のRAM
空き容量	1GB以上の空き容量 ※対象の試験によって、異なる可能性があります
ディスプレイ	1280ピクセル × 720ピクセル 以上 ※OS等の設定により表示スケールが100%を超えている場合は、ディスプレイの物理値ではなく論理値を基準とします
カメラ	リモート監視ありの試験では必須
マイク	リモート監視ありの試験では必須
インターネット接続	リモート監視ありの場合 上り / 下り 4Mbps 以上 リモート監視なしの場合 上り / 下り 1Mbps 以上 ※対象の試験によって、異なる可能性があります
その他の条件	<ul style="list-style-type: none"> JavaScriptが動作する設定となっていること Cookieが使用できる設定となっていること 認定証、スコアレポートなどのオプションを利用している場合は、Adobe® Acrobat Reader DC が端末にインストールされていること ブラウザのローカルストレージが使用可能になっていること 利用端末の時計が正確な日時に調整されていること 通信経路上の各機器設定に関して、以下を許容していること <ol style="list-style-type: none"> ① 当社サーバーとの通信・非同期通信を許可している ② HTTPプロトコルのGet/Postメソッドを許可している ③ JSON形式によるデータの送受信を許可している インストール時は、Windows管理者ユーザーでのログインが必要

資料來源：引自受訪機關網站資料（2024）

三、考試類型概述

(一) 4 種考試服務說明

受訪機關分享經組合其提供之 MASTER CBT PLUS、RemoteTesting 及 ULTRA GUARDIAN 電腦化測驗技術服務，可提供之考試服務與應用如下：

1.IBT

IBT (Internet Based Testing) 應用 MASTER CBT PLUS 系統，採網際網路型網頁版 (browser-based) 應試介面，無場地限制，惟缺乏監試機制，嚴謹度與安全性低，但應試便利性高。常用於模擬考試，補習班或大學提供模擬考，可在家進行，部分企業需員工取得某種資格時，亦會要求員工於會議室中採用此種方法應試。模擬測驗採用 IBT，主要是因為模擬考僅是一種練習，非真正考試，不擔心是否作弊。

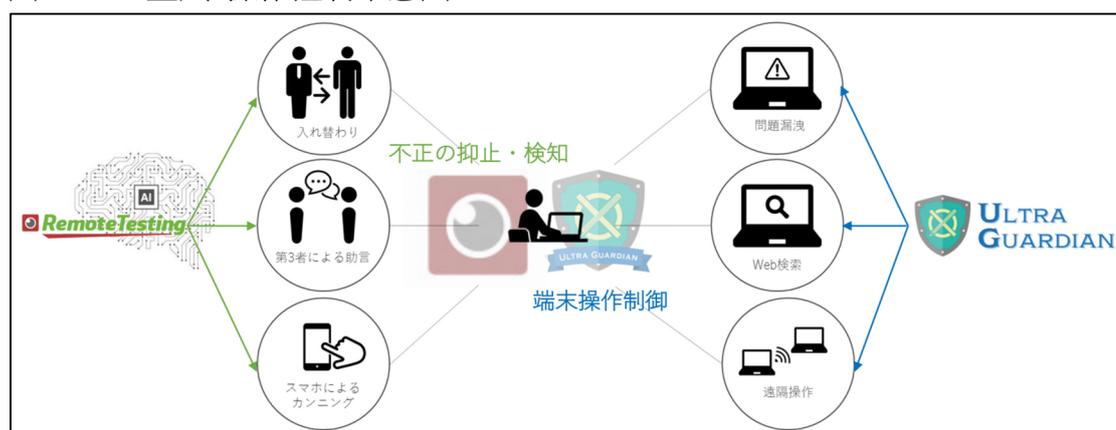
2.IBT+AI 監試+操作控制

應用 MASTER CBT PLUS、RemoteTesting 及 ULTRA GUARDIAN 服務，有 IBT 便利性，以及 AI 監試與 Windows 安全管控，提高考試嚴謹度與安全性。此種考試方式無場地限制，並適度佐以攝影機監試及作業系統安控，達防弊功能。

考試當下使用資訊設備攝影機錄音錄影，開考時要求應考人提供身分證明文件，並透過攝影機確認是否為本人應試，驗證身分後進入下一個考試階段；考試當下全程錄音錄影，畫面右邊是應試畫面，左邊是錄影畫面，應考人自己可以查閱被錄製的影像，考試結束後運用 AI 來判別是否作弊。

若 AI 發現有異常，受訪機關將擷錄異常畫面作成報告，並可依需求提供細部查閱錄影影像 (如應考人複誦題目，看看是否有人告知答案等)，報告將提供給主辦考試機關 (委託者)，由委託者自行判斷應考人是否作弊。

圖 11：AI 監試+操作控制示意圖



資料來源：引自受訪機關網站資料 (2024)

此種考試多運用於大學入學英文能力分班考試，或是企業升級考試，舉例來說多益校園團體測驗 (Institutional Program, IP) 即採此種應用，其測驗內容與

一般多益檢定相同，但不提供證書申請，成績結果僅限用於主辦考試機關（委託者）參考；至於分班與分級考試，分班考試後若進入不符合能力的班級，對受測者來說可能因為不適任而成為一種風險，或是分級後負責不適任之工作而更加辛苦，故施測者並不會想要作弊取得非能力等級之成績。

3.會場型電腦化測驗

就自辦的電腦化測驗考試管理服務部分，受訪機關2013年與日本通信紙株式會社、日本出版販売株式會社共組 J-Testing 考試管理服務。其服務內容包含提供應考人報名、會場設置、考試會場安排、電腦化測驗考試、考後試卷評閱、成績結果通知等營運支援。

J-testing 之會場係於認證過的學校或機構電腦教室來辦理考試，有專人擔任監試工作，一年施測之應考人次約15,000人次，多為連續型考試（期間限定，3天或一週等），較少辦理一次式同時之考試；J-testing 傾向提供民間、工作導向需求的資格考試，目前沒有承作國家級的資格考試，主因係因 J-testing 採題庫隨機選擇試題，而國家考試在意試題公平性，題庫須有大量高質量之試題，爰之。

4.特殊會場型電腦化測驗

特殊會場型係租借會議室或大型場地，由應考人自備電腦（Bring Your Own Device, BYOD）或租賃電腦來辦理同時間的考試，可搭配專人監試，另可使用 ULTRA GUARDIAN Windows 安全管控程式防止應考人啟動自備電腦之程式來作弊，嚴謹度較高。惟實務上受訪機關並未實際有採用 BYOD 為施測工具之案例。另外，受訪機關分享：在日本，基於公平性考量，基本上沒有太多考試採用 BYOD 此種施測工具與方式。就公平性議題，舉例來說，每個應考人的螢幕規格不同，在解析度、亮度上會影響圖片試題之解讀；如解析度不同，圖表與動畫呈現有差異，又如亮度會影像超音波圖片的解讀等問題。

(二) 4 種考試綜整比較

表 3：e-communications 公司技術服務之4種考試比較表

項目	IBT	IBT+AI 監試 +操作控制	會場型 電腦化測驗	特殊會場型 電腦化測驗
嚴謹度	30分△	50分~70分△	100分○	100分○
監試	無	運用 AI 監試 + Windows 安全控管	現場監試	運用 Windows 安全控管
考試系 統軟體 與功能	MASTER CBT /PLUS	RemoteTesting + ULTRA GUARDIAN	J-Testing	ULTRA GUARDIAN
便利性	100分○	100分○	70分△	40分△
地點	居家、會議 室、網咖等	居家、會議室、 網咖等	特定試場 全國考試中心	租用會議室、 大廳等
應試 工具	電腦、手 機、平板	電腦	設置於試場之電腦	電腦、平板 (BYOD 或租賃)
日期	考試期間內無設限		須預約地點 (時間)	須預約地點

資料來源：引自受訪機關提供資料及本報告綜合整理（2024）

四、其他電子化服務與趨勢

數位學習無時空限制，可透過電腦、平板及智慧型手機增加學習的機會；數位學習理論中認為在學習過程中，若得以適時就學習內容加以驗測學習成果（施予考試），將增加學習內容的長期記憶。爰此，受訪機關提供電子學習雲端平台（SAKU-SAKU）、防止詐欺性電子學習課程（Sakutes AIMONITOR）等服務。

(一) SAKU-SAKU

提供數位學習平台及線上應試雲端服務平台，教學端得以輕鬆上傳數位學習影片內容，設定禁止快轉、禁止調整播放倍速等控制；並查看全體學習進度、狀態及課程學習結果，且以 Excel 格式建立教學成果報告。

(二) Sakutes AIMONITOR

數位學習課程前驗證學習者身分，課程期間搭配數位學習雲端平台、線上應試及監試功能，教導學習內容並驗測學習成果，並防止學習與驗測過程舞弊與詐欺，藉以深化數位學習的效果。此服務無需安裝特殊應用程式，惟須備有瀏覽器及和網際網路連線，初次使用須以智慧型手機註冊個人資訊（臉部照片、身分證）；參加課程前須確認網際網路連線，並檢查數位學習之設備，且使用攝影機拍攝學習者臉部照片，與註冊資料進行 AI 比對；數位課程開始線上學習與驗測時，系統即時偵測學習者是否離席或有多人學習等情境（監視畫面亦同步顯示於學習者螢幕上），並採取必要控制措施；學習者可隨時查閱學習狀態，數位學習管理者亦可查閱課程整體學習成果。

肆、 Prometric Japan 公司考試服務

一、 受訪機關概述

Prometric 為跨國性考試管理服務公司，擁有30多年考試服務的經驗，員工數約2,400人，在12個國家計16個辦公室、8,000個會場，一年辦理700萬人次考試；Prometric Japan 公司為美國 Prometric 公司1991年於日本設立之子公司，員工數約169人，同 CBT-Solutions 公司支援各階段試務工作，此外，若客戶有命題需求，也提供相關專業支援；在日本有160個考試會場，約4,600席座位數，一年辦理250萬人次考試，日本子公司是除了美國之外，擁有最大市場的子公司，目前 Prometric Japan 公司已進入向國外發展的階段，而支援日本勞力引進的日語檢定考試（JFT Basic）已在日本全國各地施測。

Prometric 公司所提供的考試服務從應考人報名、選擇考試、繳費核銷、考試地點的安排、電腦化測驗考試（系統與監考）、成績評閱，以及試題分析，提供一系列完整的服務。就全球觀之，提供的指標性全球電腦考試包含財金類（如美國註冊公認會計師,Certified Public Accountant,CPA；財務規劃師, Chartered Financial Analyst,CFA），IT 認證類（如 Intel、Adobe 等認證），教育類（如美國大學委員會 College Board 所主辦的學術能力評估測試 Scholastic Assessment Test,SAT，該考試平台及試題皆由 Prometric 提供，命題已運用 AI 技術輔助命題）；語言類考試：如托福等；另 Prometric Japan 公司於日本提供之指標性客戶如 Recruit-MS 公司開發之 SPI3考試等。

受訪機關表示該公司通過 ISO 9001國際品質保證標準，且 ISMS 通過 ISO IEC 27001認證，透過標準作業程序，重複舉辦考試累積大量經驗，因此創造高水準考試服務品質；當然辦理過程中亦出現少量錯誤，然透過該錯誤調整標準程序、精進試務作業，精益求精累積更多經驗，進而強化更高水準的服務品質。

二、 電腦化測驗考試類型與演進

Prometric 有兩種考試類型：CBT（會場型）及 IBT（網際網路型），其定義的 CBT 是指會場型電腦化測驗，施測時間可為一次式同時考試及連續型考試。

受訪機關網站的專欄資料提及，電腦化測驗推動過程中，國際上高風險的資格考試、入學考試，因成績結果影響甚大，因此多採一次式同時間會場型考試；但隨著 IRT、CAT 等測驗理論演進，這些考試部分也發展為連續型考試，主要係因連續型考試不再受限應試座位數量，並且可以降低一次式考試發生偶發事件的風險，對於大量應考人的考試或解決異常狀況影響應試權益之擔憂，有相當的助益。

會場型考試有所限制，蓋因會場之設置仍侷限於供給需求概念，因此偏鄉地區因經濟效益考量，通常沒有設置會場，致偏遠地區應考人考試地點受限制，且可能因舟車勞頓影響考試結果；此外，2020年新冠疫情或天災的影響，觸發大眾注意到會場型考試之風險；再者，相關研究提出在非熟悉環境中應試

恐因考試焦慮致難發揮實力，因此，採用網際網路、無地點區域及時間限制的 IBT 開始受到關注。

IBT 除解決舟車勞頓、考試焦慮等問題，亦降低建置會場或參與會場認證之資訊設備、場地優化等投注成本，並且可以讓應考人隨時隨地參與應試，不再受應試座位數所侷限。然而，IBT 剛開始推動時，因其作弊機率之隱憂，僅適用於低風險、成績結果不影響重要決策之考試。

惟隨著資訊設備與技術之發達，為強化 IBT 防弊機制，開始運用攝影機及 AI 等資訊技術防止舞弊，由一開始透過攝影機由專人遠端即時監試，後來鑑於人工監試之人力將成為考試規模另一個瓶頸，進而佐以 AI 技術，採考試後檢查來強化防弊措施，因此 Prometric 推出 ProProctor 遠端監控 IBT 服務。

綜整來說，IBT 線上監試服務機制與功能，一般會強制設定應試電腦瀏覽器採全螢幕顯示、停用非考試相關軟體、停用部分鍵盤操作等安全管控，且隨著考試作弊機制演進，更要求禁止遠端桌面、多個顯示器、畫面擷取、虛擬攝影與錄音設備等管理機制。

受訪機關專欄資料提及，CBT 及 IBT 各有其適用之考試，除了客觀適用性外，亦受主辦考試機關國情文化等心理層面所影響，舉例來說，部分日本考試機關認為應考人採一次式同時會場型考試即表示公平性，此觀點與臺灣辦理國家考試非常相似。此外，社會亦有 IBT 考試影響個人隱私等議題之討論，因此，採用何種考試方式，皆須綜整多方評估後採用。

三、考試會場與報到流程概述

(一) 會場認證

受訪機關現場表示，會場之設置以應考人角度出發，目的在提供舒服不受干擾、減輕應試壓力的人性化考試空間設計，有統一標準，包含顏色、座位大小、監視設備等配置要求，參酌網站上資料，其會場認證程序如下。

Prometric 考試會場有兩種類型：一種為 Prometric 授權測試中心 (Authorized Prometric Testing Center, APTC)，一種為 TOEFL iBT 考試中心。APTC 係經 Prometric 認證之會場，所屬員工須經監試人員培訓，目前日本 47 個都道府縣約有 160 個 APTC，主要支援 CBT 考試。另一種類型係支援 TOEFL iBT 英語能力測驗考試 (含電腦口試)。

為確保考試的嚴謹性、公平性，Prometric 規範統一之 APTC 會場規格；會場設置第一階段由有意願者提出申請，Prometric 依據區域需求考慮是否受理設置新會場，並嚴格檢核申請方個資保護落實程度，以及審查資產、財報、樓層及教室配置、網路與軟硬體設備規格等，若通過，則進入第二階段。

第二階段為實地勘察，確認申請方真實狀況，以及交通、週邊情況、考試動線、考試環境及相關規格等。若通過兩階段審查，則簽署合作程序，並要求申請方依據 Prometric 規範準備軟硬體設備及參與培訓。最後經過驗證審查後，則列為考試會場，並參與支援各次 CBT 考試。此外，Prometric 將每年進行會場

審核，俾維持相對應品質。此認證程序，與臺灣國家考認證試場相似，所不同者，Prometric 每年審核合作方提供會場的品質，臺灣國家考認證試場則採效期制，認證合格有效期限5年，臨屆期前須重新提出再次申請，並由審查小組辦理實地審查。

（二）考試報到流程

參加 CBT 考試的應考人，須於網路報名階段填選考試會場及時間，並依據預定之時間抵達會場。抵達會場時，會先於報到處接受身分驗證，確認為應考人本人（詳下圖①）；經確認具應試身分後，進入置物室，將非考試相關用品擺放於置物櫃，並上鎖（詳下圖②），並進入等候區等候（詳下圖③），開始考試時進入會場（詳下圖④），考試當下會有監試人員採監視器（詳下圖⑤）監試或直接進入會場監試。

圖 12：Prometric 會場配置圖



資料來源：引自受訪機關網站資料（2024）

圖 13：Prometric 考試流程圖



資料來源：引自受訪機關網站資料（2024）

考察 Prometric Japan 當日，該公司提供參訪直營會場（御茶ノ水ソラシティ会場）之機會，惟因試務安全，不開放攝影或照相。我們由身分驗證開始，依序進入配置附鑰匙置物櫃的置物室、等候區，然後進入考試會場週邊的監試區。據該公司表示，為加強標準化並降低應考人焦慮，該公司全球會場皆採統一設計，包含色調、座位大小等等；每兩個座位對應一個監視器，監試影像直接傳到公司總部，分公司則需要特別申請方能看到影像。負責講解會場設置的管理人員有多年管理試場經驗，自信地回應我們的許多問題，進入試場整理的人員也都按照標準作業程序完成工作。由於維運高品質的會場需要不少成本，Prometric Japan 表示該公司非常積極地在引入更多考試使用會場。

四、人工智慧在考試服務之應用

Prometric 運用 AI 技術在線上監試與輔助命題，並有自己的伺服器與模型。Prometric 2年前買下美國專做 AI 命題的 Finetune 公司，運用大型語言模型（Large Language Model,LLM）技術，提供給 College Board 做命題應用，惟 AI 輔助命題之結果仍須由專家檢視正確性；該技術已運用於教育、財金、IT 等領域。

AI 大型語言模型若採開放式系統（如 ChatGPT），其好處是資料可以隨時更新，但會有資料外溢之安全性議題；Prometric 有自己的大型語言模型，並非採開放式系統，自硬體設備到語言模型都是自己設置擁有，所建置每一個 LLM 模型是針對客戶需求建立，不會與其他客戶混在一起。該公司表示，基本上每個應試科目每種題目類型需要5,000-10,000個題目，約20分鐘，即可完成 AI 輔助命題訓練，開始產出試題，且不同語文（例如中文）應該不是特別問題。

五、進階考試應用之建議

就辦理一次式同時大規模會場型電腦化測驗，Prometric 有實際同時辦理 15,000 人會場型考試的例子，其運作模式係使用該公司應試系統及安控程式，並搭配 BYOD 的方式來辦理。因 BYOD 之推動須有強健網路環境，且因應應考人攜進之各色電腦要有詳細標準作業程序（SOP）並搭配詳細的檢核表（checklist），逐步踏實落實執行，方能穩定辦理考試；是否適用臺灣國家考試，相關細節須研議。

就考察團詢問申論式試題非純文字作答介面之建議時，Prometric 回覆技術上或許可嘗試結合鍵盤與繪圖板方式加以解決，惟日本分公司總裁 Mr. Hagenbucher 表示，電腦化並非考試之目的，實務上應循序穩健之推動程序，上開需求允宜衡酌採紙筆測驗即可。

伍、小結

一、電腦化測驗應用於多階段徵才的第一階段考試

- （一）日本商業機構招募人才，第一階段常常需要參加所指定之考試（例如，SPI3、玉手箱、GAB 等，可採電腦化測驗方式進行考試），對於應試結果之要求各商業機構有所不同，並由各商業機構辦理第二階段選才。
- （二）日本公務人員人才招募區分為中央機關與地方政府兩大類。
 1. 中央機關公務人員由人事院負責第一階段命題（目前採紙筆測驗），及格者須於有效期限內向有意服務之機關應徵，後續由各用人機關負責第二階段人才招募。
 2. 受訪機關分享近年來部分地方公務人員考試第一階段採用 SPI3 成績辦理徵才，主要係公務部門徵才時間在商業機構徵才之後（即依據公私部門徵才的時間序，日本學生畢業後，會先遇到商業機構徵才，後續是公務部門徵才），鑑於商業機構多廣泛採 SPI3 進行第一階段徵才，地方政府為擴增招攬人才機會，因此也漸漸思考採 SPI3 來進行公務部門第一階段選才，除了避免造成應徵者須參加多次考試的困擾，主要也希望能讓優秀人才進入公部門。

二、應用電腦化測驗的考試多為資格考試，競爭性較低

- （一）日本選才（公務人員或商業機構人力招募）程序採多階段辦理，電腦化測驗考試常屬第一階段考試，且大多數為資格考試，取得資格者可以進入第二階段考試。
- （二）此種人才招募機制，第一階段及格人數無限制，因此沒有競爭上議題（非零和），應考人錙銖計較動機較低。
- （三）因第一階段的成績可能被運用在分派職務，若應考人作弊造假或取得高

分，其代價可能對未來職涯造成困擾，相對的降低作弊動機。

三、尚無辦理一次式同時於固定會場之大型考試案例

- (一) 日本電腦化測驗考試依施測地點區分固定會場（第一型）、臨時會場（第二型）、在家考（第三型）；依據施測時間區分為一次式同時考試及連續型考試（全年或期間限定）。
- (二) 搭配施測地點採不同監試方式，第一型採現場專人監試，且應考人非考試相關物品不得帶入考試會場，考試安全性及嚴謹度最高；第二型依據需求採現場專人監試或遠端專人監試，第三型依據需求採遠端專人監試、AI 輔助監試，考試安全性及嚴謹度最低。
- (三) 就電腦化測驗應用上，第一類型多屬嚴謹度與安全性較高的考試，應考人報名時需選擇考試會場，若連續型考試，則須再選擇考試時間；第二類型多為公司內部徵才或升等之考試，第三類型若採無監試方式辦理的考試，則多為模擬考、數位學習之成果考試。
- (四) 就受訪機關分享考試之辦理方式，中央機關公務人員考試為紙筆測驗，SPI3 舉辦方式之施測地點採第一、二、三型，TOEIC 施測地點採第一型、第三型，都有監試。
- (五) 日本電腦化測驗考試需求及應用多屬資格考試或檢定考試，非屬高競爭性，且多為連續型考試，因此考試當下發生偶發事件之因應方式則是提供應考人免費再參與同考試區間的考試，電腦化測驗偶發事件發生之風險值低，且考試無法進行之容忍度高。
- (六) 本次受訪三個電腦化測驗考試機構被問及一次可以多少人同時採固定會場型考試時，一開始都答不出來，這可能顯示這不是他們需要面對的場景。進一步估計或參考網路資料後，Prometric 文獻上提及約 2,000 人、CBT-Solutions 公司則為 1,000 人同時應試之規模。

四、推動 BYOD 須因應公平性議題

在日本，基於公平性考量，並無太多考試採用 BYOD 此種施測方式。據參訪機構人員表示，若應試科目試題包含受設備規格有所影響之內容（如解析度、亮度等），將因自備電腦規格不同影響試題之判讀或處理，致引發公平性議題，爰需妥善規劃推動策略。此外，BYOD 之推動在技術層面沒問題，但須有強健網路環境，並搭配標準作業程序（SOP）及與詳細完備之檢核表（checklist），逐步踏實落實執行，方能穩定辦理考試。

五、 人工智慧輔助命題仍需專人驗證正確性

此次參訪之3個公司皆運用 AI 協助網際網路型考試之遠端監試；另 Prometric 則運用 AI 在輔助命題（已有多個實際案例，如 SAT），惟 AI 輔助命題等成果仍需由專人驗證正確性。該公司有 AI 專業團隊（先前併購 AI 命題公司），並建置自有大型語言模型（Large Language Model,LLM）及軟硬體設備，以防止資料外溢之安全性議題。訓練 AI 輔助一個應試科目之命題約需5,000-10,000 個試題，20分鐘即可完成訓練，且語言不是訓練之核心問題。

第四章 融入性向測驗內涵之基本能力測驗

對於初任公務人員或是私人企業的人職評量，日本公私部門皆著重於應考人的能力及態度，並採取這些面向相對應的評量工具，以挑選符合組織所需的人才。本次考察在過往各次有關日本公務人員考試制度考察報告的基礎上，更深入對準日本公私部門於基本能力測驗的運作方式、實施結果分析以及背後的思考脈絡，觀察日本公私部門招募用人的策略，以供考選部研議參考。

壹、日本國家公務人員考試基礎能力測驗

一、基礎能力測驗概述

日本國家公務人員招募係採多試取才，二試最常見。第一試均包含「基礎能力測驗」，多數類科在第一試亦包含專門測驗（測驗題型）；第二試的測驗相當多樣化，也是依據其考試種類而有所不同，有申論型式的專業科目測驗、政策論文測及面試，如表 4。

表 4：日本國家公務人員考試各階段測驗種類、題數時間、配點比率及內容一覽表-以2024年綜合職大學畢業程度考試為例

考試階段	測驗種類	題數 時間	配點 比率	內容
第一試	基礎能力測驗 (選擇題)	30題 2小時20分	2/15	· 知能領域24題 · 知識領域6題
	專業科目考試 (選擇題)	40題 3小時30分	3/15	
第二試	專業科目考試 (申論題)	2題 3小時	5/15	各考試類別所需專業知識筆試
	政策論文考試	1題 2小時	5/15	
	面試		3/15	個人面試，內容包括性格、人際溝通能力等
英文考試				外部英語測試(托福(iBT)、多益聽力和閱讀測試(僅限公開考試)、雅思、實用英語能力考試)，並根據分數等在總分上加15或25分。
註：1 第一試依照基礎能力考試、專業科目考試的順序進行。				
2 第二試時將進行性格測驗，作為面試的參考。				

資料來源：日本人事院公布之2024年國家公務人員考試資訊

日本人事院認為，在基礎能力測驗表現佳者，能力與學習潛能足夠，有較大可能在未來從事各項工作時，有足夠的知識與能力面對時代變遷，處理各種不同的事務。人事院官員認為，國家公務人員在專業知識外，服務一般民眾更應有廣泛性的知能與常識，不能脫離民眾的生活太遠。因此，公務人員考試第一次測驗的應試科目中必須有基礎能力的評量，通過基礎能力測驗的應考人（部分加上其它不同型態的測驗題型式專業科目的專門測驗）才能進入第二次測驗，且綜合職與一般職公務人員基礎能力測驗納入總成績計算，具有門檻性質但也計入總成績計算。

二、施測內容

2012年以前的公務人員考試的第一次測驗中的基本能力測驗稱為「教養測驗」，改革後的新制公務人員考試所使用的測驗名稱為「基礎能力測驗」，包括「知能」與「知識」兩大領域。

（一）基礎能力測驗的內涵

基礎能力測驗於知能領域的命題內容包括文章理解、判斷推理、數的推理（含資料解釋）；知識領域部分則包括自然科學、人文科學、社會科學（含時事）。試題並沒有進行預試，而是根據試題審議會參與者和審議會決策過程中的經驗來做出決定和調整，所以是以參加的專家們的經驗來做分析。

（二）基礎能力測驗的題數及施測時間

過去基礎能力的編製，依不同類別進用考試配合不同的應試資格，題數不同，最少30題，最多50題，應試時間自1小時30分至3小時不等。例如國家公務人員大學畢業以上者，無論是綜合職或一般職知能題數量多於知識題數量，最少的比例是 2:1，最多的比例是 4:1。

惟自2024年起，基礎能力測驗的內容有所變動，知識題的權重下降，國家考試大學畢業以上的綜合職考試，基礎能力測驗題目統一改為30題（知能領域24題、知識領域6題），只有高中畢業程度仍維持40題的題數。除了題數減少外，知識領域的問題更專注於時事題，其內容係對於社會議題感興趣的應考人能回答的程度，同時亦新加入與資訊相關的問題，如表 5。

對於基礎能力測驗題數減成30題的原因，日本人事院官員表示，第一個是大學畢業程度的應考人越來越少，第二個是基礎能力測驗很難準備，如果應考人感到有壓力負擔，可能就不會參加國家公務人員考試，因此把此一科目的試題減成30題。

表 5：日本2024年起國家公務人員考試基礎能力測驗變動情形

2024年起新制	2023年以前
研究所畢業・大學畢業程度 30題/2小時20分 知能領域 24題 文章理解 10 判斷・數的推理（資料解釋） 14 知識領域 6題 自然・人文・社會（時事）・資訊 6	研究所畢業 30題/2小時20分 知能領域 24題 文章理解 8 判斷・數的推理（資料解釋） 16 知識領域 6題 自然・人文・社會（時事） 6 大學畢業程度 40題/3小時 知能領域 27題 文章理解 11 判斷・數的推理（資料解釋） 16 知識領域 13題 自然・人文・社會（時事） 13

資料來源：日本人事院官網2024年3月15日最新消息/2024年日本國家公務人員考試說明

（三）基礎能力測驗設計的理念

日本人事院官員表示，日本公務人員招募考試是在二戰後進行，人事院在此發展過程中發揮核心作用，人事院設定目標，依照社會的發展設計考試制度。由於社會發展來看，更傾向重視人的特質，因此對人的評量趨勢是強調知能而不是知識。

知能領域的題目評量目的是衡量應考人一般思考能力，不限於專業領域的思維，也衡量推理能力。「知能領域」中，有關文章理解、判斷推理、數的推理與資料解釋等試題命題藍圖的規劃，是針對國家未來的發展方向，人事院根據最近的實施考試的結果，考慮明年考試內容的方向。人事院網站上也提供範例問題。

（四）測驗向度與工作職能的關聯性

文章理解、判斷推理、數的推理與資料解釋這些向度均包含於普通性向測驗的向度之中，日本人事院認為這些測驗的向度與一般思考能力有相關性，與專業科目無涉，其內容涉及個人綜合理解和解決問題的能力，國家公務人員正是需要這些技能，因為行政部門需要能夠靈活應對日益複雜的社會需求。有關性向測驗中其他向度，例如記憶力和空間理解力等，人事院認為從測驗執行的角度來看，記憶力測驗很難實施，而空間的掌握能力，只會用於機場航空駕駛的測驗中，對於一般國家公務人員的業務內容而言，可能不是必要的。

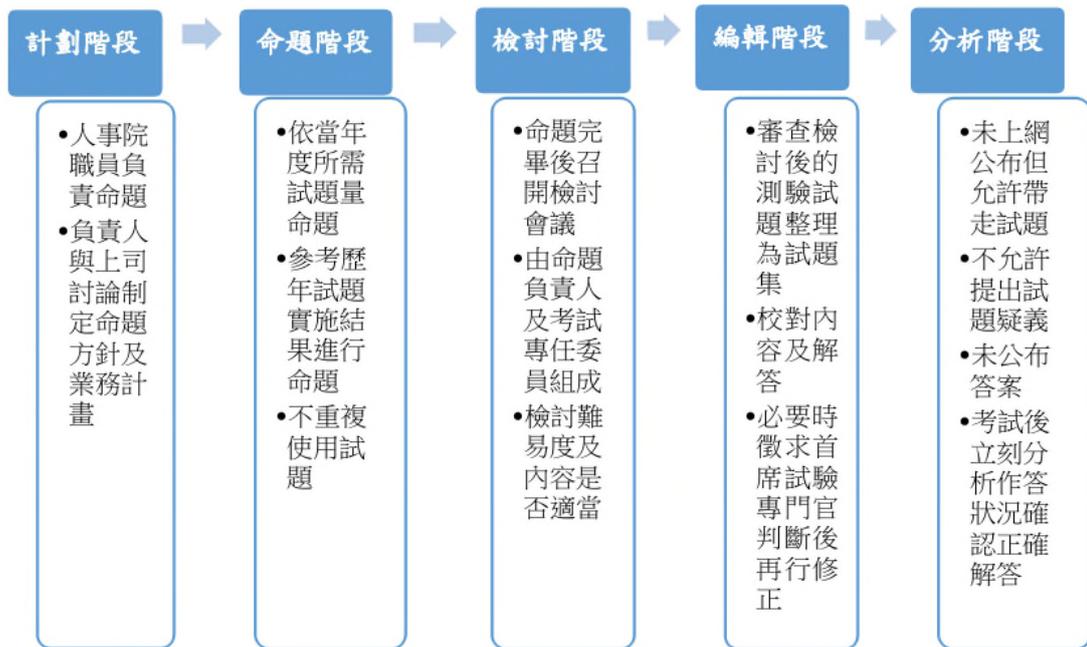
三、命題作業及相關問題

日本國家公務人員考試的命題者來源有三，一是大學教授，二是行政機關的職員，三是人事院人材局內部的員工。人事院對新員工進行初步的培訓，並在實作中得到經驗繼續學習。

(一) 試題開發

日本人事院測驗部門有40人，試題並未建立題庫，相關作業流程如圖 14

圖 14：人事院命題作業流程圖



資料來源：引自2023年日本公務人員考試「基礎能力測驗及新教養測驗編製」與「面試實施」考察報告第10頁。

(二) 編製知能領域試題的問題

在討論會現場，日本人事院官員表示，對於人事院而言，要如何將考試的試題內容設計與實際工作所需要的知識技能之間的距離不要太遠，是一個努力的方向。

四、關於施測結果的討論

(一) 試題可以重覆使用但會避免短期內再度出現

考察團詢問試題的難易度攸關考試錄取的結果，日本是不是也有應考人於考後反映試題越來越難，到這種情形有怎麼樣的對應？日本人事院官員的回答是，雖然人事院不會直接回應應考人在考試後的意見，但是確實會透過各種管道收集這些訊息，以作為未來出題的一些調整。有些試題已經使用過一次，會重新整理並且重複使用。

（二）短期應試課程的效果

在日本補習班是非常興盛的，日本人事院官員表示，補習班對於解題及提高分數確實有一定的影響，從政府的角度來說，不能限制應考人去補習班練習；但是從確保人才的角度來看，學生們普遍認為公務人員考試需要過度的準備，這才是一個問題。關於知識領域的測驗，屬於成就測驗，可以透過不斷的背誦記憶和理解，反覆練習以提高成績，人事院官員表示，依據其經驗，對於知能領域的測驗試題練習效果並沒有像知識領域來的好，雖然還是有一些提升，但知能並不是短期之內就可以馬上加強，因此人事院現在的方向是希望將知識的比重移轉到知能的方面。

（三）考試結果的分析運用

日本國家公務人員考試目前只採用紙筆測驗的方式，還沒有引進電腦化測驗。日本人事院官員表示，關於考試結果分析是基於準確率跟判別能力這兩個基本的指標，同時也會針對年齡、大學教育程度進行交叉列表，除此之外，不會另外對應考人進行特別的調查。

對於本次考察所詢問和科目之間的考試結果相關性，人事院官員表示，一般而言，第一試專業科目表現較優的應考人，在基礎能力測驗的表現，兩者之間考試結果的相關性是呈現正相關；至於「知能領域」與「知識領域」的表現則無法確認知能領域和知識領域之間的相關性。至於知能領域向度「數的推理」和「文意理解」之間的相關性，因為是評測不同的技能，所以人事院官員認為相關性並不高。

（四）有效排除不適任的人選

考察團關心知能領域此一測驗工具是否能有效排除不適合擔任公務人員的應考者，因為日本人事院對於考試設定了30分的切點，任何科目的成績在該科目30分以下就算是不及格，亦即每個科目只要有一科低於30分以下的話，應考者就會被排除掉，即使該名應考人於其他科目的分數特別高也不行。

人事院官員表示既然訂定這樣的標準，就有一定的排除作用，其後再跟其他類型的測驗結合起來，例如性格測驗、論文考試，更可以有效地排除不適任的人。最後，通過考試的人仍是要到各部會進行「官廳訪問」的過程，由用人機關選定其實際所需用的人。

（五）實施基礎能力測驗與入職工作表現之相關性

基礎能力測驗自2012年開始實施，考察團詢問實施之後與實際的工作表現之間的關係，人事院官員表示這是一個非常困難的問題，如果要採取科學的方法，應該要看考試結果和員工入職之後的工作表現之間的關聯性，但是目前各部會很難向人事院分享其員工於機關內的工作表現，因此無從得知其中的相關性。

五、性格測驗的運用

考察團認為日本與我國有相似的補習文化，面對以性格測驗作為面試的輔助參考資料，補習班是否會採用一些方式來教導應考如何應對性格測驗嗎？日本人事院官員表示，國家公務人員考試第一試採知能、知識性的測驗，第二試則會加入性格評估的部分。性格測驗旨在檢視應考人的積極性和社會性，性格測驗的結果也會作為第二試面試官的參考資料。由於性格測驗是基於應考人自己表達其性格如何，所以不能說可靠性很高，只能作為參考的材料，因此人事院官員表示，在日本並沒有聽說補習班針對性格測驗特別作解題的情況。

六、新興科技的導入

考察團對於新興科技包括 AI 的興起，於各行各業都有很多討論，並且帶來很大的影響，日本人事院對於試題導入 AI 的技術有什麼樣的想法，未來會朝向什麼方向發展，對此一議題，具有資訊背景的人事院官員表示，人事院對於導入 AI 的技術並沒有一個具體的政策，就其個人的意見，只是想知道是不是可以使用生成式人工智慧來建立測驗的問題，或者是命題後的勘誤，甚至是自動評分，這些是其個人有興趣的議題。

另外，人事院官員表示人事院的考試中，對於有資格參加第二試的應考人，有一些測驗的問題是事先給定的，例如政策論文測驗，會讓應考人事先知道施測的問題將會來自於政府的白皮書之類，因此有些人可能會想像會問題的內容，用生成式人工智慧預先寫成的論文，並且記憶背誦這些內容，於考試時寫成答案，對於此種情形，其對此持謹慎的態度，認為這也是新興科技對於未來帶來的一大挑戰。

貳、私部門的基本能力測驗—SPI3 的介紹

一、SPI3 能力測驗概述

SPI3是企業綜合適性測驗（Synthetic Personality Inventory）的縮寫，1974年發表第一代的SPI，2002年開始有第二代SPI2，到2013年改版為SPI3，由Recruit-MS公司開發，已被許多中小型公司及大型企業採用。本次考察之SPI3，包括測量潛能（而非學科成就）的能力測驗，以及測量性格的性格測驗，即用以衡量求職者的性格、工作適應力、融入組織的難易度等的測驗。在日本考科，「適性測驗」是指學科之外的測驗，相當於我國典試法第14條所指稱之「心理測驗」；在臺灣學術界，適性測驗指adaptive test，特別是電腦適性測驗，本次考察的是日文意義下的適性測驗，並非adaptive test。

根據Recruit-MS公司表示，SPI3之中的能力測驗所要衡量的，即是任何工作類型都需要的「智力」，具體來說，就是在實際工作中需要具備的能力，例如理性思考問題、設定目標並依照目標行動、有效率地處理情境等。這種能力不是

透過死記硬背、準備工作就能輕易提升的，而是需要透過長期的日常努力來培養的。換句話說，對複雜事務的組織和決策能力、學習新事物的能力、應用現有知識的能力，這些技能都是職場工作所必需具備的基本能力。

根據 Recruit-MS 公司的網站資料顯示，該公司會定期檢查 SPI3衡量的能力與員工加入企業後是否適配，並且證明它們之間存在關係，SPI3為企業公平招募選拔提供科學、客觀資訊的測驗。至於求職者關心的合格標準問題，Recruit-MS 公司表示能力測試所需的分數水平因委託評測的公司而異，沒有統一的測試結果錄取標準，而是從「是否達到招募用人公司所需的能力水平」的角度來看待。

二、施測內容

SPI3大致可分為兩種：性格測驗和能力測驗。

（一）性格測驗

性格測驗是確定求職者性格的測驗。這項測驗旨在從多個角度了解應考人日常的思考和行為方式，以及他/她持有什麼樣的價值觀。具體來說，應考人將看到兩個選項以及假設各種情況的問題，並選擇適用於那一個。性格測驗並不是分數越高越好的測驗類型。高分低分都有各自的性格特徵，是判斷應考人性格的指標。

（二）能力測驗

能力測驗客觀地衡量應考人的基本能力，其透過「語言性」和「非語言性」兩類問題來衡量應考人溝通能力、思考能力以及獲取新知識和技能的基本能力。語言領域的問題係衡量準確掌握和理解單字意義和故事要點的能力，而非語言領域的問題則係衡量數字處理能力和邏輯思考的能力。

表 6：SPI3能力測驗分類說明表

類別	衡量內容	題數及時間 (採適性測驗 (CAT))	備註 (對應性向測驗之向度)
語言	衡量是否能理解單字的意思和故事文章的要點。	基本上約為 40 個問題，作答時間約為 30 分鐘。	文章理解
非語言	測試數字處理和邏輯思考能力，並提出各種各樣的問題，從簡單的計算問題到需要從圖表中的信息得出準確順序和分解的推理問題。	非語言領域的問題以國中程度設計，題目數量約為30題，作答時間約為30分鐘。	數的推理
			判斷推理
			空間把握
			資料解釋

資料來源：參考 Recruit Management Solutions 官網 SPI3介紹內容，自行整理

三、施測的方式與題庫管理

(一) 施測方式

SPI3實施測試有4種不同類型的考試方式，第一個是會場型，在考試中心舉行，第二個是線上型，第三個是企業內部考試 (in house)，就是在公司內部設置的會場實施考試，第四個就是紙筆測驗。考察討論的是前三個項目。因為前三個項目都是用電腦來進行，應考人來參加考試，都是由題庫提供試題，考試是採用適性化測驗 (CAT) 方式，依應考人答題的情形，給予更難或更簡單的試題。

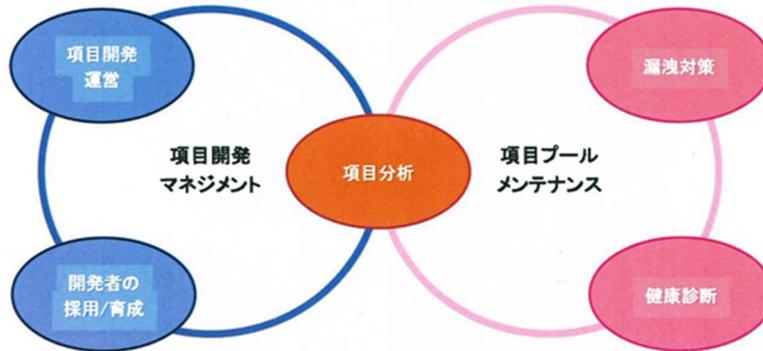
(二) 以管理導向建置兩個題庫

會場型的考試會在主管的監督之下，但是網路版沒有監督機制，而企業內部考試 (in house) 是在企業內部進行，由他們內部的員工監試，這是一個很大的問題，因此假設這些問題目會外洩的風險，所以線上的考試或者企業內部考試 (in house) 的考試的題庫會跟其他考試方式的題庫分開，亦即會場型單獨一題庫，線上型和企業內部考試 (in house) 型是另一個題庫。

SPI3盡可能避免使用外洩的試題，所以經常會檢查是不是有任何內容被洩漏，並且不再使用任何在網路上已經公開的試題。會場型的考試，有兩種方式監考的方式，一種是真人去監考，一種是用電腦的攝影機去監考。因此這兩個題庫的差別不是在於試題的建置方式，而是在於後續的管理方式，差別主要在此，其試題的難度是一致的。

四、命題開發及維護

圖 15：SPI3試題開發及維護示意圖



資料來源：引自受訪機關簡報資料（2024）

根據圖15 SPI3試題開發及維護示意圖，管理試題的運作方式，左邊是指新試題的開發，還有出題者的聘用和培訓，右邊的部分則是針對現有的試題的維護，一是防止外洩，二是試題的品質診斷，因為題目一直使用，題庫中的試題之有效性就會降低，如何防止試題的有效性降低，需要不停的修正和調整試題。

（一）試題開發著重試題的開發管理及命題者的進用培訓

新試題的實際開發是由外部專家來完成試題，現在大約有30位左右的專家在幫公司出題。試題開發者的管理對於試題的永續開發極為重要，因此會從「試題開發管理」和「試題開發人員的聘用和培訓」的角度做為開發的主軸。當公司收到這些專家們所出的題目之後，會開始進行一個會議來檢討，然後做分析採用。

因為發現試題有洩漏的情形，就會將其完全從資料庫刪除，一直刪除的結果，題庫的試題會越來越少，Recruit-MS 公司表示，會委託外部的專家每一個月都要命題，所以公司會一直接收新的試題，經過討論後後，會以試行的方式給應考人測試，每一個應考人會得到一題或兩題新試題，測驗結果就會去比較應考人在其他考題上面答題的情形，就可以去推測新試題的難易度。

（二）題庫維護著重試題的防止洩漏及品質診斷

關於試題題庫的維護，新試題的擴展是前提，但仍會採用收集資料、運用方法進行維護，並將重點放在定期檢查試題品質的做法，從「試題洩漏對策」和「試題健康檢查」的角度做為維護的主軸。

有關「試題洩漏對策」，因為企業會要求應徵者參加SPI3考試，所以Recruit-MS 公司第一個方式是鼓勵企業採用求職者前一次參加別家企業 SPI3的結果，這樣子的話，就可以自然的減少同一個求職者考 SPI3的次數；第二個方式就是如果還是希望可以重複考的話，至少在會場型的考試（如前述 Prometric Japan 公司提供測驗試場之會場型考試），可以確保的是該名求職者前兩次考過的試題不

會出現，這個部分目前的技術是做得到的。另外，SPI3盡可能避免使用外洩的試題，所以經常會檢查是不是有任何內容被洩漏，並且不再使用任何在網路上已經公開的試題，如經發現，會將其從題庫中刪除。

同時，試題分析是連接「試題開發」和「題庫維護」的樞紐，試題分析為評估試題品質提供了證據訊息，透過與題目開發者分享結果，可以促進對好問題的共識，有助於試題開發者的命題發展。

（三）命題預設初級中學畢業的程度

在設計 SPI3考題的時候，基本上設定在初級中學畢業可以理解的程度。設定在初級中學，是因日本的義務教育年限為小學及中學校（初級中學）。命擬試題時，著重在問題的解決能力，不會非常困難，主要是測驗應考人推論、論理的能力，這是會請命題者要注意的重點。

五、關於施測結果的討論

（一）短期應試課程成效有限

一般來說，性向測驗無法透過短期密集練習而提升成績，考察團詢問 SPI3的能力測驗是否在短期之內重覆施測下，應考人的成績也不會有多大的改變呢？Recruit-MS 公司表示，他們作過研究，先讓受測者參加 SPI3考試，後續分成兩組，一組於考後給他們一本參考書，繼續做模擬試題，另外一組則是什麼事也不做，經過一週之後，再讓這兩組人參加第二次考試。而再次考試的結果，繼續做模擬題那一組受測者的分數，雖然確實有小波動，但大概就是兩到三分的差距，成績差別並不大。

（二）基本能力測驗包含語言、論理及數學三個測驗向度的能力

Recruit-MS 公司表示，就一個企業入職考試來講，SPI3目前實施的是基礎能力測驗（GAT），也就是在商業上應具備的基本能力，屬於一般人入職時的測驗標準，內容的難易度基本上設定為中學程度，例如中學的語言能力、數學能力等這些問題。就內涵來看基礎能力，測試三個部分，第一個是言語，第二個是論理，第三個是數量，所以要有數學的能力，還要有語言的能力，再來就是分析申論的能力。而設定中學程度的理由，是因為日本的義務教育是到中學（國中），所以 SPI3以日本全部的國民義務教育所都具有的程度，並以此為前提命擬試題。

（三）運用「偏差值 X 信度 X 效度」作為分析試題品質的方法

從品質上的角度來看，大約從三個點來檢查，一個是標準值，標準值就是以他們這一年來參加的應考人表現成績的平均來計算，在日本來說，這個叫做偏差值的概念；另外一個就是信度，就是各個公司對他們施測結果信賴的程度，會以各種角度來確認信度，主要是用 α 值（alpha）作為信度的基礎。另外一個就是效度，Recruit-MS 公司要知道員工在職場的表現，會定期在各個產業裡面做進行調查，知道這個測驗的結果和他們考試想要要求的能力是不是符合；

同時也會請企業提供員工的工作表現，例如人員的評價、升遷速度、工作的實際績效，透過這些資訊的收集，檢查與他們的測試之間是否存在關係。

依據 Recruit-MS 公司網站的資料，公司成立40年來，累積海量的應考人作答資料，與應考人入職後的人員評價進行比較，驗證作答成績與入職後人員評價之間的相關性。因此，SPI3的優勢在於其能力測驗和性格測驗的高度有效性。再者，應考人對於 SPI3，即使進行多次測試，結果也不會發生顯著的變化，這是基於「可靠性」，還具有高水準的「標準化」，使應考人能夠相互定位。

六、地方自治體招募人才採用 SPI3 的施測結果

過去橫濱、大阪都有採用 SPI3，對於地方自治體使用的情形，Recruit-MS 公司表示，SPI3並沒有根據公務人員的特性改變試題的內容。地方自治體接受 SPI3 的評測結果，主要目的是要跟民間公司競爭，爭取優秀的人才，希望到私人公司應徵的求職者，也能夠到公部門申請。由於日本一般民間公司就職的時間會比地方自治體考試的時間早一點，地方自治體的考量就是這些優秀的學生既然已經考過 SPI3，那麼接下來如果公務人員考試也接受 SPI3的話，應考人不用特別準備的地方自治體的考試，他們也許會想到公部門試試看。但值得注意的是，地方自治體只採用會場型和企業內部考試（in house）SPI3的評測結果，並不會採用線上型的考試成績。

在本次考察蒐尋資料過程中發現，過往資源較豐裕而自辦命題作業的地方自治體東京都，自2024年起秋季公務人員考試的第一試，亦開始採用 SPI3的測驗結果。

七、性格測驗的運用

（一）以實際經驗的積累發展人格量表

SPI3有18種性格因素作為人格量表，更準確的說，這是根據特質理論。由最常見的 MMPI 和 MBTI 為基礎，Recruit-MS 公司研究的起點是思考作為一個日本公司，要尋找什麼樣的性格特質的員工，所以性格測驗更像是透過實際經驗的積累而發展和創造出來的。

考察團詢問 SPI3中的性格測驗，是否每個公司都有一些調整關於主題中的量尺，例如是把18個因素全部分成單獨的量尺，還是以四個大項計算出來的量尺分數？Recruit-MS 公司表示，一切都是共通性的，不是以公司來劃分，但是第一個答案是每個人都有自己的優點，沒有一個整體的觀點，因為每個人測的是他自己的個性，每個人的個性都有不同的地方，所以他們想要展現的是這個考試的結果，每一個人的個性在各種不同的方面的程度是怎麼樣，所以並不會特別去把每一個細項把它加成一個大分數。

（二）性格測驗以呈現真實自我為主

因為在臺灣看到這樣的情形，考察團詢問在日本，有應考人針對性格測驗去補習嗎？Recruit-MS 公司的回答很有啟發性，過去是有這樣子的情形，但那是很久以前的事，大概是在2005年左右，大家會為了考 SPI3寫很多的模擬題，那個時期可能是有的，但是現在公司並沒有出書說明性格測驗。Recruit-MS 公司強調性格測驗應以個人的性格為主，亦即求職者不應該特別為考試準備，才能夠顯現出來最直接、最正確的測驗結果。因為現在企業用性格測驗的檢測結果當成求職者的人事資料，求職者進到公司之後，會依照 SPI3 評測的結果分發工作單位，因此，如果在測驗中不是真實的反映自身的個性，就容易分配到一個不是屬於適合自身個性的工作，這樣的結果對當事人來說，也是很麻煩、很辛苦的一件事。Recruit-MS 公司沒有出版這些模擬題，也不會把考古題出版，現在也對外宣傳請應考人要依照原本的自己的個性來回答性格測驗的問題，如此一來對應考人的就職才是最有幫助的。

（三）性格測驗做為面試的參考資料

關於性格測驗在作為甄選資料時，可能有社會期許或假裝完美（faking good）的部分，Recruit-MS 公司表示，在開發時性格測驗的試題時，有組成研究小組，處理此一問題。研究將受測者分成實驗組和對照組，要求實驗組在回答題目時，依照社會的期許回答問項；而對照組則是依照自己的個性回答問項。之後再依據施測的結果，進行分析，觀察那些試題會導致差異；同時該公司亦參考國外學者的研究結果，並透過在日本實際收集的數據、驗證並轉化為結果來完成。

Recruit-MS 公司表示，就結果來說，並不會特別告訴委託評測的公司求職者的性格測驗的評測結果是以符合社會期許的方向去作答，他們只會提醒委託評測的公司，求職者的實際情況還是要再透過面試進行確認。

八、過去建立 SPI3 的契機與未來發展及新興科技的導入

（一）發表研究成果或實施數據以取得外界的認知及信任

考察團對於 Recruit-MS 公司這麼一個兼具品質及規模的測驗機構，詢問早期建置 SPI3的契機，如何建立知名度以及取得各界的理解與信任？Recruit-MS 公司表示，公司成立的時候內部已經有一些是心理學專業的專家，一開始就十分重視開發研究，當時開發者主要的運作模式就是把他們的研究成果或是實施考試所得的數據，主動於學會論壇或是研討會發表，讓更多的人知道 Recruit-MS 公司提供基本能力測驗評量的服務。所以一開始著力點就不只是進行研究開發，同時還會對於實證的數據寫成研究成果，透過學術論壇的發表，爭取曝光度，並基於施測的成績進而取得外界的信任。

（二）對於新興科技應用於命題或監試的功能尚在觀察

因為現在國際局勢的變化，不管是應用新興科技（如人工智慧）或是因應

商業版圖的關係，SPI3對未來有沒有什麼特別的計畫？Recruit-MS 公司表示，有關人工智慧的應用，就現在這個時間點來說，目前還沒有實際的運用。比如說人工智慧的監試功能，其實已經有研究過，但是就技術來講，認為還是有困難，所以就已經先放棄這個方向，未來的發展的可能就是出題的方式，該公司認為生成式人工智慧技術都已經存在，以後也許會更有效率的出題。

參、小結

在本次考察國外制度時，考察團意識到處理相同議題可能需考量社會文化差異，因此問題宜放在此脈絡下審視。秉持這樣的原則，對應本次考察有關性向測驗與性格測驗施測信效度的相關討論議題，以及對於多元評量測驗所欲探詢的問題意識，臚列以下觀察結果：

一、日本公私部門招募員工的特色

（一）「基本能力測驗 X 性格測驗 X 面試」是篩選人才的基本組合

在職場上，能力與態度是職場的兩大核心競爭力，它們共同影響一個人在工作中的表現和長期發展。在能力方面，無論是日本人事院或者是 Recruit-MS 公司，咸認為因應工作所需具備的能力，並不是透過死記硬背、事前準備就能輕易提升的，而是需要透過長期的日常努力培養的基本能力，亦即應用現有知識解決問題以及學習新事物的能力。在態度方面，包括對工作的積極性、責任感、抗壓性，以及與他人合作的精神，公私部門均透過面試，以性格測驗為輔助參考資料，採用多元評量的方法，篩選出適合的人才。

（二）基本能力測驗與專業科目評測不同的能力面向

雖然本次考察的主題是基礎能力測驗，但是仍可以發現日本公務人員考試仍是有列考專業科目，在第一試有40題選擇題形式的專業科目考試，第二試則有申論題型的專業科目以及政策論文，代表基本能力測驗和專業能力測驗評測的面向對於日本人事院而言並不相同，以公務人員工作及身分，以及掌握國家資源分配權的特質，所須的能力是在基本能力測驗之外，尚須兼具專業知識能力，各自評測的目的不同。而沒有基本能力測驗的專業科目，有可能錄取到具有專業知識但缺乏跨域理解和解決問題能力的人。

再者，公部門施測的基礎能測驗和 SPI3的能力測驗，評測內容皆包含閱讀理解和數字處理能力和理論思考能力，可以透過選擇題的方式設計試題，可大量地在短時間內分辨能力高低，通常可做為第一階段篩選的評量工具。

（三）知能領域不容易靠短期應試課程改善成績

屬於成就測驗者，可以透過不斷的背誦記憶和理解，反覆練習以提高成績，日本人事院官員表示，依據其經驗，對於知能領域的測驗試題練習效果並沒有像知識領域來的好，雖然還是有一些提升，但知能並不是短期之內就可以馬上加強，因此人事院現在的方向是希望將知識的比重移轉到知能的方面。

Recruit-MS 公司則透過科學實驗的方法，先讓受測者參加 SPI3 考試，再將受測者分成兩組，一組於考後給他們一本參考書，繼續去做模擬試題，另外一組就是什麼事也不做，經過一週之後，再讓這兩組人參加第二次考試。而再次考試的結果，繼續做模擬題的受測者的分數，雖然確實有小波動，但大概就是兩到三分的差距，成績差別並不大。

（四）性格測驗對於工作分發的輔助功能

日本國家公務人員考試實施性格測驗旨在檢視應考人的積極性和社會性，性格測驗的結果也會作為第二試面試官的參考資料。由於性格測驗是基於應考人自己表達其性格如何，所以不能說可靠性很高，只能作為參考的材料，因此日本人事院官員表示，在日本並沒有聽說補習班針對性格測驗特別作解題的情況。

Recruit-MS 公司強調性格測驗應以個人的性格為主，亦即求職者不應該特別為考試準備，才能夠顯現出來最直接、最正確的答案。因為現在企業用性格測驗的檢測結果當成求職者的人事資料，求職者進到公司之後，會依照 SPI3 評測的結果分發工作單位，因此，如果在測驗中不是真實的反映自身的個性，就容易分配到一個不是屬於適合自身個性的工作，這樣的結果對當事人來說，也是很麻煩、很辛苦的一件事。所以 Recruit-MS 公司沒有出版這些模擬題，也不會把考古題出版，現在也對外宣傳請應考人要依照原本的自己的個性來回答性格測驗的問題，如此一來對應考人的就職才是最有幫助的。

（五）透過考試設計排除不適任人員

考察團關心知能領域此一測驗工具是否能有效排除不適合擔任公務人員的應考者，因為日本人事院對於考試設定了30分的切點，任何科目的成績在該科目30分以下就算是不及格，亦即每個科目只要有一科低於30分以下的話，應考人就會被排除掉，即使該名應考人於其他科目的分數特別高也不行。

人事院官員表示既然訂定這樣的標準，就有一定的排除作用，其後再與其他類型的測驗結合起來，例如性格測驗、政策論文考試，更可以有效地排除不適任的人。最後，通過考試的人仍是要到各部會進行「官廳訪問」的過程，由用人機關選定其實際所需用的人。

（六）命題者培訓的重要性

公私部門實施針對處理工作所需技能的基本能力測驗，針對應考人個人綜合理解和解決問題的能力進行評測，測得其未來發展潛能。基本能力測量的目標係對應各項技能之面向，依評測重點設計試題內容，用以反映受試者能力的程度，是以試題的良窳與命題者對於評測內容的了解程度有高度相關，因此命題者的招募及培訓十分重要。

日本人事院對新員工進行初步的培訓，並在實作中得到經驗繼續學習。對於 Recruit-MS 公司而言，新試題的實際開發是由外部專家來完成試題，現在大約有30位左右的專家在幫公司出題。試題開發者的管理對於試題的永續開發極

為重要，因此會從「試題開發管理」和「試題開發人員的聘用和培訓」的角度做為開發的主軸。

二、對我國國家考試的啟發

近年國家考試論述的主軸政策，列入加強多元測驗評量，其背後的原因在於主政者意識到，找到好的公務人員，甚至好的人才的時候，不應該只侷限在專業能力，因此考試院及考選部規劃在知能領域評量和性格測驗方面多加考量。此次考察有機會近距離觀察公私部門對於基本能力測驗的觀點及施測的實際經驗，透過兩個不同的視角，在考試公平、有效性的前提下，對於人才篩選工具有許多一致性的看法。

事實上，透過國家考試評量內容的設計有助於帶動學習重點的重新選擇，在公私部門職場工作特徵的界限逐漸模糊的現今，日本人事院和 Recruit-MS 公司兩者實施基本能力測驗的經驗啟發關於國家考試導入多元評量方式的3個思考點：第一，在工作強調能力和態度並重的公務職場，初任公務人員入門條件是否須重新檢視？第二，國家考試適合導入性向測驗何種向度？可以透過那些途徑？第三，對於多元評量方式的導入，可以透過那些途徑與利害關係人對話以取得理解與信任？或許可以從本次考察的內容中，得到值得參考的答案。

三、未來國際合作的可能性

對於本次考察性向測驗議題的兩家機構，人事院為公家機構，需在考量兩國外交下進行交流；編製 SPI3的 Recruit-MS 公司是日本大型的測驗機構，考察討論中認為測驗需要考慮到當地文化，似乎沒有表現對國外拓展的興趣，惟後續彼此交換意見仍屬可能。

第五章 考察心得與建議

壹、電腦化測驗

一、宜針對現行電腦化測驗的需求變動擬定對策

(一) 考選部一次式考試與日本的連續型考試之對比

考選部舉辦之國家考試，皆為一次式考試，亦即，同一考試所有應考人皆於同時間應試，因此，電腦試場座位數需高於應考人數。考選部每年舉辦數個一次式電腦化測驗考試，座位數則需以數個考試之最大應考人數規劃。舉例來說，如果考選部舉辦5次考試，應考人數分別為2,000、5,000、6,000、10,000與15,000，則需至少備有15,000個座位（未考量因應設備故障之備用座位需求）；而以每次考試2天計算，15,000個座位中，每年有2天使用15,000個位置、8天使用10,000個以下位置，另有355天使用0個位置，需求變動（Demand Variability）很大；座位數需求很高，但多數考試不需這麼多座位，座位平均使用次數很低（目前為2-3次/年）。

由於電腦試場設置成本高，考選部無法自行維持眾多座位（僅設置少數），而需向學校（大學或高中）借用，能否辦理考試操之在人；且為考試安全，學校需配合於一定期限內維持設備不變，而借用可能影響日常營運（教學），或修繕規劃（寒暑假），學校借用之成本高；考選部每年借用平均僅使用2-3次，給學校之經費誘因不高。

由於一次式考試需讓所有應考人於同時間考試，因此需要作快速處理偶發事件之準備，也因此，需備有較多之備用座位。

相較於此，日本的連續型考試，則是在一定期間內，讓應考人依試場開放日期與開放座位數報名考試，應考人分散於不同日期，座位數可少於應考人數。在處理不同考試應考人數引起的座位需求變動方面，由於連續考試將考試人次分攤在區間內，所需總座位數不高，但座位使用頻率高，營運效率較佳。

同時，由於營運效率較佳，硬體與人員能有效利用，因此部分日本電腦化測驗公司有自營或固定合作之電腦化測驗場所（稱為會場、測驗中心），亦有意願投資提升試場環境與人員訓練；同時因為連續使用，降低考試前檢查成本，人員亦能快速累積排除問題的經驗。

連續考試「技術上」容許應考人多次應試（個別考試可能有區間內次數限制），提供應考人更多考試機會；同時，沒有「應該」何時考試的議題，如遇設備或網路故障等偶發事件無法繼續考試時，不須擔心需等候至次年才能應試，可安排於近期內再次應試，不影響應考人權益，考試安全韌性較高。

(二) 日本推動策略與經驗無法直接擷取或全數採用

前述連續考試雖有種多優點，但仍須搭配條件，方能採用。首先，連續型考試以隨機抽題或是電腦化適性測驗方式，讓使用不同題目的應考人之成績位於同樣尺度，這需要搭配足夠大的題庫，各題有客觀之試題參數，以能「連

續」出題。

其次，社會需要接受考不同試題的連續考試仍屬公平；在日本，電腦化測驗多屬第一階段考試，且大多數為資格考試，其結果僅為後續階段或分班、分級之參考，考試風險較低，因而應考人作弊動機較低。

此外，連續型考試應試當下發生偶發事件可以再應試一次因應，容錯度較高，與國家考試專門職業及技術人員醫事人員採電腦化測驗之一次式同時考試之性質不同，爰其應試架構、推動策略與經驗無法直接擷取或全數採用。

(三) 應試系統與架構相容之前題下得著手研議與他考試機構聯盟

前述說明考選部一年舉辦多個一次式同時會場型電腦化測驗考試，各考試對電腦應試座位需求數不同，高峰與低峰的變動過大，因而有需求變動問題。在不變動一次式電腦化測驗考試方式下，聯合其他考試機構共同辦理電腦化測驗考試，將可降低需求變動，提升使用效率，並累積辦理次數與經驗，增加考試服務品質與辦理穩定度，或許在應試系統與架構等環境相容之前題下，得著手研議。

(四) 具質優量多試題與國人觀念改變之前提下得推動連續型考試

連續型考試優點有：不須「同時」借用大量足數應試座位、遇無法即時處理之偶發事件（如即使是電力、網路）因應度高、會場硬體設施與監場人員得充分被運用，以及多次辦理之經驗累積將增加考試辦理品質等，惟其前提條件須有大量經客觀驗測之高品質、具衡鑑度試題，且施測之對象及國情民俗須具備連續型考試具考試公平性之接受度，上開前提條件可為國家考試電腦化測驗未來努力的方向。

二、 人工智慧用於國家考試

(一) 人工智慧監試無立即引用之需求

AI 監試技術用於在家考試或其他遠端考試之情境，惟日本受訪機關皆表示其嚴謹度遠低於會場型專業人員現場監試，且目前應用亦不廣泛，對臺灣國家考試電腦化測驗而言，目前並未採在家考試或遠端考試，爰 AI 監試尚無立即引用之需求。

(二) 若需人工智慧輔助命題，允宜建立自有或共有大型語言模型

由於國家考試須確保試題安全，如使用非自屬之伺服器，則有特別注意保密之議題。就 Prometric Japan 公司分享運用 AI 協助命題之經驗，鑑於開放式系統會有試題外溢之安全性議題，爰該公司自建大型語言模型（LLM），並就客戶需求建立專屬 LLM 模型；雖考選部目前尚無此應用規劃，且 AI 發展須投注大量成本與人力資源，惟其不採開放式架構，而建立自有或與其他考試機構共有獨立伺服器與大型語言模型之安全性概念，可納入未來推動之規劃。

三、日本律師、檢察官等公務人員考試採電腦化測驗值得關注

於本次參訪中，受參訪機構對於風險較高之考試則皆採連續型會場考試，並經詢問日本現行情形，沒有看到採用一次式會場考試之案例，與考選部辦理之電腦化測驗方式有不少差異，因此，日本之電腦化測驗經驗未能直接用於考選部辦理之電腦化測驗。

據了解，日本法務省正規劃律師、檢察官、司法官公務人員考試，於2026年辦理電腦化測驗，人數約10,000-14,000人。由於競爭性較高，是否採一次式考試或其他辦理方式，或其應試架構、試務程序與辦理結果，以及應考人反饋意見，皆值得持續關注及進一步考察。

貳、融入性向測驗內涵之基本能力測驗

一、心理測驗之使用

(一) 多元評量是世界趨勢

根據典試法第14條，國家考試「各種考試得採筆試、口試、心理測驗、體能測驗、實地測驗、審查著作或發明、審查知能有關學歷經歷證明或其他報請考試院同意之方式行之。」為鑑別應考人的多元評量方式提供法律規範的依據。惟考選部辦理的國家考試長期以專業科目筆試為主，著重於評量應考人的專業知識或學習成就，現行部分考試或類科有納入口試，以進一步評估應考人的綜合表現。

考選部前常務次長曾慧敏於2023年國家人力資源論壇第27期「再思公務人員考試心理測驗角色」文內表示，心理測驗雖列於得採用之考試方式中，惟考試院僅在1950-1965年間試辦、應用與發展，採用的心理測驗則包括智力測驗與性向測驗。余民寧教授亦於2015年國家菁英季刊第11卷第4期「再議心理測驗在國家考試上之可行性」一文中提及，多位學者曾建議國家選才使用心理測驗，並可考量採用性格測驗。

相較於此，日本的公私部門廣泛採用著眼於發展潛能或基礎能力的基本能力測驗（屬性向測驗），並透過多元評量的方式，結合性格測驗與面試，以篩選符合組織需求的人才。在當今社會中，環境變化快速且日益複雜，除了專業知識和學習成就外，應考人的學習能力、溝通表達、適應調整能力與工作態度也至關重要。多元評量正是因應這些需求的有效方式，能更全面地反映應試者的實際工作潛力。

(二) 應辨識公務人員職場的基本能力向度

基本能力測驗應涵蓋職場所需的基本能力向度。2024年，日本人事院增強了知能領域的測驗題數，題目內容包括文字理解、判斷能力及數據推理（資料解釋）。SPI3的能力測驗則涵蓋語言和非語言兩大面向，前者衡量準確掌握和理解單字意義和故事要點的能力，後者則係衡量數字處理和邏輯思考的能力；性

向測驗則通過多種向度來評估不同技能，以測量應考人的潛在發展能力。我國公務人員考試可分為行政類和技術類，涵蓋多種考試類科，所需的工作技能亦相當多樣。為確定哪些屬於共同的基本能力，宜透過工作分析（job analysis），對準基本工作需求，進而選擇適合的性向測驗向度。

（三）性格測驗不計分，係作為面試的參考資料

性格是影響工作績效的重要因素之一，但於採用性格測驗進行人事篩選時，因測驗依賴應考人的自我陳述，容易受到假裝完美（faking good，即應考人可能傾向於過於正面的自我表現）的影響，從而降低測驗結果的可靠性，並可能削弱其鑑別效果。

在日本的國家公務人員考試中，第一試主要評估知能與知識性內容，第二試則會加入性格評估的部分。性格測驗的目的在於了解應考人的積極性和社會性，其結果作為第二試面試官的參考資料。因此，日本僅將性格測驗視為輔助工具，使面試官能從多個角度綜合觀察應考者。在不影響第一試筆試成績的前提下，將性格測驗結果作為第二試口試的參考資料，有助於面試官更全面地評估應考人，減少假裝完美的影響，提升口試結果的客觀性和公正性。

二、心理測驗編制團隊

（一）基本能力測驗命題技術須要長期培訓

基本能力測驗和專業科目測驗的能力面向並不同，對於人才的篩選各有其功能。日本國家公務人員考試的命題者來源有三，一是大學教授，二是行政機關的職員，三是人事院人材局內部的員工。SPI3 的命題與檢查等流程，則由30位外部專家與 Recruit-MS 公司內部人員共同組成。針對性向測驗中的文意理解、判斷推理、數的推理及資料解釋等各面向，其取材設計涉及綜理解及運用，雖未必涉及高深的專業學識理論，但屬於高層次的命題技術，因此命題者的招聘及培訓是非常重要一環。我國現行國家考試專業科目的評測，屬於成就測驗，其試題內容多著重於記憶及理解，對於性向測驗的綜理解及運用所需的高層次命題技術仍屬陌生，未來國家考試導入性向測驗或融入性向測驗內涵之基本能力測驗，須將命題者的聘用及培訓納入考量。

（二）專業團隊維運確保試題品質與考後不公開的必要性

考試院曾於1950-1965年間試辦、應用與發展心理測驗，惟於1967年取消；曾慧敏於前開「再思公務人員考試心理測驗角色」一文中亦說明，取消的原因之一，即在於心理測驗之實施應進行預試確認其難易度與信效度等，與國家考試之保密要求不相符。依此來看，如國家考試擬納入心理測驗，以專業團隊維運，以確保試題品質實為重要考量。

以日本人事院經驗，性向測驗是由專人專責逐年進行，研究每次施測結果，並回饋到下一次考試。此種專責的維運組織或單位，方能確保試題的標準化及內容適切性，並維持評量應考人未來潛能的有效性。

同時，國家考試公務人員考試一年舉辦10次公務人員考試，每次考試大多分不同等級，不論是實施性向測驗或是融入性向測驗的基本能力測驗，所需之試題量龐大，從試題的開發、檢查、修正以及分析回饋，考選部如要採用如日本人事院或 SPI3的融入性向測驗的基本能力測驗，需及早規劃命題人才。

再者，有鑑於試題取材及內容涉及綜合理解及運用設計，屬於高層次的命題技術試題，未來可經整理後再次利用，因試題具有特殊性，宜考量考後試題不予公布，以避免試題快速耗盡。

附錄 1【考察計畫】

赴日本考察電腦化測驗與性向測驗發展現況與趨勢

壹、考察目的

國際教育測驗發展趨勢，電腦化測驗對人才招募及能力評鑑有實質推動之意義；國家考試局負為國掄才及衡鑑專門職業及技術人員執業資格之重任，工欲善其事、必先利其器，考選部自93年起推動電腦化測驗，陸續推動15種專門職業及技術人員類科採行電腦化測驗；為因應國內外資訊領域快速發展，符應民意對數位化考試之期望，爰規劃赴日本考察推動電腦化測驗之測驗機構，藉由學習國外電腦化測驗考試架構與方法，強化國家考試效度與效能。

性向測驗為測量特殊能力或能力群而設計之工具，可以用來預測個人在職業上成功的可能性，亦即測量個人在某領域的能力或傾向，例如普通性向測驗包含一般學習能力、語文性向、數的性向、空間性向等9項，各種施測內容因應職務需求而調整變化。日本公私部門皆廣泛採用性向測驗的基本因素編製測驗項目，做為篩選人才的施測工具，如招募中央公務人員的人事院、為招募地方公務人員供題的人事測驗研究中心以及招募自成一格的東京都廳等皆編製有類似性向測驗的施測試題，做為公務人員考試第一試筆試的科目內容如文章理解、判斷推理、數的推理與資料解釋等；至提供企業招募人才測驗亦有專業公司編製性向測驗，如 Recruit Management Solutions 公司之 SPI3，採用心理學的人格理論，基於日本社會文化而設計的性向測驗及人格測驗。

為了深入了解日本電腦化測驗在人才招募及能力評鑑之發展現況與趨勢，以及日本公私部門採用性向測驗篩選人才與培育人力之施測內涵與實務要項，規劃赴日本考察訪問，期精進國家考試方法與技術，俾利我國公務人力適才適所，並精準篩選專門職業及技術人員，為國家社會公共利益或人民生命財產嚴格把關。

貳、考察重點

考察計畫包含電腦化測驗與性向測驗兩大主題，其中電腦化測驗係為學習日本當地不同電腦化測驗應試系統功能與特色、應試與監考程序、試場規模與設置標準；並了解電腦化測驗施測對象、規模、辦理考試方式，以及不同考試方式間之選擇；另亦探討電腦化測驗系統架構、考試安全性與穩定性(如電力供應與試題保密)等議題，且更進階觀摩新興科技施測之推動，以及 AI 在考試上之應用。

至性向測驗主題，則規劃了解知能(ちのう)施測內涵以及其命題藍圖規劃、觀摩試題取材來源以及相關命題者的來源和培訓，並探討性向測驗施測結果之信度及效度，最終考察採用性向測驗之機構與實施效益。

參、擬參訪機關（構）

- 一、株式会社シー・ビー・ティ・ソリューションズ（CBT-Solutions）
〒101-0022 東京都千代田区神田練塀町3 AKS ビル6階
- 二、株式会社イー・コミュニケーションズ（e-communications）
〒106-0032 東京都港区六本木7丁目15-7 新六本木ビル SENQ 六本木704
- 三、プロメトリック株式会社（Prometric Japan）
〒101-0062 東京都千代田区神田駿河台4-6 御茶ノ水ソラシティ アカデミア5F
- 四、株式会社リクルートマネジメントソリューション
（Recruit Management Solutions Co., Ltd.）
〒108-0023 東京都港区芝浦3-16-16 住友不動産田町ビル東館 4F
- 五、人事院
〒100-8913 東京都千代田区霞が関1-2-3

肆、考察團員：計3人

- 一、領隊：鄭政務次長中平
- 二、團員：數位國考及資訊管理司楊科長淑如、測驗發展司歐陽科長安琪

伍、考察期間

113年9月16日（星期一）至9月20日（星期五），共5日。

附錄 2【考察議題訪綱】

壹、電腦化測驗參訪議題

(中文)

一、說明

中華民國憲法將考試權獨立於行政權之外，明定中華民國公務人員之任用資格與專門職業及技術人員之執業資格均應通過國家考試。考選部承擔國家考試行政任務，為我國獨一無二的考選行政機關。

國際教育測驗發展趨勢，電腦化測驗對人才招募及能力評鑑有實質推動之意義；考選部自93年起推動電腦化測驗，陸續推動15種專門職業及技術人員類科採行電腦化測驗，本次規劃赴日本考察推動電腦化測驗之測驗機構，藉由學習國外電腦化測驗考試架構與方法，強化國家考試效度與效能

二、建議參訪議題

(一) 電腦化測驗考試方式與服務有哪些？如：

株式会社 CBT-Solutions 公司提供之考試方式與種類計有：CBT / IBT / Remoty AI+ / Online Test Center 等；

株式会社イー・コミュニケーションズ (e-communications 公司) 提供之考試方式與服務計有：CBT / IBT / MASTER CBT PLUS/ Remote Testing / Ultra Guardian / J-Testing ；

1. 請分享各種考試方式與服務之系統架構，包含：軟硬體設備環境與架構、網路架構。
2. 各種考試方式與服務屬於考場型 (<日>會場型、<英>test center) ?或是遠距型/居家應試型?。
3. 各種考試方式與服務適用之情境、優缺點與利弊分析。
4. 各種考試方式與服務之穩定性、備援性、安全性、公平性。
 - (1)如何確保試題機密性?如何確保考試環境安全性?如何確保作答資料不遺失?如何確保作答資料不可否認性?系統備援與備份、網路安全與應用 (有線網路、無線網路)、資安防護、個資保護等議題如何處理。
 - (2)可以支援之考試最大規模為何?如何驗測大規模考試壓力與瓶頸?如何因應?
 - (3)因應偶發事件之配套措施為何 (如應試設備故障、網路異常、電力中斷、地震與颱風等天災)?可否分享其標準作業程序 (SOP)。
5. 各種應試系統功能與特色，包含試題型式 (如多媒體)、作答介面 (單選、複選、簡答、配對、申論等、支援多國語言?支援製表與繪圖等)、內建輔助工具 (如計算機、法條查詢等) 與外接輔助工具 (如手

寫板等) 有哪些, 是否可提供應考人應試系統模擬練習網站與操作說明。

6. 各種考試試務作業流程: 如: 開始辦理考試、應考人報名、排定應試地點、應考人到考、監試作業、成績回傳、成績計算與榜示、考試結束、統計資料分析、試題品質分析等程序。
7. 各種考試方式與服務之監試作業:
 - (1) 監試方式與範圍, 採用之技術。
 - (2) 應考人身分驗證程序與方法。
 - (3) 是否有監試系統? 請分享監試系統系統架構, 包含: 軟硬體設備環境與架構、網路架構。
 - (4) 監考人力評選標準及培訓方式。
 - (5) 監考人員擔任的工作內容與作業方式。
 - (6) 監考人員聘用與評鑑。如何確保監考人員品質?
8. 採用機構 (<日>試驗團體、<英>test owner) 選擇不同考試方式與服務之關鍵要素。
9. 應考人 (<日>受験者、<英>test taker/examinee) 選擇不同考試方式與服務之關鍵要素。
10. 所實作之考試清單 (<日>試験一覽)
 - (1) 考試清單列表 (舉例): 包含考試名稱、運用之考試方式 (如 CBT 或 IBT)、採用機構 (test owner) 名稱與性質 (如公家機關、職業團體、教育機構等)、推動歷程、施測對象與規模 (應考人數)、考試性質 (如競爭型考試、能力認證考試)、應考人參與該考試之目的與收穫等。
 - (2) 是否有應考人提出訴願或對考試結果提出懷疑 (如非其作答資料、內容、成績之爭訟?) 若有, 可否舉例說明處理方式、提出之客觀佐證及答辯資料
 - (3) 請就各種考試方式與服務, 各舉一個最具代表性或最為廣泛應用之考試, 並分享其特色。

(二) 考場型 (<日>会場型/、<英>test center) 電腦化測驗考試環境:

1. 考場規模: 全國考場分布情形、考場數量與規模、同時可支援之應考人數等。
2. 考場認證規範
 - (1) 認證程序。
 - (2) 認證標準 (如電力供應、網路穩定、考試安全)。
 - (3) 認證有效期限? 再認證要求與程序? 是否有定期認證規範?
 - (4) 考場建置 (安裝設定) 步驟。
 - (5) 考場檢測程序 (確認考場符合標準)。
 - (6) 權利義務關係: 是否簽訂契約? 權利義務內容? 罰則?

(7)考場維運與管理規範。

(8)支援辦理考試之酬勞與費用。

3. 技術人力之背景及人數需求。
4. 監場人力之背景及人數需求。
5. 考試環境準備程序（如多久以前辦理環境準備與測試）。如何確定考試環境正確性?是否需逐考試安裝設定?是否針對特別應考人執行個人化設定嗎（如付費輸入法安裝等）?
6. 應試系統異動或程式改版之處理程序。
7. 因應政策推動或技術更新（如作業系統廠商不再支援舊版作業系統），須辦理系統更新與測試程序。

(1)全面性辦理所有考場作業系統升級或改版、應試系統或工具軟體更新之作業程序。

(2)如何確保異動或改版辦理完成且正確。

(三)線上閱卷服務：申論題部分。

1. 對線上閱卷的看法（如優缺點、提升試務作業效率及優化考試效度等面向）?
2. 對線上閱卷運用人工智慧自動評閱的看法?
3. 可否分享現有或規劃中線上閱卷之系統架構（包含硬體設備環境與架構、網路架構）、已推動或規劃推動之規模與應用範圍，以及對評閱品質之控制（如不同老師間、同一老師不同時間具一致性評閱標準）。

(四)新興科技考試方式之推動與規劃

1. 是否提供或規劃新興科技考試方式，如 BYOD（Bring Your Own Device）、BOOD（Bring Our Own Device）、BYOP（Bring Your Own Proctor）等。
2. BYOD/BOOD 網路架構採有線網路或無線網路連線的考試安全性與穩定性。
3. 多元應試環境（如各種電腦廠牌、各種作業系統與版本等）下，BYOD/BOOD 考試可用性（適用所有多元環境）與穩定性之想法與構思。

(五) AI 在考試上之應用：如智慧型監考、試題內容調校等。

1. 是否提供或規劃 AI 在考試上之應用，其服務項目、功能及系統架構（包含硬體設備環境與架構、網路架構）為何。
2. 系統的正确性與可用性。
3. 應用後之效益與成果（實例或預期）。

日本語

一、説明：

中華民国憲法では、試験権は行政権から独立し、公務人員の任用資格や専門職及び技術者の資格は国家試験に合格する必要がある、と明記されている。考選部は、国家試験の行政責務を担う、我が国における唯一無二の試験行政機関である。国際的な教育試験が発展する中で、パソコン受験を推し進めることは人材募集と能力評価においては実質的な意味がある。2004年から、考選部はパソコンでの受験を開始し、現在15種類の専門職及び技術者の試験に導入している。今回は国家試験の効果や効率を強化するため、日本におけるパソコン試験の導入試験機関の視察を計画し、海外のパソコン試験の枠組みや方法を学んでいきたい。

二、視察の議題：

(一) パソコンでの試験方式及びサービスについて

株式会社 CBT-Solutions 公司では CBT / IBT / Remoty AI+ / Online Test Center 等の試験方式やサービスを提供。

株式会社イー・コミュニケーションズ (e-communications 公司) では CBT / IBT / MASTER CBT PLUS/ Remote Testing / Ultra Guardian / J-Testing 等の試験方式やサービスを提供。

1. 各種試験のソフト・ハード設備、作動環境等のサービスシステムの枠組みを教えてください。
2. 各種試験は会場型（テストセンター）、遠距離型、または在宅型か。
3. 各種試験の適用環境、またはそのメリットとデメリット。
4. 各種試験の安定性、不可抗力による再試験の可否、安全性と公平性について。

(1)試験問題の機密性、試験環境の安全性をどのように確保しているのか。また、解答用紙の紛失や解答用紙の否認防止をどのように確保するのか。システムの復元やバックアップ、ネットセキュリティと応用（有線/無線ネットワーク）、情報セキュリティ、個人情報保護などはどのようにしているのか。

(2)対応できる試験サービスの最大規模はどれぐらいなのか。また、大規模な試験に伴うプレッシャーやボトルネックなどには、どのように対応しているのか。

(3)突発事件が発生した場合の対応策としてどのようなものがあるか。（例えば、試験設備の故障、ネット回線の異常、電力供給の中断、地震や台風などの天災）標準作業手順をご教示いただけ

ないか。

5. 各種試験システムの機能や特色を教えてください。例えば、試験の方式（マルチメディア）、回答のインタフェース（単一回答、複数回答、簡潔回答、マッチング、記述、多言語対応、図表やグラフの作成など）、内蔵された補助ツール（電卓や法規条文の検索など）、及び外部ツール連結機能（手書きパッド）等の機能搭載状況。また、受験システムのシミュレーションサイトや操作マニュアルの提供は可能か。
6. 各種試験の事務作業プロセスについて。試験の準備、受験申込、試験場手配、当日の受付、試験監督作業、成績の転送、採点及び公表、試験終了、統計資料の分析、試験内容の品質分析など。
7. 各種試験の監督作業：
 - (1)監督方法及び範囲、搭載している技術。
 - (2)受験者の本人確認プロセス及び方法。
 - (3)監視システムが搭載されているのか。監督システムのソフト及びハード設備、ネット環境などの構造教えていただきたい。
 - (4)試験監督の人選選びの基準と育成方法。
 - (5)試験監督の仕事内容及び作業方法。
 - (6)試験監督の採用と評価はどのように行われ、試験監督の質をどのように確保しているのか。
8. 試験団体が異なる試験方式やサービスの利用を決める主な理由は。
9. 受験者が異なる試験方式やサービスを選ぶ主な理由は。
10. 試験一覧
 - (1)試験リスト：試験名、試験方式（CBT か IBT）、試験団体名称や性質（公的機関、職業団体、教育機構など）、推進プロセス、受験対象や規模（受験者数）、試験の性質（競争型試験、能力認証試験）、受験者の参加目的及び収穫等。
 - (2)これまで受験者から提訴、或いは成績疑義の申請を受けたことがあるか（解答用紙、内容、成績の訴訟等）。もしあれば、どのように対処したか、可能であれば客観的な証明及び答弁資料の共有をお願いしたい。
 - (3)あらゆる試験方式とサービスがある中で、最も代表的な、または一番広く利用されている試験を例に挙げ、その特徴を教えてください。

(二) 会場型パソコン試験での試験環境：

1. 試験場の規模：全国試験場の分布、試験場の数量と規模、同時に対応できる受験者数等。
2. 試験場の認定規定

- (1)認定プロセス
 - (2)認定基準（電力の供給、安定したネットワークの接続、試験時の安全）
 - (3)認定の期限。再認定の要求及びプロセス。定期的に認定が必要な規定があるのか。
 - (4)試験場の設置（セッティング）のステップ
 - (5)試験場の検査プロセス（試験場が基準に合っているかどうかの確認）
 - (6)権利と義務：契約を交わしているのか。権利と義務の内容は。罰則があるのか。
 - (7)試験場のメンテナンス及び管理規定。
 - (8)試験開催の協力費と関連料金。
 - 3. 技術担当のバックグラウンド及び必要人数
 - 4. 試験監督のバックグラウンド及び必要人数
 - 5. 試験環境準備プロセス（どれくらい前から環境を整え、テストを行うべきなのか）。また、試験環境が正常か否かはどのように確認するのか。関連のセッティングは試験ごとに必要なのか。特殊な需要がある受験者向けにカスタマイズ設定を行っているのか。（課金で入力ツールの導入）
 - 6. 試験システムの変更、またはバージョンアップの処理プロセス
 - 7. 政策推進または技術更新（オペレーティングシステムメーカーが既存バージョンシステムを支援しないなど）に伴い、システムの更新とテストを行っていく。
 - (1)全試験場のオペレーティングシステムのアップデートやバージョン変更、試験システム、ソフトウェアツールの更新作業のプロセス。
 - (2)どのようにして移行やバージョン変更が正しく完成しているかを確保するのか。
- (三) オンライン採点サービス：記述問題
- 1. オンライン採点に対する考え方（メリット、デメリット、そして試験事務の作業効率向上、試験効率の最適化等）。
 - 2. オンライン採点において人工知能を駆使して自動的に行うことに対する考えは。
 - 3. 既存、または既に計画中のオンライン採点システムの枠組み（ハード設備環境や枠組み、ネットワークのフレーム）、既に推進済みか推進を計画中の規模や応用範囲、評価の質をどのようにしてコントロールをしているのか（例えば、異なる先生同士、または同じ先生の異なる時間帯の採点が一致する基準）の共有をお願いしますでし

ようか。

(四) 新たな試験方式の推進と計画

1. BYOD (Bring Your Own Device)、BOOD (Bring Our Own Device)、BYOP (Bring Your Own Proctor) など、新たな試験方式を提供している、或いは計画中か。
2. BYOD/BOOD のネットサービスで、有線または無線のネット回線を使用している試験の安全性と安定性。
3. 多様な試験環境 (パソコンのブランド、各種のオペレーティングシステムとバージョン) のもと、BYOD/BOOD 試験の適用性、安定性に関する考え。

(五) 試験における AI の利活用：スマート型試験監督、試験内容の調整等。

1. 試験に AI を導入しているか、または計画しているか。そのサービス項目、機能、システムの枠組みはどのようなものなのか。(ハードウェア設備、環境と枠組み、またネットワークのフレーム)
2. システムの正確さ、可用性。
3. 利活用後の効果及び成果。(実例、または予測。)

貳、RECRUIT MANAGEMENT SOLUTIONS 公司參訪議題

(中文)

一、「SPI3」的試題命擬方式與流程相關問題

1. 試題的命製作歷經哪些過程呢？可否提供簡要的流程圖？
2. 人類的能力包羅萬象，單一測驗不可能全部納入，總要取捨。在基礎能力檢查中，如何區分「語言」與「非語言」問題，各自包含哪些能力種類？
3. 以上決策，是考量了哪些因素，或是根據什麼理論，或是參考什麼測驗來決定的呢？
4. 如何決定題目的良莠？例如：透過預試，取得參數？或是應考人回饋？或是用人機關回饋？
5. 如果有預試如何辦理？受試者的來源為何？如何防止試題外洩？
6. 使用過的題目，是否不再使用？或，使用多少次之後會有淘汰的機制？
7. 應考人去補習或練習坊間的模擬試題，是否對於 SPI3 考試造成影響或困擾？對於「應考人透過練習而增加考試的表現」貴公司的看法為何？

二、有關「SPI3」的命擬委員

1. 研發試題的人是該領域的學者專家（如教授）嗎？還是公司內部的研發人員？
2. 公司內部的研發人員，有哪些專業背景？如：數學系？英文系？心理系？測驗專長？...等
3. 是否能夠與具有命題經驗的研發人員直接對話？

三、有關「SPI3」的運用

1. 企業或機關使用 SPI3 之後有哪些回饋？使用 SPI3 測試挑出來的人，是他們預期要的人選嗎？如果有些差距，SPI3 可以做什麼調整呢？
2. 對於施測的結果進行哪些分析？除了項目分析（item analysis）之外，有沒有其他質性或量化分析（例如與應考人的背景變項做交叉分析、應考人的測施滿意度問卷等）？
3. SPI3 與政府採用的基礎能力測驗或新教養測內容架構相仿，近年來地方自治體有逐漸改採 SPI3 的趨勢，請問 SPI3 的優勢有哪些？
4. 是否有依據地方自治體的用人需求，而客製化試題設計與內容？
5. 其他：性格檢查也是一樣的問題：人類的性格有許多面向，SPI3 納入哪一些項目，是考量了什麼呢？

日本語

一、「SPI3」試験問題の出題方法及び流れに関する質問

1. 試験問題の選定の流れは？簡単なチャートがあれば、ご提供願いたい。
2. 人間の能力は多種多様で、単一の試験ですべてをカバーすることは不可能であり、取捨選択が欠かせない。基礎能力の検査の中で、「言語」と「非言語」の問題をどのように区別し、各々どのような能力を含んでいるか。
3. 以上の選定は、どのような要素や理論、または実際の試験を参考にしているのか。
4. 設問の良し悪しをどのように判断しているのか。予備テストを通して定数を取得しているのか。受験者または採用機関によるフィードバックなのか。
5. 予備テストはどのように開催しているのか。どのように受験者を応募し、試験問題の漏洩をどう防いでいるのか。
6. 使用済みの試験問題は二度と使わないのか。または、何回使用すると、淘汰されるのか。
7. SPI3 試験は、受験者が塾に通ったり、世に出ている模擬試験の練習をすることで影響、または障害があるか。「受験者が練習を通して試験のパフォーマンスを向上させていること」について、貴社はどのように思われているのか。

二、「SPI3」の出題委員について

1. 出題者は当該分野における学者や専門家なのか（教授等）。または、会社内部の研究開発者なのか。
2. 社内の研究開発者はどのような専門的バックグラウンドを持っているのか。数学、英語、心理学などか。また、試験測定の専門はいるか。
3. 出題経験のある研究開発者と直接お会いして交流することは可能か。

三、「SPI3」の運用について

1. 企業や機関が SPI3 を利用したのち、どのようなフィードバックがあったか。また、SPI3 を通して採用した方は、希望通りの人選だったか。誤差があるのであれば、どのように SPI3 を調整できると思うか。
2. 試験結果をどのように分析しているのか。項目分析（item analysis）のほか、質的や量的分析を行っているのか。（例えば、受験者のバックグラウンドの変数をクロス分析にかける、受験者の試験に対する満足度調査アンケートなど）
3. SPI3 と政府が採用している基礎能力試験、新教養試験の内容や枠組みには近い部分があり、近年地方自治体でも SPI3 を取り入れる傾向が見られるが、SPI3 のメリットは何か。

4. 地方自治体の雇用ニーズを踏まえて、出題の形式や内容をカスタマイズしているのか。
5. 性格検査についても伺いたい。人間は個性豊かだが、SPI3 ではどういった面を取り込み、また何を根拠としているのか。

參、人事院參訪議題

(中文)

一、「知能領域」試題命擬方式與流程相關問題

1. 在日本中央及地方自治體的公務人員考試第一試筆試中的「知能（ちのう）」領域測驗，其施測目的為何？
2. 「知能領域」中，有關文章理解、判斷推理、數的推理與資料解釋等試題的命題藍圖如何規劃？可否提供範例試題供參考？
3. 前開考試內涵與公務工作的職能之間的關聯性為何？為何是這些項目而非其他能力項目呢？
4. 前開內涵的挑選及發展演變為何？試題的取材來源為何？
5. 「知能領域」的試題，在高中畢業、大學畢業與研究所畢業可報考的考試，是否依考試等級，區分相對應程度的試題？
6. 承上，命題者是否依考試等級區分命題內容？或是一開始並無區分等級，而是命題完成後，於辦理選題與組卷時，有一套區分等級的判斷標準？
7. 如何決定「知能領域」與「知識領域」在試卷中的比重？更進一步，各種細項的比重如何決定？（請參考表 1，例如「判斷推理」的題數介於 2~6 題，有什麼特別的考量？）
8. 如何決定題目的良莠？例如：透過預試，取得參數？
9. 如果有預試如何辦理？受試者的來源為何？如何防止試題外洩？
10. 使用過的題目，是否不再使用？或，使用多少次之後會有淘汰的機制？
11. 應考人去補習或練習坊間的模擬試題，是否對於「知能領域」考試造成影響或困擾？對於「應考人透過練習而增加考試的表現」政府的看法為何？

二、有關「知能領域」命擬委員

1. 相關命題者的來源為何（內部與外部）？
2. 外部的命題者在命題之前有施予什麼培訓嗎？培訓的內容與方式為何？
3. 試題研發人員，有哪些專業背景？如：數學系？英文系？心理系？測驗專長？...等
4. 是否能夠與具有命題經驗的人員直接對話？

三、有關「知能領域」試題的運用

1. 試題施測的方式為何？係採取集中電腦化測驗、紙筆作答，抑或其他方式？施測效果有無差異？
2. 對於施測的結果進行哪些分析？除了項目分析（item analysis）之外，有沒有其他質性或量化分析（例如與應考人的背景變項做交叉分析、應考人的測施滿意度問卷等）
3. 一般而言，第一試專業科目表現較優的應考人，是否其基礎能力測驗

（新教養測驗）也表現較優？兩者之間的關係是否為正相關或不一定？
是否期待兩者有中度相關？

4. 承上，「知能領域」與「知識領域」的表現，是否為正相關或不一定？是否期待兩者有中度相關？
5. 承上，「知能領域」中，「數的處理」與「文章理解」的表現，是否為正相關或不一定？是否期待兩者有中度相關？
6. 此測驗工具是否能有效排除不適合擔任公務人員的應考者？
7. 採用機構？

備註：希望與實際編製文章理解、判斷推理、數的推理與資料解釋等測驗試題之命題者面對面交流，交換意見。

表1：

試験種	一般知能分野							一般知識分野										合計出題数(問)	合計解答数(問)	解答時間(分)					
	数的処理			文章理解				人文科学				自然科学			社会科学						時				
	数的推理	判断推理	空問把握	資料解釈	現代文	英文文	古文	世界史	日本史	地理	思想	文学・芸術	数学	物理学	化学	生物学	地学					法律	政治	経済	社会
国家一般職	5	6	2	3	6	5	—	1	1	1	1	—	—	1	1	1	—	1	1	1	—	3	40	40	140
国税・財務・労基	4	7	2	3	6	5	—	1	1	1	1	—	—	1	1	1	—	2	—	1	—	3	40	40	140
裁判所一般職	6	6	4	1	5	5	—	1	1	1	1	—	—	1	1	1	1	1	2	2	—	—	40	40	180
特別区I類	6	5	4	4	5	4	—	1	1	1	1	—	—	2	2	2	2	1	2	—	1	4	48	40	120
東京都I類B*	6	2	4	4	4	4	—	1	1	1	—	1	—	1	1	1	1	1	1	1	—	5	40	40	130
地方上級 全国型	6	5	4	1	3	5	1	2	2	2	—	—	1	1	2	2	1	3	—	3	6	50	50	150	
地方上級 関東型	6	3	2	1	3	5	1	3	3	2	—	1	1	1	2	2	1	3	1	3	6	50	40	120	
横浜市*	7	4	2	2	5	5	—	1	2	2	—	—	1	—	1	1	1	8	3	7	8	60	60	170	
国立大学法人等	4	4	4	1	3	3	1	2	2	2	—	1	1	1	1	2	1	2	1	2	2	—	40	40	120

※ゴシックの数字は必須解答です。 ※上記は、過去の試験情報を元に作成しています。

日本語

一、「一般知能」における出題方法またはプロセスに関する質問

- 1.日本における中央及び地方自治体の公務員試験の一次筆記試験の中にある「知能」分野の試験項目の目的は？
- 2.「一般知能」における文章理解、判断推理、数的推理、資料解釈等の出題の方向性はどのように計画しているのか。また、関連する出題例を共有していただけないでしょうか。
- 3.前述した試験の内容と公務の職能との関係性は。出題はなぜこうした能力項目で、他の能力ではないのか。
- 4.前述した試験の内容はどのように選び、どのように発展してきたのか。また、出題のよりどころは。
- 5.「一般知能」の出題内容は、高卒や大卒、院卒で受けられる試験について、試験のレベルまたは相応なレベルの出題をしているのか。
- 6.以上を踏まえ、出題者は試験のレベルに基づいて出題をしているのか。または最初からレベル分けはなく、出題完成後出題の選択時にレベル分けの判断基準を設けているのか。
- 7.「一般知能」と「一般知識」の出題比重をどのように決めているのか。また、細かい項目の比重をどのように決めているのか。(図1参照。例えば、「判断推理」の出題数が2~6問になっているが、何か特別な考えがあるのか。
- 8.設問の良し悪しをどのように判断しているのか。予備テストを通して定数を取得しているのか。
- 9.予備テストはどのように開催しているのか。どのように受験者を応募し、試験問題の漏洩を防いでいるのか。
- 10.使用済みの試験問題は二度と使わないのか。または、何回使用すると、淘汰されるのか。
- 11.受験者が塾に通ったり、世に出ている模擬試験の練習をすることで「一般知能」の試験に対する影響あるいは障害があるか。「受験者が練習を通して試験のパフォーマンスを向上させていること」について、政府はどのような考えをお持ちか。

二、「一般知能」の出題委員について

1. 出題関係者はどこから募集しているのか。(内部と外部)
2. 外部の出題者は出題前に、何か研修や育成を行っているか。また、研修方法と、内容は？
3. 出題の研究開発者はどのような専門バックグラウンドを持っているのか。数学や英語、心理学など、また、試験測定の専門はいるか。
4. 出題経験のある方と直接会って交流をすることは可能か。

三、「一般知能」の出題の応用について

1. 試験方式は。テストセンター方式、筆記試験、または他の試験方式を採用しているのか。効果に違いはあるのか。
2. 試験結果をどのように分析しているのか。項目分析 (item analysis) のほか、質的や量的分析を行っているのか。(例えば、受験者のバックグラウンドの変数をクロス分析にかける、受験者の試験に対する満足度調査アンケートなど)
3. 一般的に、一次試験の専門科目で良い成績を出した受験者は、基礎能力試験 (新教養試験) でも良い成績を出せるのか。両者の間にプラスの相関が成り立つのか、または中程度の相関が成り立つのか。
4. 上述した「一般知能」と「一般知識」の成績には正の相関が成り立つのか、または一定の中程度の相関が見込まれるのか。
5. 上述した「一般知能」における「数的推理」と「文章理解」には正の相関が成り立つのか、または一定の中程度の相関が見込まれるのか。
6. この試験で公務員に適していない受験者を有効的に排除することができるか。
7. この試験を導入している機関は？

実際に文章理解、判断推理、数的推理、資料解釈等の出題者と対面交流し、意見交換を希望。

図1：

試験種	一般知能分野							一般知識分野										合計出題数(問)	合計解答数(問)	解答時間(分)					
	数的処理			文章理解				人文科学				自然科学			社会科学										
	数的推理	判断推理	空間把握	資料解釈	現代文	英文	古文	世界史	日本史	地理	思想	文学・芸術	数学	物理学	化学	生物学	地学				法律	政治	経済	社会	時事
国家一般職	5	6	2	3	6	5	-	1	1	1	1	-	-	1	1	1	-	1	1	1	-	3	40	40	140
国税・財務・労基	4	7	2	3	6	5	-	1	1	1	1	-	-	1	1	1	-	2	-	1	-	3	40	40	140
裁判所一般職	6	6	4	1	5	5	-	1	1	1	1	-	-	1	1	1	1	1	2	2	-	-	40	40	180
特別区I類	6	5	4	4	5	4	-	1	1	1	1	-	-	2	2	2	2	1	2	-	1	4	48	40	120
東京都I類B*	6	2	4	4	4	4	-	1	1	1	-	1	-	1	1	1	1	1	1	1	-	5	40	40	130
地方上級 全国型	6	5	4	1	3	5	1	2	2	2	-	-	1	1	2	2	1	3	-	3	6	50	50	150	
地方上級 関東型	6	3	2	1	3	5	1	3	3	2	-	1	1	1	2	2	1	3	1	3	6	50	40	120	
横浜市*	7	4	2	2	5	5	-	1	2	2	-	-	1	-	1	1	1	8	3	7	8	60	60	170	
国立大学法人等	4	4	4	1	3	3	1	2	2	2	-	1	1	1	1	1	2	1	2	1	2	40	40	120	

※ゴシックの数字は必須解答です。 ※上記は、過去の試験情報を元に作成しています。

附錄 3 【考察行程表】

行程表

日期	9/16 (一)	9/17 (二)	9/18 (三)	9/19 (四)	9/20 (五)
上午	11:35 抵達		10:00-12:00 e-communications 公司	10:30-12:00 人事院	
下午		12:00-14:00 CBT-Solutions 公司 17:00-17:30 代表處	14:30-16:30 Prometric Japan公司	14:00-15:30 Recruit-MS公司	12:40 返程
晚上		18:00-20:00 工作餐敘			

第一天

日期	時間	行程說明	備註
113.9.16	04:40	考選部出發	行政大樓 1 樓集合
	05:20	臺北松山機場TSA (航廈1) 長榮櫃台報到及託運行李	秘書處江處長宗正、林羽 萱科長送機
	06:00	長榮櫃台集合	次長抵達，托運行李後， 秘書處一行人帶領次長通 關禮遇及前往貴賓室休息
	07:20	搭機飛往日本	Eva Air (BR192)
	11:35	抵達東京羽田國際機場	接機人員：駐日代表處
餐食	早餐：松山機場 午餐：自理 晚餐：自理		
住宿	HOTEL THE CELESTINE TOKYO SHIBA (東京芝賽萊斯廷酒店) 〒105-0014東京都港區芝3-23-1 TEL：03-5441-4111		

第二天

日期	時間	考察行程說明	備註/出席者
113.9.17	下午 12:00- 14:00	CBT-Solutions公司 東京都千代田区神田練 塀町3 AKSビル6階	1.11:20從飯店出發 2.拜會對象： ● 取締役 佐々木 元昭 ● 執行役員兼營業統括本部部 長 大倉 俊輔
	下午 17:00- 17:30	拜會代表處 東京都港區白金台5-20- 2	1.16:30從飯店出發 2.蔡公使明耀主持
	晚上 18:00- 20:00	工作餐敘 北大路品川茶寮	
餐食	早餐：飯店 午餐：自理 晚餐：餐敘		
住宿	HOTEL THE CELESTINE TOKYO SHIBA（東京芝賽萊斯廷酒店）		

第三天

日期	時間	考察行程說明	備註/出席者
113.9.18	上午 10:00- 12:00	e-communications公司 東京都港区六本木7丁 目15-7 新六本木ビル SENQ六本木704	1.09:30從飯店出發 2.拜會對象： ● 取締役 營業本部長 佐藤 直 人 ● 營業本部 營業部 嶋 真乃
	下午 14:30- 16:30	Prometric Japan公司 東京都千代田区神田駿 河台4-6 御茶ノ水ソラシ ティ アカデミア5F	拜會對象： ● Vice President of Prometric, General Manager of Prometric Japan（Jim Hagenbucher） ● Managing Director,Asia Paciific（Humphrey Chan） ● Chief Product and Technology Officer（Kevin Pawsey） ● 取締役副社長 二本木 威秀
餐食	早餐：飯店 午餐：自理 晚餐：自理		
住宿	HOTEL THE CELESTINE TOKYO SHIBA（東京芝賽萊斯廷酒店）		

第四天

日期	時間	考察行程說明	備註/出席者
113.9.19	上午 10:30- 12:00	人事院 於日本台灣交流協會（東京都港区六本木 3-16-33 青葉六本木ビル七樓） **經由日本台灣交流協會安排，於該協會東京考選部與人事院進行交流	1.10:00從飯店出發 2.拜會對象： -- 交流協會總務部副長 紺野紗佑里（KONNO Sayuri）主持 ● 秋庭能久（AKIBA Yoshihisa）人事院人材局首席試驗專門官 ● 竹中繪理（TAKENAKA Eri）人事院人材局試驗專門官 ● 諸星渚（MOROHOSHI Nagisa）人事院人材局試驗專門官室係員 ● 小島遙（KOJIMA Haruka）人事院人材局試驗專門官室係員 ● 中村るり（NAKAMURA Ruri）人事院事務総局国際課国際研究官 ● 柏木綾子（KASHIWAGI Ayako）人事院事務総局国際課係員
	下午 14:00- 15:30	Recruit Management Solution 東京都港区芝浦3-16-16 住友不動産田町ビル東館 4	拜會對象： ---- 出席實體會議2人 ● 行銷部經理 有本倫子 ● 研究員 Nita 先生 ---- 線上參與2人 ● 行銷部 Kowa 先生 ● 系統部門 Hirota 先生
餐食	早餐：飯店 午餐：自理 晚餐：自理		
住宿	HOTEL THE CELESTINE TOKYO SHIBA（東京芝賽萊斯廷酒店）		

第五天

日期	時間	考察行程說明	備註/出席者
113.9.20	上午 10:00	出發前往成羽田機場	10:00從飯店出發 送機人員：駐日代表處
	10:40	羽田機場第三航廈	
	下午 12:40	搭機返台	Eva Air（BR191）
	下午 15:05	臺北松山機場TSA（航廈1）	秘書處林羽萱科長接機
餐食	早餐：飯店 午餐：自理 晚餐：自理		

附錄 4 【考察座談紀錄】

CBT-Solutions 公司 訪談紀錄

時間：2024年9月17日中午12時

地點：CBT-Solutions 公司

日方：佐々木 元昭 取締役兼人事部部長 兼日本の資格・検定事業部部長
大倉 俊輔 執行役員 兼 営業統括本部 部長

以下由大倉先生進行簡介

CBT-Solutions 公司提供4種考試服務：(1) CBT (會場型)、(2) OLTC Online Test Center (線上會場, 有監試人員)、(3) Remoty AI+：全程影像錄畫, AI 判定應考人有無作弊, 採用雲端支援、(4) IBT：適用模擬考試。就嚴謹度觀之, 由輕至重依序為 IBT、Remoty AI+、OLTC Online Test Center、CBT。

IBT：只要有手機、電腦、平板皆可考試, 主要是資格考試模擬試驗, 沒有 AI 監試, 最不嚴謹。

Remoty AI+：全程錄影、錄影資料上到雲端, 再用 AI 判定 (判斷三種：判斷是否多於一位人員在考試現場、依據視線與動作或手機電腦的位置判別有無作弊、是否與報考人同一人), 資格考運用少, 此種考試運用先端技術, 尚屬發展階段, 主要用於大學研究所 (MBA) 社會人士課程需要考試時使用。運用少之原因, 是因為大家對於 AI 技術達成之公平、正確、嚴謹性尚存有質疑。

OLTC Online Test Center：線上有監試官, 1位監試官對8位應考人, 可放大特定一位應考人的監試影像, 並且監試應考人電腦螢幕操作情況 (如:使用 google 搜尋引擎查閱資料), 同樣地運用少, 主要應用於公司內部升遷升職等考試, 目前合作的客戶有2間企業採用

CBT (會場型)：國家考試要求的安全性、公平性, 即是 CBT (會場型) 推動的重點, 主要應用於資格考試。目前有340個會場, 會場設置集中在都市, 主要是看供需情況。辦理考試之規模舉例：IPA 考試約30萬人, 漢語能力檢定一年有100多萬人 (但是使用 CBT 考試方式者約6萬), 金融規劃師考試約30萬人利用 CBT 考試。

(佐々木先生：日本技職考試由中央政府機關 (各個主管省廳) 管理與訂定, 如 IPA、ECEE 主管機關為經濟產業省、FP 主管機關為厚生勞動省。)

CBT (會場型)：除應用於資格考試外, 還有應用於工作進用或適性測驗。CBT 的應用市場有3個方向, 1.資格考、2.企業內部考試 (社內應用)、3.就職進用之適性考試, 以2020年新冠疫情為分水嶺, 2020前主要應用的為民間公司, 2020後國家考試也開始採用; 國家考試因為疫情的影響, 比較難採用一次式同時考試, 因此開始受到檢討, 也慢慢地開始採用 CBT 考試 (非採同一個時間一起考試); 國家考試若要採同一時間考試, 只能用紙筆測驗方

式，日本地方政府公務人員漸進地考慮採用 CBT 考試（如東京都廳），至於中央機關國家公務人員考試仍採用紙筆測驗，無法用 CBT 來考試。日本很難考的司法測驗（律師、檢察官、司法官），為公務人員考試中最難的考試，其主管機關法務省也開始檢討 2026年開始，該考試要採實驗性質方式辦理 CBT 考試，大約10,000-14,000人將試行 CBT 考試，所以 CBT 會是未來的主流方向。

楊科長：東京都廳公務人員採 CBT 考試的規模大約有多少人？屬於一次同時間考試？或是連續型考試？

大倉先生：東京都廳公務人員考試時間為連續型考試（一週之內），規模約數千人，一次考試時間為2小時。正確考試時間無法立即回答，但因為日本就職季節為每年11月到隔年2月（因4月要入職），因此應該是於該時段內參加考試。

楊科長：連續型考試（一週之內），試題不相同，是否有應考人反映公平性議題。

大倉先生：每個應考人試題不盡相同，係運用試題反應理論（Item Response Theory, IRT）來實作，因具備大量優質試題資料庫，有良好識別度。部分考試採用適性化測驗（Computerized Adaptive Test, CAT）方式來考試，但實務上較少用 CAT（因需要龐大資料庫）。

楊科長：申論題如何實作應試介面？在網站上提到漢字檢定會應用手寫板，可否分享技術上推動的難處與經驗？

佐々木先生：漢字檢定採平板及筆來回答，回答資料經處理後，以圖像方式送給閱卷者。漢字檢定僅須回答單一文字，一題一字，閱卷者判定一字為正確或錯誤，技術上並不困難，這個考試的申論題所回答的不是整篇文章。

楊科長：推動的試題類型？申論題型？測驗題型？其作答方式？

大倉先生：主要是選擇題，可為單選題及複選題；另外有申論題，採用打字方式進行。

楊科長：中文輸入法相當多元，請教日文在申論題答題的輸入上是否有推動上的困境？應考人要求使用特殊輸入法時，貴公司如何處理類似問題？

佐々木先生：日文僅一種輸入法，透過發音、輸入，轉換成英文、平假名、片假名，從來沒有想過“使用特殊輸入法”這個問題。

楊科長：CBT（會場型）3大項應用（1.資格考、2.企業內部考試（社內應用）、3.就職進用之適性考試）所占的比率？各施測時間（1.每天皆可參與考試、

2.某一段區間參與考試、3.所有應考人同時間考試）所占的比率？同時間考試的最大應考人數為何？

大倉先生：資格考（含檢定）占70%、企業內部考試（社內應用）占20%、就職進用/適性考試占10%；每天皆可參與考試占50%、某一段區間參與考試（期間限定）占30%、所有應考人同時間考試占20%；同時考多為醫學學會委託的專科資格認定考試，最多1,000人，平均大約幾百個人。

楊科長：同時間支援1,000人的原因與考量因素？有何瓶頸？

大倉先生：我們擁有340會場6,000席應試座位，可支援各種考試，一年有150萬人次參加考試。340會場係與當地教育機構或測驗企業簽約，該會場須辦理本業之課程或研究，故6,000席應試座位是無法同時間提供考試；另外，也需要考量一天之內要因應各種不同考試（如同時間會有醫事考試，也有其他考試同時進行），故平均最大可以支撐同時間考試的人數為1,000人。

楊科長：如何確保會場型電腦化測驗考試環境準備正確無誤？

大倉先生：CBT 使用 Windows 作業系統，應試系統不是網頁版，會場需要安裝應試系統，應考人只要在應試系統登入即可考試，不需要額外的環境準備。應試系統有連網與不連網兩種方式，不連網可以事前將考題下載下來，但技術上也有一題一題下載試題方式的考試。

楊科長：公司網站上分享應試架構在考試開始時需要下載檔案安裝，係下載安裝應試系統與試題？還是僅試題內容；另是否可以分享事前下在試題且不連網的架構，如何解決硬碟故障及答案回傳的問題？硬碟壞了怎麼辦？考完網路不連，是否有容許的時間可以上傳作答結果？

大倉先生：登入後即下載試題，之後不用管網路是否連網，最後一題答完後，送至雲端（區分為立即評閱與試後評閱兩種），上傳完竣後立即刪除硬碟資料，不會有試題外洩之情況。作答結果儲存在硬碟，若當機無法考試，重開機後應繼續應試；另考後若無法連網，因答題結果儲存在硬碟中，可俟網路穩定後再上傳；近年來舉辦考試，沒有遇到檔案損壞無法讀取作答結果之案例，若真發生此種狀況，則需另案處理（如請應考人再次考試）。

歐陽科長：試題類型是否有限制？是否有相容性問題？

大倉先生：系統可以支援圖片、動畫（如醫事考試心臟跳動）等試題類型，若試題需搭配語音時，則應試環境需要特別處理。

歐陽科長：有無應考人作弊之情形、有無偵測之工具。

大倉先生：會場有監視器，考場有監場人員，一間教室約20個座位，監場人員可以充分監試，電子設備進入教室之前要放在保管區。平均一個月有1-2人

作弊。應試系統會鎖定微軟作業系統的工具列，考試時無法啟用其他軟體或上網搜尋。

鄭政務次長：國家考試在意公平性，包含形式上公平，同一個考試無法使用不同試題，且不得重複使用試題，故無法實作重複使用的題庫試題。請問日本在推動電腦化測驗過程中，是否遇到相同問題？另與社會溝通之程序為何？

佐々木先生：日本社會亦要求公平性，然新冠疫情讓 CBT 之推動快速往前發展 10 年，強迫大眾思考採用電腦化測驗考試。日本為颱風、地震（天災）發生頻率較多的國家，若一年僅舉辦一次考試，當遇到天災致一年一次考試無法進行，將影響應考人權益。

歐陽科長：貴公司配置之 IT 人力與研發機構人力？是否與測驗發展機構有所合作？

佐々木先生：公司大約 120 人，IT 開發部門人力占 1/6（20 人），目前公司代表傾向公司內部研發，但會透過研討等機會與外界學習互動、學習知識與技術。

佐々木先生：我們公司主要期望塑造電腦化測驗的文化，請問在臺灣國家考試紙筆測驗是否也會發生作弊的情形？如何防弊？

歐陽科長：紙筆測驗沒有安排梅花座，2 監場人員/試場，並使用無線電波偵測器不定時在教室偵測是否有異常作弊狀況，並有巡場主任。若應考人偷看，監場人員會現場阻止。

佐々木先生 分享：電腦化測驗每位應考人試題不一樣，無法作弊；紙筆測驗之公平性是形式上的公平性，監場人員可以選擇性看不到。至於電腦化測驗是技術上做到防弊，並以人才培育角度，提供隨時可以考試的機會，培育更多的人才。

e-communications 公司 訪談紀錄

時間：2024年9月18日上午10時

地點：e-communications 公司

日方：佐藤 直人 取締役 營業本部長

嶋 真乃 營業本部 營業部

以下由佐藤先生進行簡介

e-communications 公司業務始於辦理多益考試之法人委託，負責多益考試應考人報名、收費核銷及應考人資料庫管理。後該法人規劃將多益考試由紙筆測驗進階為電腦化測驗，開啟 e-communications 公司開發多益電腦化測驗應試系統大門，該多益考試電腦化測驗於2004年上線。惟該年在日本設置之會場數量不多，故多益推動的策略仍以紙筆測驗為主，不過此項合作開啟了 e-communications 公司與日本法人、機構推動電腦化測驗的契機。

目前 e-communications 公司提供會場型考試、透過網際網路考試（IBT）、遠端監試（RemoteTesting）及防作弊機制（ULTRA GUARDIAN）等服務；透過網際網路考試（IBT）可以在家施測，並透過攝影機監試，另防作弊機制可以禁止檢索相關資料，適度防止作弊行為。會場型之會場是借用學校或機構電腦教室來辦理考試，目前同時間考試仍為主流，不過將以電腦化測驗為趨勢。

IBT：有些公司需要員工取得某種資格，會採用此種考試方式，請員工於會議室中參加取得資格的考試，這個方式比起紙筆測驗之試題寄送、回收作答情形，試務作業簡易方便，許多公司內部會運用此種考試。IBT 亦常用於模擬考試，補習班或大學提供模擬考，可在家進行。模擬測驗採用 IBT，主要是因為模擬考僅是一種練習，測試是否有實力，非真正的考試，不擔心是否作弊。另外 IBT 之所以會蓬勃發展，主要是因為新冠疫情大家必須在家工作與學習，故若要考試，會選擇採用 IBT，舉例來說企業內部社內考試要定期進行，新冠疫情期間員工即可在家參加升等或升級考試。

遠端監試（RemoteTesting）考試當下使用資訊設備的鏡頭來錄影，考試結束後運用 AI 來判別是否作弊，考試當下並不會進行作弊之判別。IBT 可以加上 RemoteTesting 錄影監試服務，即有監考功能。透過鏡頭確認是否為本人，要求應考人提供身分證明文件，確認身分後才能進入下一個考試階段（同時開啟考試及監試功能），畫面右邊是應試畫面，左邊是錄影畫面，應考人自己可以查閱被錄製的影像。而所錄影的畫面，於考試結束後透過 AI 來分析。AI 分析的情境包含：考試場所是否有其他人在場？應考人是否離開鏡頭？應考人視線是否飄移或看其他地方（可能有查閱資料等作弊之嫌疑）。若 AI 發現有異常，會擷錄異常畫面做成報告，也會分析聲音，看看是否有人告知答案，所以是同時錄音錄影。擷錄異常畫面可以提供細部查閱錄影影像。AI 擷錄異常畫面的報告將提供給考試委託方，由委託方自行判斷

應考人是否作弊。e-communications 公司僅提供技術服務，不負責判定作弊行為之真偽。作弊與否的判定由考試委託方主政。

RemotTesting+Ultra Gurdian 適用於在家考試或會場考試，如資格考試，員工可能因工作忙碌無法到會場考，但要求此資格之公司或團體希望員工參加考試取得資格，則會讓員工在家也可以參加考試（當應考人無法去會場，但需要採用此種考試時，即會用電腦考試加上此兩種監試輔助功能）。如大學入學英文能力分班考試，或是企業升級考試。另多益即是採用此種方式考試，這種考試叫做“多益 IP 考試（TOEIC IP）”，專門提供給團體考試。分班考試若進入不符合能力的班級，或是分級後負責不適任之工作，對受測者來說因為不適任而成為一種風險，故施測者並不會想要作弊取得非能力之等級成績。目前多益企業內部的分級或分班考試，同時採用紙筆測驗及電腦化測驗。

目前我們會場型考試皆為民間性考試所採用，在日本，國家資格考試導入 CBT，多於 IT 資格考試（CBTS），目前 J-testing 沒有承作國家級的資格考試，主要係因 J-testing 試題自資料庫抓取，惟國家考試在意試題公平性與一致性，且需要在同時間施測，不適用 J-testing 期間限定之特色。J-testing 傾向提供民間、工作導向需求的資格考試。

楊科長：在日本網站上搜尋電腦化測驗考試時，J-Testing 占有排名，且網路上的資訊該系統是由 e-communications 公司開發，請問 J-testing 即是會場型考試嗎？J-testing 使用的應試系統與 IBT 底層（MASTER CBT/PLUS）相同嗎？請教 J-Testing 考試最大規模為何？一次同時間考試可支援的應考人數為多少？

佐藤先生：J-testing 一年施測之應考人次約15,000人次，較少辦理一次同時的考試，主辦的會場型考試多為期間限定（考試時間3天或一週等），目前沒有開放在家使用個人電腦來考試。

楊科長：J-testing 用 MASTER CBT/PLUS 應試系統嗎？其架構支援同時間考試之應考人數？

佐藤先生：目前在日本全國有190合作的會場，每間規模大約5-20席座位數，一天考試分為好幾班制，舉例一天可以輪8班，所以規模視考試舉行的情況及會場大小、可以輪的班次來看。

楊科長：J-testing 同時間一起開考，應考人數是否有限制？其壓力與瓶頸為何？

佐藤先生：e-communications 公司合作對象與推動考試性質並無同時間實施的需求，因為資格考試不需要同時間開考，應考人數主要依據會場可以提供的位數及可以服務的時間區段；至於目前日本其他同業推動同時間考試電腦化測驗的狀況，最大規模約5,000人。若要擴大該規模，需要再租借其他大學、機構場地。

楊科長：e-communications 公司 190個會場之總座位數？所有考試類型一年施測總人次？各類型約多少人次？

佐藤先生：一個會場最大100席座位數，因20席座位數，一天5個考試在進行，即可以容納100人次的考試。

楊科長：請說明可支援的作答類型（是否有圖表、多媒體等試題）

佐藤先生：測驗式試題可以自動評閱，簡答題可以判別文字者也可自動評閱，但申論題打字之內容則不能即時評閱。一個題目有數個可確定答案者也可自動評閱；應試畫面可分割，區分為試題區與答題區；可支援影音試題，試題可以放大。

歐陽科長：試題大小之限制？是否要先下載？

佐藤先生：目前採用影音試題比率很少，技術上一個考試1G 以下試題可以正常考試。考前自雲端先下載試題，考完上傳試題至雲端，並刪除硬碟中試題與作答資料。

楊科長：考前下載試題，其安全機制為何？

佐藤先生：試題有加密，拿到檔案無法查看試題

楊科長：若應考人反映：我看到試題與公告母版試題不相同，我作答的內容不對，遇到此問題有無建議？

佐藤先生：試題皆自資料庫出題，應考人試題皆不相同，沒有公布試題，所以沒有相關之問題。e-communications 公司主要是提供考試技術服務之公司，不直接面對應考人，應考人提列的問題皆須向考試主辦機關反映；若應考人質疑非應考人答題結果等問題，e-communications 公司可提供該應考人答題紀錄給主辦考試機關，由其回應應考人，所以 e-communications 公司目前沒有收到相關的反映事項；但提供考試服務的過程中，曾遇過應考人反映某試題內容有問題，此時也是由主辦考試機關告知該題目是否全部給分或不予計分。

鄭政務次長：期間限定施測多次的考試，每個人題目不相同，用何種方式決定應考人試題？

嶋女士：所推動的所有考試皆自題庫隨機取題，CAT 施作之技術很困難，故未實作 CAT。

鄭政務次長：每個人題目不相同，如何決定是否通過考試？是否有做試題難度之調整？還是據答對題數計分？

佐藤先生：自題庫隨機抽題；A 題庫每一題難易度相同，每個應考人自 A 題庫抽取的試題數也相同；B 題庫每一題難易度相同，每個應考人自 B 題庫抽取的試題數也相同；也就是說每個應考人試卷之試題難易度及題數都相同。通過標準由主辦考試機關決定，有些是以分數（答對題數），有些是答對多少題以上即合格。會針對難易度與答題結果提供試題分析給主辦考試機關決定，並由主辦考試機關決定及格方式。

楊科長：試題是否經過預試？

佐藤先生：試題由主辦考試機關委請專家學者命題，並提供試題難易度等參數；若主辦考試機關期望做預試，e-communications 公司可以提供預試服務，但由主辦考試機關提供50位參與預試之應考人；若預試結果與專家學者提供之參數一致，則該試題將儲存至題庫資料中；若不一致，則由主辦考試機關決定是否要調整參數或放棄該試題。

楊科長：請說明特殊會場電腦化測驗的應用狀況。就特殊會場採應考人自備電腦（BYOD）應試，其電力、網路及設備故障之因應措施為何？

佐藤先生：e-communications 公司並未採用 BYOD 為施測工具，此外，在日本，基於公平性考量，基本上沒有太多考試採用此種施測工具與方式；BYOD 之公平性考量，舉例來說，每個應考人的螢幕規格不同，在解析度、亮度上會影響試題解讀；如解析度不同，圖表與動畫呈現有差異，而亮度則會影響超音波圖片的解讀，這些皆會影響公平性。

楊科長：可否展示 Ultra Gurdian 實作結果？

佐藤先生：Ultra Gurdian 可一鍵檢核與安裝，先辦理筆電環境確認（如是否配備攝影機、麥克風），檢核過後，系統採全螢幕顯示，鎖住 Windows 開始功能鍵，並禁止擷取螢幕或截圖、禁止啟動瀏覽器搜尋資訊。

歐陽科長：請問有多少員工負責開發？

佐藤先生：負責 CBT 業務有6位員工

鄭政務次長：請分享電腦化測驗未來展望與新的計畫。

佐藤先生：在日本，考試會場不會越來越多，擴大會場已經到達瓶頸，主要是因為業界實體電腦教室越來越少（因為現在的趨勢是線上課程，非到電腦教室上課），大專校院亦少擴大建置電腦教室，多採網路及線上課程教學，不會提供電腦設備，故會場型考試在多家考試服務中心競爭之下，在日本的推動已趨於飽和，很難再加以擴大。現在電腦化測驗推動趨勢多為在家、公司內部或場地辦理電腦化測驗考試，因應此種考試方式，就遠端監控、善用 AI 推動監試功能及提供考試環境管控等機制將是 e-communications

公司技術發展的方向。日本大規模紙筆測驗考試，目前的趨勢亦規劃往電腦化測驗考試方向發展，因此 e-communications 公司也將致力於同時間考試施測系統的發展，努力提供穩定應試系統。

Prometric Japan 公司 訪談紀錄

時間：2024年9月18日下午2時30分

地點：Prometric Japan 公司

日方：Jim Hagenbucher Vice President of Prometric, General Manager of Prometric Japan（總公司副總裁、日本地區總裁）

Humphrey Chan Senior Vice President and Managing Director, Emerging Markets.（總公司資深副總裁、新興市場總裁）

Kevin Pawsey Chief Product and Technology Officer（首席產品和技術官）

二本木 威秀 取締役副社長 最高技術責任者

以下由 Mr. Chan 就 Prometric 整體公司的介紹

Prometric 擁有30多年考試服務的經驗，每年在全球180國家辦理約800萬人次的考試，總部設置於美國東岸馬里蘭州巴爾地摩市；8年前屬 ETS 分公司，後由 BPEA 收購，已不屬於 ETS；全球有2400員工，與450家類似 Recruit-MS 這種大規模客戶合作，全球會場皆可以當考場。會場的容量並沒有限制，主要原因是全球有好幾百個由 Prometric 自己建置與擁有的會場，另與全球5000個夥伴技術合作設置會場，該夥伴皆遵循 Prometric 的安全與技術規範，另自2020疫情之後，居家在線考試模式蓬勃發展，因此 Prometric 號稱考試的地點是沒有侷限的。

Prometric 大型客戶舉例如下：財金類：如美國註冊公認會計師（Certified Public Accountants, CPA）、財務規劃師（Chartered Financial Analyst, CFA，該考試一年舉辦3次，每次2星期，約有幾萬人參加）；IT 認證類：如 Intel、Adobe 等認證；教育類：如美國大學委員會（College Board）所主辦的學術能力評估測試（Scholastic Assessment Test, SAT，該考試平台及試題皆由 Prometric 提供，命題已運用 AI 技術輔助出題）；語言類考試：如托福。

以下由 Mr. Hagenbucher 就 Prometric Japan 公司的介紹

Prometric Japan 公司成立於1994，剛開始時承作國際性的考試：如托福考試、IT 認證考試（如 SUN、CISCO、微軟等），後與日本本土化企業合作，其第一個大客戶為證卷業協會（Japan Securities Dealers Association, JSDA），之後並開始建立全國的會場，隨之而來的是增加了與日本本土企業的大量合作（如 Recruit-MS 的 SPI3 考試）。目前 Prometric Japan 公司每年承作250萬人次的考試，是除了美國之外，擁有最大市場的子公司。目前 Prometric Japan 公司進入向國外發展的階段，將日本考試帶到全國各地，舉例來說，日本勞力引進的日語檢定考試（JFT Basic）。在日本，我們每天提供5000人次的考試服務，僅出現非常少量的錯誤，主要原因是因為重複辦理考試，故累積高質量的服務品質。

以下由 Mr. Chan 就亞太地區的介绍

亞太地區約有500位同事，提供的服務包含亞太地區產品開發、本地客戶服務等；亞太地區一年服務300萬人次的考試，屬全球成長最快的區域；另外，我們也因應客戶的需求，提供紙筆測驗服務，一年大約提供15萬個紙筆考試服務，大部分為中國地區，可能是因為東亞文化的特性，偏重於同時考試方式，我們在這個領域也有很多經驗，但近期的考試趨勢，已漸漸由紙筆測驗轉成電腦化測驗。此外，我們擁有培訓監場合格人力約2500位，培訓過程亦包含背景調查及定期訓練。

以下由 Mr. Hagenbucher 就 Prometric Japan 公司的介绍

日本地區的考試服務是1991年開始，我們現在有169位全職員工，分布在全國各地（大部分以東京為基地），我們訂定各項標準作業模式，有200會場，每年舉辦250萬人次的考試，我們的服務通過多項國際認證，並也符合日本標準。日本是一個簡單的國家，非常要求高品質服務，我們擁有5個大型國家考試的合約，提供服務的領域則包含 IT 考試、卡車專業司機考試、飛行員考試等考試，這些考試我們都與客戶多方配合，從無到有，從紙筆測驗轉成電腦化測驗考試，另外我們有心理測評師團隊（教育測驗專家），協助客戶命題，並提供考試服務。我們總公司有許多會講中文的員工，此外，我們考試服務的對象也在財金、醫事等領域慢慢涉入，日本分公司是 Prometric 先驅者，服務質量非常好，也做了很多本土化服務的努力。

以下由 Mr. Pawsey 就 Prometric 技術層面的介绍

我在 Prometric 服務20幾年，負責系統開發與技術服務，我們的考試系統是一種 end-to-end 的服務，從應考人報名、選擇考試、繳費、考試地點的安排、考試系統、考後成績計算，以及考試分析，提供一系列完整的服務。至於考試的分析，則包含該次考試試題質量分析、考試是否為公平的，以及是否有改善空間之建議等。至於考試會場的設計，則強調提供舒服不受干擾的環境，提供沒有壓力且人性化的考試空間。

另因應現行網路考試（在線考試）蓬勃發展，我們運用 AI 技術防止作弊，確保不會有人冒名頂替考試，並防止考試期間作弊行為。在線考試有兩種監試方式：一種是真人監考，另一種為 AI 監考，視委託者對於安控的要求，安全性要求高者採真人監考模式。

此外，透過事後統計與分析數字，顯示出我們的考試是成熟、穩定的系統。系統運用180多種技術，採模組方式，具延展性及擴充性。核心系統是由總公司負責，但會配合做在地客製化。（二本木先生）舉例來說，Prometric Japan 公司全方位的考試系統，會專為日本客戶客製化考試預約系統，並配合日本當地付款方式（如便利商店付款）作金流設計，並就日本應考人做當地資料管理，以及在地化的系統設定。

Prometric 2年前買下美國專做 AI 出題的 Finetune 公司，其所開發之大語言模型 LLC 出題技術，是提供給 College Board 做出題的應用，我們運用 AI 在監試及輔助命題。輔助命題在財金及 IT 都有 AI 的應用。

鄭政務次長：飛行員考試是否有實務考試（實地測驗）。

Mr. Hagenbucher：我們有從紙筆測驗（P&P）轉成電腦化測驗（CBT），也有 CBT 轉換成進階 CBT；有提供實地測驗，但考試日程會與 CBT 分開。IT 系統可有許多模擬環境，可以驗測應考人實務能力。

歐陽科長：在試題設計上與委託者間如何分工？專業團隊都是何種背景？

Mr. Chan：若委託者有專業出題團隊，則 Prometric 僅負責技術層面，畢竟每個領域都由其專業；不過出題技巧有通用方式，Prometric 會給建議。我們有專業出題團隊，背景為心理評測師，大部分有博士學位。

Mr. Hagenbucher：也有支持日本客戶的心理評測師，另外我們有很多亞洲的心理評測師。

鄭政務次長：有自己的大型語言模型嗎（LLM）？開放式系統（如 ChatGPT）的好處是資料可以隨時更新，但會有資料外溢之安全性議題，請問你們如何確保 LLM 與時俱進的更新。

Mr. Chan：Prometric 有自己的大型語言模型，並非採用 ChatGPT，且硬體設備到語言模型都是自己設置擁有的。我們與 ChatGPT 不同的是，我們的 LLM 是每一個模式是針對客戶需求建立的，不會與其他客戶混在一起，模型建立時要與客戶有很多互動，包含之前考題與考題的答案，並將其加入語言模型訓練。心理評測師利用這個 AI 輔助工具，讓出題效率更好，而不是取代心理評測師。

鄭政務次長：試題是採 IRT 模型，或隨機抽樣？

Mr. Chan：採線性動態（LOFT）或 CAT。

楊科長：考選部國家考試非常在意考試公平性，並採用一次式同時於固定會場中辦理考試，就10,000-15,000人同時參加電腦化測驗的大型考試，有何建議？

Mr. Chan：我們有一個實際同時辦理15,000人考試的例子，係使用我們公司的產品、運用 BYOD 的方式來辦理，並非固定會場，至於未來貴公司擬採用 BYOD 或會場型何種方式辦理一次式同時之考試，相關細節須再討論研究。

楊科長：申論題的非純文字作答（如繪圖）如何解決？就有時需使用鍵盤輸入、有時需使用繪圖板繪圖的考試，有何建議？

Mr. Pawsey：鍵盤整合繪圖板技術層面是可行的，但尚須實作測試。

Mr. Hagenbucher：若真的要繪圖，建議採紙筆測驗，有時這種方法反而比較好。新考試方式的推動，應採逐步漸進推動。

楊科長：如何解決 BYOD 在安全性、穩定性之議題？

Mr. Pawsey：技術層面沒有問題，但要有好的網路環境，且要有 SOP 與檢核表，也要逐步踏實地做好。

楊科長：請問在 AI 輔助出題上，如何達成訓練成效？耗時多久？

Mr. Chan：基本上每個科目至少需要5,000-10,000個題目，大約20分鐘，才能訓練 AI 輔助出題，不過語言不是問題，中文並沒有特別的困難。

歐陽科長：亞洲是否有公部門採用你們的產品嗎？

Mr. Chan：日本有很多公部門使用，很多高利害相關的考試都有在使用我們的產品。電腦化測驗在試卷評閱及考後分析上可以提升試務處理效率。

鄭政務次長：在不同國家上推動電腦化測驗考試，是否有遇到較為獨特的問題？如文化上差異，有些可能可以用 IT 來處理，但有些可能是根深蒂固的社會價值。請問是客戶直接處理？還是依據你們建議處理，可否分享？

Mr. Chan：我們承作2個政府有關的例子，一個是愛爾蘭的駕照考試、歐洲公務人員考試，我們做了許多前期工作，包含製作愛爾蘭駕照考試之教育訓練教材、定義考試流程等。CFA 特許財務分析師（一年30多萬人）是一個5年的合約，包含考試平台開發、考後評分系統等一系列合作。

日本人事院 訪談紀錄

時間：2024年9月19日上午10時30分

地點：日本台灣交流協會

日方：交流協會總務部副長 紺野紗佑里（KONNO Sayuri）主持

秋庭首席試驗專門官 能久（AKIBA Yoshi-hisa）人事院人材局首席試驗專門官

竹中 絵理（TAKENAKA Eri）人事院人材局試驗專門官

諸星 渚（MOROHOSHI Nagisa）人事院人材局試驗專門官室係員

小島 遥（KOJIMA Haruka）人事院人材局試驗專門官室係員

中村 るり（NAKAMURA Ruri）人事院事務総局国際課国際研究官

柏木 綾子（KASHIWAGI Ayako）人事院事務総局国際課係員

鄭政務次長：秋庭首席試驗專門官，還有日台交流協會的副長以及各位朋友們，非常高興各位接受我們的參訪。我們是來自臺灣的考選部。簡單的介紹一下臺灣的考選部。考選部是一個國家機構，主要管理兩類考試，一是全國的公務人員考試，二是負責專門職及技術人員考試。我們意識到，人才應該是多方面的，所以我們也期待在這方面就是可以多加的考慮。到目前為止，臺灣測驗還沒有實施性向測驗和人格測驗，近期我們在某一些公務人員的考試裡面加入了第二階段的口試。目的是希望可以了解職能跟性格的部分。但是我們相信，如果可以透過標準化的筆試實施性向測驗，會是一個更好的方式，可以拓展我們的選項。所以希望這一次就此部分向各位多多請教。

秋庭首席試驗專門官：日本人事院負責國家公務人員錄用考試，考試招募的目的是選拔能夠引領政府未來有才華的年輕人，所以我們的目標是選拔頂尖人才進行業務的推動。二戰之後國家人事院成立，作為一個公平中立的人事管理組織，對人事院來講，公務人員招募考試是我們最重要的職責之一。今天希望你們能了解我們的努力，並且對臺灣考選部未來考試的發展有所幫助。

一、「知能領域」試題命擬方式與流程相關問題

1.在日本中央及地方自治體的公務人員考試第一試筆試中的「知能（ちのう）」領域測驗，其施測目的為何？

秋庭首席試驗專門官：知能領域的題目目的是衡量一般思考能力，不限於專業領域的思維，也衡量推理能力。

2.「知能領域」中，有關文章理解、判斷推理、數的推理與資料解釋等試題的命題藍圖如何規劃？可否提供範例試題供參考？

秋庭首席試驗專門官：如何規劃每個面向的方向，首先是針對未來的發展方向，主管機關正根據最近的實施結果考慮明年的考試內容的方向，國家人事院網站上也提供了範例問題。

3.前開考試內涵與公務工作的職能之間的關聯性為何？為何是這些項目而非其他能力項目呢？

秋庭首席試驗專門官：有關關聯性的問題，誠如我一開始所回答，這些測驗的向度與一般思考能力有相關性，與專業科目無涉，其內容涉及個人綜合理解和解決問題的能力，國家公務人員正是需要這些技能，因為行政部門需要能夠靈活應對日益複雜的社會需求。有關性向測驗中其他向度，例如記憶力和空間理解力等，從測驗執行的角度來看，記憶力測驗很難實施，而空間的掌握能力，只會用於機場航空駕駛的測驗中，對於一般國家公務人員的業務內容而言，可能不是必要的。

4.前開內涵的挑選及發展演變為何？試題的取材來源為何？

秋庭首席試驗專門官：關於一般知能領域各等級試題的內容，係根據過去的測驗結果和經驗來決定未來出題的方向。關於級別的劃分，有高中畢業生和大學畢業生等級，分為綜合職、專門職和一般職，對應公務人員工作的性質，一般職和綜合職有不同的職能測驗，整體來說，綜合職大學畢業程度使用的對象的考題是最困難的，一般職高中畢業程度試題是最簡單的。

5.「知能領域」的試題，在高中畢業、大學畢業與研究所畢業可報考的考試，是否依考試等級，區分相對應程度的試題？

秋庭首席試驗專門官：我們要求應考人明確考試的級別，出題者根據考試級別來命擬試題，之後透過評審會議和審核流程進行調整，使試題的難易度和考試等級相當。一般是由出題者依照級別出題，接下來會進到會議來審查題目，確認其是否符合不同等級之程度，看看是否有需要調整的地方。

6.承上，命題者是否依考試等級區分命題內容？或是一開始並無區分等級，而是命題完成後，於辦理選題與組卷時，有一套區分等級的判斷標準？

秋庭首席試驗專門官：一般知識題的權重會根據情況來修訂，例如從今年開始，國家考試大學畢業以上的基礎能力測驗題目由40題改成30題。

歐陽科長：請問基礎能力測驗開始這個題數減成30題，是什麼原因？

秋庭首席試驗專門官：首先，大學畢業的應考人越來越少，第二個是準備基礎能力測驗很難學習，給應考人帶來壓力負擔，就可能不會參加這個考試，所以就把試題減成30題。

7.如果有預試如何辦理？受試者的來源為何？如何防止試題外洩？

秋庭首席試驗專門官：一般來說，試題並沒有進行預試，而是根據試題審議會參與者和審議會決策過程中的經驗來做出決定和調整，所以是以參加的專家們的經驗來做分析。但是我們有為高中畢業程度的考試進行一些預試，例如有關事務能力的數學/算數測驗就有預試，實施方式是直接和高中合作，進行預試，考試結束後，試題和答題卷會全部回收，這樣就可以防止問題流失。

鄭政務次長：臺灣公務人員考試的試題沒有進行預試的原因，是臺灣的文化非在意考試的公平性，如果採行預試，會有一小批人知道試題內容，社會會擔心公平性的議題。

8.使用過的題目，是否不再使用？或，使用多少次之後會有淘汰的機制？

秋庭首席試驗專門官：有些問題已經使用過一次，我們會重新整理並且重複使用。

9.應考人去補習或練習坊間的模擬試題，是否對於「知能領域」考試造成影響或困擾？對於「應考人透過練習而增加考試的表現」政府的看法為何？

秋庭首席試驗專門官：在日本補習班是非常興盛的，補習班對於解題及提高分數確實有一定的影響。從政府的角度來說不能限制應考人去補習班練習，從應考人的角度來看，是沒有問題的；而從確保人才的角度來看，學生普遍認為公務人員考試需要過度的準備，這才是一個問題。

鄭政務次長：一般對於試題的理解，就是知識領域的部分透過不斷的練習，對於考試成績會是有效的，但如果是知能領域的部分，透過不斷的練習，是不是仍然可以展現他的潛能？

秋庭首席試驗專門官：知能領域試題練習效果沒有知識領域來的強，但是我認為它確實是有一定的提升也就是說，去補習班對知識的部分可以有所提升，而知能測驗的部分，沒有那麼可以針對，人事院現在的方向也是希望從知識的比重轉到知能的方面。

鄭政務次長：請問補習班會用一些方式來教學生怎麼應對性格測驗嗎？

秋庭首席試驗專門官：公務人員的考試第一階段比較是採知能、知識性的測驗，第二階段就會比較轉到性格的部分，採行面試的考試，性格測驗在檢視應考人的積極性和社交性，性格測驗的結果也會作為第二階段性格測驗面試官的參考資料。由於性格測驗是基於應考人自己表達其性格如何，所以不能說可靠性很高，只能作為參考的材料，因此我在日本並沒有聽說補習班針對性格測驗特別作解題的這種情況。

二、有關「知能領域」命擬委員

1. 相關命題者的來源為何（內部與外部）？

秋庭首席試驗專門官：出題者大概有三個方向，一個是大學教授，第二是一些獨立的行政機關的職員，第三個就是人事院人材局內部的員工。

2. 外部的命題者在命題之前有施予什麼培訓嗎？培訓的內容與方式為何？

秋庭首席試驗專門官：我們對新員工進行初步的培訓，並在實作中得到經驗繼續學習。

3. 試題研發人員，有哪些專業背景？如：數學系？英文系？心理系？測驗專長？...等

秋庭首席試驗專門官：人事院測驗部門有40人，至於是否試題是否有不同的發展方向嗎，實際的狀況沒有針對新型試題進行研究，還是以之前的出題方式，現階段並沒有一個新的方向，還是在現在的範圍之內。

歐陽科長：請問現在在面臨知能領域的這個部分會遇到什麼樣的問題呢

秋庭首席試驗專門官：要如何在考試的時候這個將試題的內容與實際工作之後所需要的知識距離不要太遠，是人事院努力的方向。另外，就出題者來說，知能領域仍是以現有範圍來命題，但也希望可以有一些新的方式，但就結果來說，能展現多少也是現在所面臨的問題。對於在準備入職考試的人消除這些問題是有意義的，例如可以透過職業的需求創建新的問題，在討論問題的時候，可以藉由情境的討論題，討論題的內容實際上是一個解決方案。

鄭政務次長：這陣子對於新興科技有很多的討論，會不會想在知能領域這個部分導入生成式人工智慧，請問人事院有什麼樣的一個想法，您覺得未來會朝向什麼方向去發展

秋庭首席試驗專門官：人事院並沒有一個具體的政策，就我個人的意見而言，只是想知道是不是可以使用生成式人工智慧來建立測驗的問題，或者是命題後的勘誤，甚至是自動評分，這些是個人非常有興趣的。另外，人事院的考試中有一些類型的問題是事先給定的，就是有資格參加第二階段考試的應考人，考試內容包括個人的作業、指導和職前考試，這些會給定一個方向，準確的來講，就是像政府的白皮書之類的，會讓應考人事先知道問題將會來自於這裡，考試題目就是請大家現場寫有關於白皮書的內容，因此有些人會想像可能會出什麼樣的問題，然後用生成式人工智慧寫成的論文，並且記憶背誦這些內容寫成答案，本人對此持謹慎的態度，這也是未來的一挑戰。

歐陽科長：是否有應考人會反映試題越來越難，該怎麼樣應對？

秋庭首席試驗專門官：雖然人事院不會直接回應應考人在考試後的意見，但是我們確實會透過各種管道收集這些訊息，然後作為未來出題的一些調整。

三、有關「知能領域」試題的運用

1. 試題施測的方式為何？係採取集中電腦化測驗、紙筆作答，抑或其他方式？施測效果有無差異？

秋庭首席試驗專門官：日本國家公務人員考試目前只採用紙筆測驗的方式，還沒有引進電腦化測驗。

2.對於施測的結果進行哪些分析？除了項目分析（item analysis）之外，有沒有其他質性或量化分析（例如與應考人的背景變項做交叉分析、應考人的測施滿意度問卷等）

秋庭首席試驗專門官：關於第一試測驗結果分析是基於準確率跟判別能力這兩個基本的指標，同時我們也會針對年齡、大學教育程度進行交叉列表。另外就不會對應考人做特別的調查。

3.一般而言，第一試專業科目表現較優的應考人，是否其基礎能力測驗（新教養測驗）也表現較優？兩者之間的關係是否為正相關或不一定？是否期待兩者有中度相關？

秋庭首席試驗專門官：關於考試結果的相關性，兩者之間是有正相關的關係。

4.承上，「知能領域」與「知識領域」的表現，是否為正相關或不一定？是否期待兩者有中度相關？

秋庭首席試驗專門官：關於第四個問題 無法確認知能領域和知識領域之間的相關性。

5.承上，「知能領域」中，「數的處理」與「文章理解」的表現，是否為正相關或不一定？是否期待兩者有中度相關？

秋庭首席試驗專門官：數的推理和文意理解之間的相關性，兩者確實是評測不同的技能，所以本人認為相關性並不高，只有一定程度的相關。

6.此測驗工具是否能有效排除不適合擔任公務人員的應考者？

秋庭首席試驗專門官：因為我們設定了30分的切點，任何科目的成績在該科目30分以下就不及格，就是每個科目只要有一科低於30分以下的話，應考人就會被排除掉，即使其他科目的分數很高特別好也不行。

鄭政務次長：在臺灣是不能有一科是零分。非常謝謝秋庭首席試驗專門官先生讓我們了解他們的做法從這個觀點可以重新考量我們的標準，非常有幫助。

秋庭首席試驗專門官：所以既然訂定了這樣的標準，本人認為是有一定的排除作用，其後再跟其他類型的考試結合起來，像是性格測驗，還有論文考試，都可以更有效的排除不適任的人。最後，通過這個考試的人仍是要到各部會面試，他們會經過官廳訪問的一個過程，由機關選定他們實際所需用的人。

歐陽科長：請問從2012年開始實施基礎能力測驗，想要知道實施之後跟實際的工作內容 還有這個能力測驗之間的關係

秋庭首席試驗專門官：我認為這是一個非常困難的問題，如果要採取科學的方法，應該要看員工入職之後的工作表現和測試結果之間的關聯性，但是各部會的工作人員很難向我們分享他們的表現的一些信息，所以他們自己對

員工的表現也沒有共享這些資訊。

鄭政務次長：事實上臺灣也遇到類似的困難，非常感謝分享。

秋庭首席試驗專門官：未來有機會收到各部會關於2012年後實施基礎能力測驗的意見，也會請人事院的專家表示意見，這一點確實在評估當中。

鄭政務次長：人事院人才局在計畫命擬試題時，是否會與其他公務機關進行討論，以了解應該要怎麼設計這些試題？因為國際局勢跟科技的變化，他們的想法會不會因此有一些改變？或是人事院針對這些問題會更主動地與其他機關討論？

秋庭首席試驗專門官：目前各公務機關並沒有針對試題的設計相互交換意見。

鄭政務次長：今天非常感謝日本人事院的協助與經驗分享，希望我們下次還有機會再會面。

Recruit-MS 公司訪談紀錄

時間：2024年9月19日下午2時

地點：Recruit Management Solutions 公司

日方：出席實體會議2人

- 行銷部經理 有本倫子
- 研究員 Nita 先生

線上參與2人

- 行銷部 Kowa 先生
- 系統部門 Hirota 先生

研究員 Nita 先生：實施測試有4種不同類型的考試方式，第一個是會場型，在考試中心舉行，第二個是線上型，第三個是 in house，就是在公司內部設置的會場實施考試，第四個就是紙筆測驗。今天討論的是前三個項目。因為前三個項目都是用電腦來進行，應考人來參加考試，都是由題庫提供試題，只是會場型單獨一題庫，線上型和企業內部（in house）型是一題庫。會場型的考試會在主管的監督之下，但是網路版沒有監督機制，而 in house 是在企業內部進行，由他們內部的員工監試，這是一個很大的問題，因此假設這些問題目會外洩的風險，所以線上的考試或者 in house 的考試的題庫會跟其他考試方式的題庫分開，SPI3盡可能避免使用外洩的試題，所以經常會檢查是不是有任何內容被洩漏，並且不再使用任何在網路上已經公開的試題。會場型的考試，有兩種方式監考的方式，一種是真人去監考，一種是用電腦的攝影機去監考，目前這樣子方式只有在會場型實施。因此這兩個題庫的差別不是在於建置方式，而是在於後續的管理方式，差別主要是在這個地方。

研究員 Nita 先生：這三種類型都是使用 CAT 的方式，題庫的試題每一個都有提前收集好的數據，比如說難度、等級等，因此當應考人參加考試時，回答正確的時候，就會出現更難的問題；如果答錯這一題的話，又會下降他的難度。

楊科長：我們在 Prometric Japan 的考試中心有看到 SPI3的專屬試場。

研究員 Nita 先生：SPI3的測驗中心有外包給兩家不同的公司，一個就是 Prometric，是採真人監考的方式，另一個是 JJS，採用監視鏡頭監考的方式。

歐陽科長：請問一下這兩個題庫的試題的難易度有什麼不一樣嗎？

研究員 Nita 先生：試題的難度是一致的，如果在網路上面看到了，就是試題外洩的話就會把它從資料庫刪除，不能再使用，所以是以管理為導向。

歐陽科長：如何確認試題的難易度是一致的？我們想要問一下你們實際上面是如何去創建和維護這些事情的？

研究員 Nita 先生：這就是試題開發的示意圖，管理試題的運作方式，左邊是指新試題的開發，還有出題者的進用和培訓，右邊的部分是針對現有的試題，一是防止外洩，第二是因為題目一直使用，它的有效性就會降低，如何防止試題的有效性降低，需要不停的修正和調整試題。新試題的實際開發是由外部專家來完成試題，現在大約有30位左右的專家在幫公司出題。當公司收到這些專家們所出的題目之後，會開始進行一個會議來檢討，然後做分析採用。

鄭政務次長：請問剛剛講到試題的曝光率，會嘗試著去調整，會調整哪一些項目，是試題內容還是它的難度？

研究員 Nita 先生：試題有洩漏的話就是完全從資料庫刪除，一直刪除的話，題庫的試題會越來越少，會委託外部的專家每一個月都會出題，所以會一直接收新的試題，之後會以試行的方式給應考人測試，每一個應考人會得到一題或兩題新放進去的試題，測驗結果就會去比較應考人在其他考題上面答題的情形，就可以去推測新試題的難易度。

鄭政務次長：試題的曝光有兩種方式，一種是應考人自然的看到，第二種是洩漏到應考人之外，如漏到網路之上。因為通常來講，新開發的試題的洩漏通常是很自然的曝光，例如我們管控大概只有10%的應考人可以看到這個試題，通常在談 CAT 的時候，會有一個曝光管理的這件事，通常不會讓太多的學生看到這個試題，同一個試題大概只有10%的人會看到。這邊所指的洩漏對策是指那一種形式？

研究員 Nita 先生：因為公司會要求應徵者參加 SPI3考試，如果應考人已參加 A 公司委託評測取得 SPI3的評量結果，之後 B 公司也希望應徵者有 SPI3的評量結果，本公司會鼓勵 B 公司考慮是否直接採用 A 公司委託評測的結果，如此一來，就可以自然地減少同一個應考人參加 SPI3的次數。第二個方式就是如果應徵者或是 B 公司還是希望另外參加 SPI3的話，至少在會場型的考試，可以確保的是該應考人前兩次考過的試題不會出現，這個部分目前是做得到的。

楊科長：企業內部（in house）考試會不會有重複試題的問題？

研究員 Nita 先生：線上或是 in houses 在企業內部自己施測的話，就沒有這樣子的功能。企業自己可以決定，如果對結果比較嚴格的話，可以要求應考人去考試中心應試，但是如果企業沒有要更嚴格的考試結果，也可以透過線上考試。

提問議題：研發試題的人是該領域的學者專家（如教授）嗎？還是公司內部的研發人員？

有本倫子經理：目前這三十位的試題開發的專業背景，以心理學領域居多。

提問議題：人類的能力包羅萬象，單一測驗不可能全部納入，總要取捨。在基

礎能力檢查中，如何區分「語言」與「非語言」問題，各自包含哪些能力種類？

研究員 Nita 先生：就是一個公司基本入社考試來講，目前實施基礎能力測驗 GAT，GAT 基本上就是在商業上面的基本能力，所以是一般人的入社的時候的一個測驗標準，內容的難易度基本上是中學的程度，比如說中學的語言能力、數學能力等這些問題而已。就內涵來看基礎能力，測試三個部分，第一個是言語，第二個是論理，第三個是數量，所以要有數學的能力，還要有語言的能力，再來就是分析申論的能力。為什麼會設定中學的程度，因為日本的義務教育是到中學（國中），所以是以日本全部的國民義務教育所都具有的程度，以此為前提。

提問議題：對於施測的結果進行哪些分析？除了項目分析（item analysis）之外，有沒有其他質性或量化分析（例如與應考人的背景變項做交叉分析、應考人的測施滿意度問卷等）？

研究員 Nita 先生：從品質上的角度來看，大約從三個點來檢查，一個是標準值，標準值就是以這一年來參加的應考人表現成績的平均來計算，在日本來說，這個叫做偏差值的概念；另外一個就是信度，就是各個公司對他們施測結果信賴的程度，會以各種角度來確認信度，主要是用阿法值來確認信度，作為信度的基礎。另外一個就是效度，本公司想要知道員工在職場的表現，會定期在各個產業裡面做進行調查，知道這個測驗的結果和考試想要要求的能力是不是符合，也就是說會以實際公司的工作情形，員工加入企業之後，在企業裡是否有良好的表現，透過這些資訊的收集，檢查與測試之間是否存在關係。

歐陽科長：測量的方式是採用質化還是量化的方式

研究員 Nita 先生：每一家公司都會分析和測驗結果上，例如人員的評價、升遷速度、工作的實際績效，評估對他們公司很重要，是不是真正有用那這個用數字來查看是一種形式，主要還是看見那邊提供什麼樣的資料，而當然是他們的評價的方式。

歐陽科長：過去橫濱、大阪、神奈川都有採用 SPI3，現在的情形是怎麼樣呢？他們對於試題的使用也沒有回饋？他們對於試題的組成有沒有特別的需求呢？

研究員 Nita 先生：地方自治體對 SPI3 考試的內容並沒有要求改變，因為在日本一般民間公司就職的季節跟地方自治體考試的季節來說，一般的民間企業會比較早一點，地方自治體會晚一點，地方自治體的考量就是這些優秀的學生既然已經考過 SPI3 了，那麼接下來如果公務人員考試也接受 SPI3 的話，這些優秀的人也許會想要試試看，這就是地方政府的考量）

鄭政務次長：SPI3 是怎麼樣判斷一個人的性格，它的依據是什麼？

研究員 Nita 先生：SPI3有18種性格因素作為人格量表，更準確的說，這是一個特徵理論，如果從18個方面來看，基本上來說，使用最常見的 MMPI 和 MBTI，Recruit-MS 公司研究的起點，是思考作為一個日本公司，要尋找什麼樣的性格特質的員工，所以性格測驗更像是透過實際經驗的積累而發展和創造出來的。

鄭政務次長：請問每一個公司都會分別呈現18個分數還是分成四個分數？每一個公司直接拿到18個分數之後自己再加，我想知道 SPI3測試是否每個公司都有一些調整關於主題中的點數，例如是把18個項目全部分成單獨的點數，還是以四個大項計算出來的點數？

研究員 Nita 先生：一切都是共通性的，不是以公司來劃分，但是第一個答案是每個人都有自己的優點，沒有一個整體的觀點，因為每個人測的是他自己的個性，每個人的個性都有不同的地方，所以他們想要展現的是這個考試的結果，每一個人的個性在各種不同的方面的程度是怎麼樣，所以並不會特別去把每一個細項把它加成一個大分數。

鄭政務次長：請問性向測驗是不是需要補去補習？日本是不是有這樣的情形？

研究員 Nita 先生：本人認為過去是有這樣子的情形，但那是很久以前的事，大概是在2005年左右，（大家會為了考 SPI3寫很多的模擬題，那個時期可能是有的，但是現在）現在公司並沒有出書來說明性格測驗，但是用測驗的結果來決定何時加入一家公司的情形，越來越多而且用 HR 的數據來測謊的情形也越來越多，公司也會用這樣子的結果來分配工作，不論是向他們開發考試的企業來說，或者是對一般的企業來講，都還是強調一個人的個性為主，就是應考人不應該特別去為了考試準備，才能夠顯現出來最直接的最正確的答案。因為 SPI3的結果現在很多公司把它當成一個人的人事資料，所以你進到公司之後，他會依照你的 SPI3 評測的結果分發你的單位，所以如果不是真實的反應你的個性，然後去考 SPI3的結果出來，就容易到一個不是屬於適合你個性的話，這樣子對當事人來說也是很麻煩、很辛苦的的一件事。所以公司沒有去出版這些模擬題，也不會把考古題出版，現在也對外宣傳就是要依照原本的自己的個性來回答這些問題，這樣對應考人的就職才是最有幫助的。

歐陽科長：他們在網頁上面也說到他對 SPI3的評測事前準備再多，成績上面表現也不會有太大的改變，請問他們是怎麼樣設計試題怎麼樣做到的呢？

研究員 Nita 先生：第一個就是之前談到的18個項目，就是不應該做太多，因為性格不僅會被正確的評測出來，而且之後也會被人事部門使用。第二個就是他們的出題中，其實可以分析出來，其實他答題可能就不是依照他自己的個性，比如說他題目答案裡面最好那個是什麼，他就選那個，（應考人回答的方式比如說他是以社會對一個表現好的人的期望這個方向去答題，這種他們是可以分析出來的）系統是可以抓出來的。

鄭政務次長：性格測驗裡面有一個 Social disable 量表，其內容是自己發展，還是參考國外的一些資料訊息？

研究員 Nita 先生：為了這個研究，會有一個實驗組和對照組，公司會要求有一部分的人在測這個題目的時候全部都是依照社會的要求來回答題目；有另外一部分的人，就是依照自己的個性回答題目，然後這個所得到的結果，再分析說是在哪一個項目或是哪一些在18個項目裡面，有哪一些會有一些特殊的表現，就是用自己所收集到的資料來做分析。就結果來說，公司並不會特別說這個人的回答就是完全以符合社會期待的方向去寫，只會提醒委託的公司，未來在面試階段再來確認求職者的情形。

歐陽科長：進一步的詢問，在能力測驗的部分，應考人在短期之內重覆受測的成績會有多少的改變？

研究員 Nita 先生：這是一個非常困難的話題。應考人在短時間內重覆考試的成績是不是會改變，公司的做法是先找一批受試者，讓他們先參加 SPI3 考試，之後把他們分成兩組，一組就是繼續去做模擬試題或者是補題目，另外一組就是什麼事也不做，那這兩組在一週之後，再考試的結果，就是差兩到三分，確實會有差別，有唸書的那一組確實會有差別，那一組可能是因為他們對於這樣的試題比較熟悉，有可能是因為已經使用過 CBT 這個考試的方式，雖然結果確實有差異，大概就是會有兩到三分的差距，但是還是差不多的。

歐陽科長：請問 SPI3的難易度是如何設定？

研究員 Nita 先生：因為在設計考題的時候，基本上能力測驗的題目難度就是設定在國中畢業的程度，所以試題的難易度的基準就在那裡，設計試題時，著重在問題的解決能力，因為難度就是設定在國中畢業的程度，不會再更難，反而是要測驗應考人的推論的能力、申論的能力，所以這個在 SPI3 出題的時候要注意的地方

鄭政務次長：請問現在你們的產品在業界使用的時候，可以分項選擇能力測驗跟性格測驗，還是一定要做能力測驗跟性格測驗？另外公家機關對於考試採取 SPI3的方式，他們比較接受是會場型，還是完全無限制？

研究員 Nita 先生：關於第一個問題，現在的產品其實有能力測驗和性格測驗都包含在 SPI3 測驗裡面，公司提供的服務基本上是包在一起的，只是說企業要怎麼樣去採用他的結果，只重視他的能力的部分，還是重視性格的部分，就是由企業自己去決定；另外一個問題關於線上或者是在公司企業內部施測結果的採用，地方自治體只接受會場型和企業內部的考試的結果

鄭政務次長：我們看到一個這麼有品質有規模的測驗機構，不曉得貴公司早期是有一些專業的人員，意識到這些需求才創建這家公司，還是公司意識到這樣的需求然後集體的投資一個公司。因為我們看到有很多公司願意採

用，地方自治體也願意繼續使用，這樣子的規模的效果，所以我們想要知道最初的時候，這樣子的規模效果是怎麼樣達成的？

研究員 Nita 先生：Recruit-MS 公司當初成立的時候，內部已經有一些是心理學專業的專家，一開始就來重視開發研究，主要就是把研究成果或是實施考試所得的數據，會在一些學會或是研討會發表，這樣就會讓更多的人知道公司有提供這樣子的服務，所以一開始的著力點就不只是自己在研究開發，還會把這樣子的結果發表，推廣公司有這樣的能力提供服務。

鄭政務次長：因為現在的國際局勢的變化，不管是應用新興科技或是因應商業版圖的關係，Recruit-MS 公司對未來也沒有什麼特別的計畫？

研究員 Nita 先生：就新的技術，其實新興科技如人工智慧的應用，就現在的時間點來說，目前還沒有實際的運用。比如說人工智慧的監試功能，其實他們已經有研究過了，但是就技術來講，認為還是有困難的，所以就已經先放棄了這個方向，未來的發展的可能就是出題的方式，生成式人工智慧技術都已經存在，以後也許會讓更有效率的出題。

鄭政務次長：在臺灣這幾年我們也意識到，找到好的公務人員，甚至好的人才的時候，不應該只侷限在專業能力，所以臺灣考試院也希望就是在知能領域的部分和性格方面的多加的考量，因為這個原因就有這個機會來參訪 SPI3 的運作方式，非常感謝 Recruit Management Solutions 公司的協助。

附錄 5 【參訪行程剪影】

一、拜會 CBT-Solutions 公司，與取締役 佐佐木元昭先生（左二）等人合影



二、拜會 e-communications 公司，與取締役 營業本部長佐藤直人先生（左二）等合影



三、拜會 Prometric Japan 公司，與 Mr. Jim Hagenbucher（Vice President of Prometric, General Manager of Prometric Japan，左二）、Mr. Humphrey Chan（Managing Director, Asia Pacific，左一）等人合影



四、拜會人事院，與交流協會總務部副長 紺野紗佑里女士（右一）、人事院人材局首席試驗專門官 秋庭能久先生（右六）等人合影



五、拜會 Recruit Management Solutions 公司，與行銷部經理有本倫子女士（左一）、研究員 Nita 先生（左二）合影



六、拜會駐日代表處與李逸洋代表（左三）、蔡明耀政務副代表（右二）、何仲民組長（右一）合影

