政治科學論叢 / 第 105 期 / 民國 114 年 9 月 / 頁 1-38

爲何公務人員不接受考績乙等？
設計思維下的議題重構與創新*

蘇偉業**、賴怡樺***、張詠涵****

摘　要

我國公務人員考績制度長期遭受甲等浮濫，而公務人員卻抗拒考列乙等之困局折磨。政府權宜地實施甲等比例設限遏止浮濫，卻沒有解決問題。本研究以設計思維途徑重新建構問題框架，將問題聚焦在：爲何公務人員不接受乙等？透過針對公務人員實施的情境式問卷進行回溯性推理，以及針對臺灣、香港兩地文官與老師的作文評核實驗，本研究發現臺灣可能普遍存在對乙等之偏見，認爲乙等是表現不良的符號。這是理性假設及一般人力資源管理理論無法解釋的現象。基此，政府有必要重新設計考績列等之選項，而非僅是在現有「甲乙丙丁」的框架下調整設計。

關鍵詞：公務人員考績、設計思維、問卷調查實驗、非理性、推力

---

* DOI:10.6166/TJPS.202509_(105).0001
本文初稿曾於 2024 年 TASPAA 台灣公共行政與公共事務系所聯合會年會發表。本文部分內容改寫自國立政治大學行政管理碩士學程碩士論文「爲何公務人員不接受考績乙等？我國公務人員考績設計之檢討」（張詠涵，2022）。感謝研討會評論人及期刊匿名審查人所提供之寶貴修正意見，惟文責由作者自負。
** 國立政治大學公共行政學系教授，E-mail: bennisso@nccu.edu.tw。
*** 東吳大學政治學系助理教授，E-mail: evalai920@gmail.com。
**** 國立政治大學社會科學院行政管理碩士，E-mail: annie8062@gmail.com。
收稿日期：113 年 6 月 20 日；通過日期：114 年 9 月 22 日

臺灣大學期刊資料庫

壹、緒論

「2023 臺灣燈會在臺北」全國花燈競賽評獎有五個等級：燈王、特優、優等、甲等、佳作。[1] 其中排名第四是甲等，最後是佳作。唉！爲何「甲等」之後不是「乙等」呢？所以「乙等」表示不佳嗎？

以上觀察啓示並不足以證明乙等就是不佳的觀感或認知，因爲這是一項競賽評獎，是擇優。然而，這觀察足以讓我們開啓另一扇探索之門，探索在其他方面，作爲一種評價符號的「乙等」是否不單不能與「優秀」相提並論，且淪爲「不良」、「不佳」之代表？可與這設想匹配驗證的評價機制正是我國的公務人員考績制度，因爲一般公務人員對考績考列乙等皆十分感冒、反彈，打考績的主管頗感爲難，以致我國考績評等分配出現甲等膨漲現象，但外界輿論廣泛認爲考績甲等占比過高，使考試院及行政院祭出設定考績甲等之 75% 之比例上限。不過，這比例仍然極高。

過去對公務人員考績問題的分析，大多以主管的「理性受限」作爲解釋的切入點，致使主管不得不採取「潛規則」，以工作表現無關的標準來分配甲、乙等的基準。然而，「理性受限」不可能是我國機關主管所面對的獨有問題，但我國公務人員考績最高評等占壓倒性多數比例的情況，在其他國家實屬罕見。基此，先不論對解決問題是否有幫助，該解釋能否刺中問題要害也成疑問。

近年流行的設計思維（design thinking）理念與方法已進入公領域的政策／制度設計。其強調針對如上述考績的棘手問題處理，應先質疑既有的問題框架，透過回溯性推理（abductive reasoning）加以重建，跳出既定的思考框框（thinking out of the box），溯源特定問題的根本所在，並由此發展出具創意的解決方案。

本研究正是要針對我國公務人員考績制度的特殊現象進行再檢視，透過對問題的批判性反思，重新定位問題焦點，從「甲等」轉移到「乙等」。設問：爲何公務人員不接受考績乙等？爲此，本研究進行了兩階段的實驗

[1] 相關資料可參看臺北市政府教育局網頁（臺北市政府教育局終身教育科，2023）。

臺灣人學學術期刊資料庫

性調查。首先，透過針對公務人員實施情境式問卷以進行回溯性推理，找出導致這問題的根源；接續依第一階段的實驗結果，調整考績列等的設計，進行第二階段實驗，針對由臺灣與香港文官及老師組成的三組測試樣本，實施中文作文的評核模擬試驗。臺灣實驗組依新設計的評等進行評核，而臺灣及香港華人的兩組對照組則依現行評等（甲乙丙丁）進行評核。第二階段的實驗進一步確認了第一階段實驗的發現，同時也爲推力式（nudging）的列等新設計進行了初步驗證。

基於研究發現，本研究認爲，我國公務人員普遍不接受考績考列乙等可能是基於固有偏見，人力資源管理理論或國外經驗皆無法合理解釋此現象。因此，政府有必要重新設計考績列等的選項，而非僅是在現有「甲乙丙丁」的框架下調整設計。

貳、設計思維的思考邏輯與公共治理

設計思維是本世紀初崛起並迅速發展的一種強調以人爲本（human-centered）的設計途徑，其從最初應用於商業領域，逐漸擴散到公共領域，於近十多年延伸成爲社會創新、公共服務設計及公共政策制定背後的創新泉源之一（Bason, 2014, 2018; Brown, 2019; Brown & Wyatt, 2010）。各國成立的政策實驗室大多採用設計思維作爲創新的重要途徑（McGann et al., 2018; Olejniczak et al., 2020）。

不過，公共政策學者是近年才特別關注到設計思維在政策設計之應用，並在 2020 年於期刊 *Policy & Politics*（48 卷 1 期）出版專刊討論，嘗試與現存的政策設計理論做對話與重新區隔分類（Hermus et al., 2020; Howlett, 2020; Lewis et al., 2020; van Buuren et al., 2020）。其中，Lewis 等人（Lewis et al., 2020）與 van Buuren 等人（van Buuren et al., 2020）就認爲，作爲一種公共政策途徑，設計思維導向處理公共問題的創造性理念（creative ideas），是一種要求跳出既定的思考框框的創造性藝術（creative art）。這有別於強調選擇已有的政策工具來達成既定目標，重視技術專業能力（technical expertise）的傳統理性（rational）政策設計途徑；也有別於強調利害相關者共同設計（co-design）或協力設計（collaborative design），重視對話與建立民主共識

臺灣人學學術期刊資料庫

的參與式（participatory）設計途徑。值得注意的是，過去討論的政策創新，實多聚焦在擴散（diffusion）過程（Berry & Berry, 1999），或促進政府創新的條件（Bekkers et al., 2011），而沒有討論如何設計創新政策，特別是「無中生有」（*ex nihilo*）的創造，這是政策學者一早就提出的批評（Alexander, 1982）。而設計思維眞正提供了公共創新的途徑與工具。

誠然，公部門有其保守性，不利於創新（如官僚規避風險的傾向、官僚繁瑣程序規範的約束），但透過實踐，設計思維實務者已發展出一些特殊策略來舒緩或排除運用障礙（如透過委託政府外部設計者、強化文官對試錯的心理建設），且一些先行者（如丹麥與荷蘭）已累積了不少成功經驗（Brinkman et al., 2023）。其他西方國家也陸續投入試驗，如澳洲及紐西蘭就有效利用設計思維途徑來改善醫療及社會服務之政策（Mintrom & Luetjens, 2016）。曾擔任著名丹麥政府創新試驗室 Mindlab 執行長的 Christian Bason (2018)，依其實務經驗提出政府運用設計思維的四項信條：(1) 要有實驗精神；(2) 要挑戰現狀；(3) 要重視公民，以人爲本；(4) 設計過程要盡量具體化，使人可感知。

設計思維之落實充滿著各種工具與技巧，但設計思維之所以能導向創新之核心是在於其背後不同的思考邏輯，工具與技巧僅是協助創新設計的輔助。設計思維在處理問題上傾向採用回溯性推理，即有別於傳統理性途徑常採用的演繹與歸納邏輯（Lewis et al., 2020; van Buuren et al., 2020），其並非建基於既存定律或經驗法則，而是透過可獲取的資訊推論出問題「暫時性」的最佳判斷。此邏輯無意發展通則化理論或一般性的解決方案，因爲一個在某處可行的方案，不見得可適用於他處。而任何回溯性推理所得出的答案都是暫時性的，資訊更新即可加以推翻。於此邏輯之下，設計者思維不會侷限在既有定律或可觀察到的多數經驗，刺激對問題解讀的突破，繼而促進對問題處理的不同設計（Dorst, 2011）。

與回溯性推理一致，設計思維強調對問題的反思式框架重建（reflective reframing），亦即對既有問題的認知抱持懷疑態度，不斷針對問題或現象重新思考，挑戰傳統或一般看法，重建對問題或現象的認知框架（Hassi & Laakso, 2011）。Bason (2018) 也強調政府要創新，必須要用不同方式來思考問題，重塑問題框架。這方法對一些長期無法有效處理的公共問題，提供

臺灣大學學術期刊資料庫

了另類處理途徑。設計學者也認爲設計思維的運用應聚焦在一些可能長期被錯誤定義的棘手問題（wicked problems）（Buchanan, 1992）。這些反思往往會挑戰演繹與歸納法的邏輯推理，採用以問題當事人爲中心的視角，進行特殊解讀性的分析，追溯特定問題的特定根源（Clarke & Craft, 2017）。如此的改變也常關連到行爲經濟學（behavioral economics）的非理性或心理面向（如認知偏誤（cognitive bias）問題）。當中，推力手段與設計思維常會重疊在一起（Hunnes, 2016）。以我國而言，一些問題如教育改革、少子化，皆有可能需要在構思解決方法前，重新檢視問題的本質，重建問題框架（如揚棄簡單的理性假設）。本文所處理的公務人員考績問題，正是我國幾十年來一直爭議膠著，無法有效解決的問題。

設計思維除了是一套處理問題的思考途徑，在實務上也發展出一套操作程序與處理問題的工具，可因應情況需求變通運用。設計思維在程序上一般包含五個步驟：同理建立（empathy building）、問題定義（problem definition）、發想（ideation）、原型製作（prototyping）及測試（testing）。這五個步驟可二分爲兩項目標，即針對問題的探究（前兩個步驟）及針對問題的解決（後三個步驟）。這五個步驟並非簡單的線性（linear）程序，問題探究與解決方案可以共同演化（co-evolution），即對問題的定義與解決方案的探索可以互爲因果，在操作過程當中共同演變（Dorst, 2019）。

設計思維在公部門之應用一般是聚焦在民眾服務及對民眾有影響的措施上（參看 Bason, 2014），而本研究則聚焦在官僚系統本身，會以設計思維邏輯重構對公務人員考績議題的框架，嘗試跳出既定的思考框框，並利用設計思維常用的工具進行問題探究。而在探究問題根源的同時，本研究也發掘出解決問題的可能之道（即共同演化邏輯），並進行初步驗證。

參、公務人員考績的議題框架：從甲等到乙等的聚焦轉移

一、我國公務人員考績列等制度之演變

早自南京國民政府成立初期，我國已建立考績制度，對受考人員以列等（評分）方式進行評價，但並未明列比例限制。由於出現考績寬濫問題，

臺灣大學學術期刊資料庫

1943 年公布的「非常時期公務員考績條例」首次限制比例，考列八十分以上之人數（即最高之一等）不得超過機關考績總人數三分之一。抗戰勝利後 1945 公布的「公務員考績條例」維持該規定：考績評等分五等，考列一等人數不得超過機關考績總人數三分之一；1949 年公布「公務人員考績法」，仍規定各機關年考列一等人員，不得超過各該官等受考人數三分之一（徐有守、郭世良，2019）。

1954 年修正考績法，在一等之上加列特等，五等以下加列劣等（免職）。並規定各機關年考，各職等人員均得考列特等一員，各該職等參加考績人數在五十人以上者，得加列一員，在一百人以上者，得加列二員，但以工作學識操行均確屬特別優異，有事蹟可舉者爲限，如無適當人員，不得濫列。各機關年考一等以上員數，不得超過各該職等受考人數三分之一，如各該職等有餘數，滿二員時，得加列一員。惟 1962 年修法又廢除了上述的特等及劣等，除機關長官、副長官另行斟酌處理外，其餘人員維持考列一等人數不得超過同職等各該職務受考人數三分之一（徐有守、郭世良，2019；許道然、林文燦，2015）。

1967 年實施公務職位分類制度，另行公布「分類職位公務人員考績法」，適用於職位分類人員，考績等次分甲、乙、丙、丁、戊五等；後於 1969 年修法刪除戊等，並廢除考列甲等人數不得超過機關考績總人數三分之一規定。這是 1943 年對最高考績等次設限後，第一次棄設上限（徐有守、郭世良，2019）。惟不過幾年，在 1971 年復以行政命令方式再度設限，「考績考列甲等者，以參加考績（成）人數三分之一爲原則，最多不得超過二分之一」（許道然、林文燦，2015）。但現實是考列甲等者在 1972 年首次過半，達 51.2%（施能傑，2024，頁 159）。

1987 年建立新人事制度，將簡薦委與職位分類制兩制合一，考績評等沿用職位分類制下的甲、乙、丙、丁，並再次廢除考列甲等比例的原則限制，只規定考列甲等條件。1987 年全國考列甲等人數比例馬上出現斷裂式攀升，從 1986 年的 51.3% 躍升到 79%，並逐步攀升至 2000 年的 86.7%（施能傑，2024，頁 159-160）。最終，銓敘部與行政院人事行政局於 2001 年與中央及地方主管機關人事機構協商獲致結論，「對考列甲等人數比例限制，長期制度上宜修正公務人員考績法予以明定，並已列入修法計畫中，其限

臺灣人學學術期刊資料庫

制以二分之一爲原則，最高不超過三分之二；在尚未完成修法前，先循首長用人考核權方式，採漸進原則處理……甲等人數比例，先以 50% 爲原則，最高不超過 75%」（許道然、林文燦，2015）。嗣後，每年銓敘部及人事行政局（今爲人事行政總處）首長皆以聯名箋函方式請中央暨地方各主管機關配合，機關考列甲等人數比例「以 50% 爲原則，最高不超過 75%」。自 2001 年後，全國公務人員考績考列甲等的比例大約徘徊在 75% 上下。考試院也多次透過提出考績法修正案，規範考績列等比例，但迄今仍未通過相關修正。

二、甲等比例膨漲討論之困境與盲點

從上文回顧可見，對考列一等或甲等比例設限，是對考績寬濫明快而有效的回應。然而，透過一刀切的比例設限，顯然並未考慮不同機關的狀況，而實施相對性績效評等法（relative performance rating）必然產生不公平的觀感（Roch et al., 2007; Schleicher et al., 2009）。爲了順從甲等比例限制，我國機關主管常常會以所謂「潛規則」，即非依工作表現優劣，改以其他與工作表現無關的因素，分配員工年終考績的甲、乙等次。這是一直爲人詬病的不公平陋習（陳敦源，2010；張瓊玲，2009）。

惟回顧歷史，上文已指出，我國曾兩次廢除考列甲等人數上限。第一次是當時新成立的人事行政局所主張，認爲機關若全體人員都守法守紀，品行優良，勤懇奮發，工作成績效優良，不應對這些機關考列甲等的人數設限，於是廢除考績法內的相關比例限制（徐有守，2006）。第二次廢除甲等人數上限，是在新人事制度建立後，改以明定考列甲等的具體條件，即採用絕對績效評等法（absolute performance rating）。自廢除上限後，考列甲等人數馬上暴增。無可否認的是，依我國過去經驗，一旦廢除比例，甲等的膨漲程度使人瞠目結舌，是故人事主管機關不得不踩煞車。即便是現行甲等上限 75%，也會有人質疑比例過高。依最新銓敘部統計，2021 與 2022 年考列甲等比例更從 2020 年的 75.6% 突然衝破到 79.5% 及 80.5%。[2] 一般民眾難以理解爲何絕大多數四分之三的公務人員都可以被評爲最高等

[2] 參看銓敘部（未註明）統計資料。

臺灣人事學學術期刊資料庫

次的甲等，特別是對政府施政感到不滿的民眾而言。

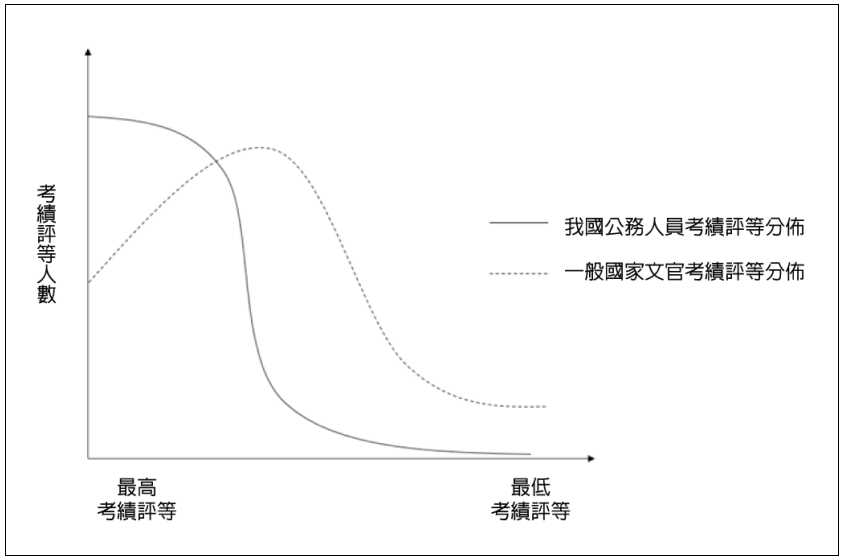
不論是否對最高等次設限，一般會理解組織工作人員的考績等次應呈現鐘形曲線分布（不一定是常態分布）較爲合理，即優秀、表現一般及表現差劣三種人員中，應以表現一般者占多數（徐有守，1996；Collings, 2017）。然而，考核者的考績評等寬濫是普遍的行爲表現（Lin & Kellough, 2019），因此給了政府實施考績等次比例分配或限制一個合理辯護。事實上不只我國，即便是歐盟也有一些國家實施針對文官考績等次的比例限制，包括德國、馬爾他、葡萄牙、義大利及拉脫維亞（Staroňová, 2017）。與我國相同文化圈的日本與韓國也有相關的設限。日本文官考績分五等級，最高 A 等以 5% 爲上限，B 等以 20% 爲上限（Otani, 2023）。韓國文官考績分四等級，最高 S 級（代表優秀）以表現前 20% 爲上限，表現前 20%～50% 爲 A 級（代表傑出）（即不一定有 S 級），表現後 50%～90% 爲 B 級（代表正常），表現墊底的 10% 爲 C 級（代表令人不滿意）（Kong et al., 2013）。由此可見，一般國家的文官考績皆呈現鐘形曲線分布。相比之下，我國的 L 型分佈則頗爲異常，非但最高等次的甲等占比最高（一般高達 75%），然後丙等僅占 0.1%～0.2%、丁等幾乎是零，看似只有甲、乙兩個等次（示意圖請參看圖 1）。

從以上討論可見，我國公務人員考績制度處於一種兩難困局。考績等次設限與不設限都會帶來不盡人意的結果。然而，國內在學術討論上並未嘗試破解該困局，且最終傾向二擇其一選邊站的作法。一些學者一早就注意到我國公務人員考績寬濫的問題，並以一般考績理論會提及的考績謬誤（rating error）作爲切入點（林水波，1990；施能傑，1992a，1992b）；陳敦源、林靜美（2005）進一步以人的有限理性（bounded rationality）來解釋該謬誤，認爲組織成員之間處於「不完全契約」（incomplete contract）狀態，因此主管與屬員對考績分配存在一定的默契，接受以「潛規則」的折衷方式來打考績（陳敦源，2010；陳敦源等人，2011）。

誠然，有限理性的解讀無法爲上述困局解套，且有限理性實足以有效處理甲、乙二分法的考績評等分配，不應成爲使用「潛規則」的藉口（曾玟儀、蘇偉業，2017）。此外，考績的有限理性理應非我國獨有，爲何國外未曾出現如此甲等（或最高等次）膨漲的景象？與我國同一文化圈的日、

臺灣大學國家發展研究所

圖 1
我國與一般國家文官考績評等分布的比較示意



資料來源：本研究自行繪製

韓兩國也實施考績評等設限，也未聞兩國出現如我國的相關爭議？有限理性的普遍性並不能解釋我國的特殊性。

即便大家承認人的「理性受限」，但「理性受限」無法解釋爲何公務人員不接受乙等，只接受甲等的原因。近年國外研究引進相關心理學的認知偏誤分析，提出一些可能在公部門常出現的考績偏誤狀況，其實是上文考績謬誤的延伸研究，如定錨效應（anchoring effect）與月暈效應（halo effect）（Belle et al., 2017），但這些通則性發現無助於理解本個案。在沒有破解這疑惑之下，問題解決之道就是立基於「甲等是唯一的理性選擇」之假設下設想。一些學者認爲應加強考績的競爭性（余致力等人，2003）；另一些學者則加以反駁，認爲考績的公平正義才是重點（張瓊玲，2009；陳敦源，2010）。前者主張仍應爲甲等設限，但要引入團體績效，並在甲等之上加設等次，進一步拔尖；後者因應公平性及行爲扭曲（即以潛規則考績）的考量，對設限的作法有所保留。

以上圍繞著「甲等」爲議題框架的討論有一個重大盲點，即當各國文官考績的最高兩等次多屬優良表現時，爲何我國公務人員不接受第二等次

臺灣大學學術期刊資料庫

的乙等？上述的競爭論或公平正義論好像都默認乙等就是不良的假設。我們應否轉向解釋爲何大家不接受乙等，以破解這困局呢？

三、框架重建：轉向聚焦乙等

前考試院院長關中在一次對行政院人事主管會報致詞中，作了以下一段頗爲玩味的評論：

> 考績法的規定很清楚，考列乙等給予半個月的獎金，並晉本俸（年功俸）一級，已達年功俸最高級者給予一個半月獎金。就結果而言，他們獲得晉級也拿到了獎金，應該具有一定的鼓舞作用。也就是說，乙等的人即使沒有特殊表現，至少是稱職。但實際上，因爲有四分之三的人是考列甲等，而丙、丁等少到幾乎不存在，被考列乙等的人被認爲是績效不佳的人，他們也自認爲是被懲罰，不但忿忿不平、抱怨連連，申訴的情形也到處可見（林立曼，2009，頁 10）。

這難得聚焦在乙等的官方評論揭露了三項啓示。第一，在制度設計上，考列乙等並非表現不佳，是「稱職」的，可晉俸級，有績效獎金；第二，雖說考績分爲四等，但實質上只有甲、乙兩等；第三，因此員工認爲考列乙等就是代表績效不佳。

這三項啓示實可轉化爲回答「爲何公務人員不接受考績乙等？」的切入點。評論反映制度設計與現實存在矛盾，就是考列乙等並非表現不佳，但公務人員並不認爲如此。這不見得是理性假設可以解釋的。這涉及了多個層次的問題。第一，理性上的利益層次仍然存在。考列甲、乙所獲得的利益有實質上的落差，包括「獎金、陞遷、訓練進修機會」[3] 與「職務調動機會」[4] 等。第二，進入心理層次，就是實質上的甲、乙兩等制，使考列乙

---

[3] 相關參加升官等訓練的規定皆有明列考績條件，即最近三年年終考績 2 甲、1 乙以上（參看公務人員任用法第 17 條第 2 項）。機關選送公務人員進修的條件也有要求服務成績優良，即最近二年年終考績 1 甲、1 乙以上（參看公務人員訓練進修法施行細則第 13 條）。

[4] 法規上並未以考績等次來限制公務人員職務調動，但實務上許多機關在辦理外補徵才時，常將應徵人員資格條件設定爲近三年或五年考績均考列甲等，或最近五年考績中，

臺灣人事行政期刊資料庫

等者自認表現墊底而致自我形象受損。更甚者，乙等早就被「污名化」，成爲公務人員心目中表現不佳的符號。

上述思辯已初步建構了新的議題框架，下一步我們則須針對不接受乙等的可能因素進行驗證。驗證會以實驗性調查的方式進行，嘗試追溯公務人員不接受考績乙等的根本所在。

肆、研究設計

爲探討不接受乙等的根源，本研究採用實驗方式進行研究。就工作考績問題，一直有不少研究以實驗方式探討管理者的考核行爲與員工對考核之反應，還有各種干預因素如何影響考績結果，例如強制分配（forced distribution）評等機制（Berger et al., 2013），還有以實驗方式探討考績上的認知偏誤問題（Belle et al., 2017）。而跳脫考績問題，從積極面看，實驗干預也可以利用人的認知偏誤來導正負面行爲，例如利用人對左位數偏誤（left-digit bias）[5] 來修正行車時速限制，有效遏止超速行爲（Rubaltelli et al., 2021）。

而本研究因應我國在考績上的特殊課題，即不接受乙等之原因，專門設計了兩階段的問卷調查實驗（survey experiment）。第一階段實驗以「受評者」角度析究該問題的根本原因，研究發現大多數樣本不接受乙等，並非是基於可改變的條件性反應；故第二階段實驗改從「考核者」角度確認是否存在該特殊性，透過臺灣與香港樣本比較受測者的反應是否存在地域上的顯著差異，並透過改變評等選項（一種導正的干預），測試是否對臺灣樣本產生顯著影響。

須三年以上考列甲等。

5 左位數偏誤是指人對最左位數字有過度敏感的問題，在商品價格設定上常利用例如 $5.00 改爲 $4.99 來勾引消費者購物。而內文所指利用這偏誤來導正行車超速問題，就是將時速上限從例如每小時 50 公里微調爲 49 公里，以驅使駕駛減速。

臺灣大學圖書館期刊資料庫

一、第一階段實驗

就第一階段實驗，本研究以設計思維常用的五個爲什麼技術（The five whys technique）作爲藍本，自行設計一份情境式問卷進行測試。該技術由日本豐田佐吉發明，用以探索問題根源，定義問題的工具。其方法是針對問題現象，透過提問「爲什麼」，並就回答持續追問「爲什麼」，以找出問題的因果關係，並發現最有可能導致問題的根本原因（root cause）（Serrat, 2017）。所謂五個爲什麼，未必是問五次，不過一般情況下問五次通常便足以找到問題的根源。而多次之回答會代表從淺層到深層之原因追究。

考量實施面對面訪談十分耗時且成本太高，本研究採問卷量表方式設計實驗，就各種可能導致人員不接受乙等的原因分成不同層次進行測量（猶如五個爲什麼技術的追問層次）。另爲儘可能排除受測者個人經驗與現況的影響，問卷內容設定一些特定情境，要求受測者依該情境下填答問卷（非他們的現實情境），以抽離特定機關或工作場域，並使所有受測者處於同一狀態。在填問卷前，受測者會被告知：第一，問卷標的爲年終考績；第二，考績各等次未設人數比例限制；第三，受測者自認今年工作表現不在單位內的前 25%，但並未犯任何錯誤，也未受到處分，且相關工作都有如期如質圓滿完成。

就以上情況，問卷中列出一系列條件或情況，以陳述句方式呈現其在某條件下（不）願意接受考列乙等，受測者勾選對各陳述題項的同意度，以 10 點量表方式測量（1 爲完全不同意，10 爲完全同意）。問卷設計有 11 個題項，除題項 1 爲在上述情境下就接受考列乙等外，其餘 10 題可分爲三類：在第一、二類中，受測者願意在條件改變下接受乙等，第一類條件是利益性，涵蓋上述提及的獎金、陞遷、訓練進修機會與職務調動機會；第二類條件是心理性，改變考列乙等等於變相考績墊底的設計，包括乙等變成多數、設定丙等比例、刪除現行公務人員考列甲等的一般條件。[6] 第三

[6] 公務人員考績法施行細則法第四條列有考列甲等的一般條件，受考者只須符合兩項條件即得被考列甲等。在這些一般條件之下，只要人員沒有犯錯，沒有遲到、早退或曠職紀錄，大多可以符合考列甲等的條件。

臺灣人事行政學術期刊資料庫

類是受測者在任何條件改變下都不願意接受乙等，這部分分成兩個原因：(1) 同事心目中對乙等的否定；(2) 自己對乙等的否定。

至於該 11 個題項的順序，以無條件接受乙等排在最前，接續是利益條件改變，第三是心理性條件的改變，最後是心理性絕對不接受乙等。以上排列是由顯然直接的物質因素到隱性深層的心理因素，形成四個層次的題項。對第一、二類條件，若受測者的同意度高，代表其對乙等的否定並非絕對性，可以在條件改變下接受；第三類則相反，代表無論任何條件的改變都不接受乙等。

二、第二階段實驗

第一階段實驗發現過半樣本不接受考績乙等是無條件性的反應，我們有必要確認這無條件性是否爲本地脈絡所導致的。爲此，本研究進一步與國外情況做比較，驗證這是否爲我國的特殊偏見。然而，由於各地文官考績制度不同，我們不可能在國外複製相同的實驗，起碼不見得會用「甲乙丙丁」來評等，惟僅針對「甲乙丙丁」的等次格式進行比較驗證是有可能的。這驗證不必是工作考績的情境，因爲如果某一地方在評核作爲上存在對乙等的偏見，其也會出現在其他場域。這偏見不僅是存在於被評者的角色，就算轉換爲評核者，也一樣會出現。第二階段實驗改透過評核作文的方式，以評核者的角度進行實驗。因作文評核與工作考績在形式上具高度相似性，都是對表現產出作出評價，兩者皆可採用等次進行優劣排序。此外，作文評核工作不涉及任何利害關係，有助排除實質利益干擾，純粹觀察對乙等等第的心理傾向。

第二階段實驗以另一個華人社會——香港，作爲與我國比較的對象。由於本測試不能挑選並非使用中文之地區，且須使用繁體字爲宜，基此我們才可能測試同一項評核工作。香港是僅有少數可進行這測試之地方。這項實驗就是要測試兩地華人是否在使用「甲乙丙丁」等次評等偏好上有所差異。同時，實驗中也嘗試測試等次格式的改變，是否會對我國受測者產生評核行爲上的變異。

本實驗所實施的作文評核，是從我國教育部國中教育會考近 5 年國文寫作測驗的作文樣卷中，選取六至三級分的樣卷各 3 篇（六級分爲最好），

臺灣大學學術期刊資料庫

共 12 篇，邀請臺、港兩地的文官與老師進行評核。實驗透過網絡問卷方式進行，分爲三個評核組：其中臺灣對照組與香港實驗組的受測者皆使用「甲乙丙丁」評等，其間以「地域」爲實驗處遇（treatment）；臺灣實驗組以「傑出、優良、尚可、待改進」評等，其與臺灣對照組之間的實驗處遇則爲「評等方式」。「傑出」與「優良」明顯都是正面評價描述，倘若以這兩個等次代替甲、乙等是否會產生不同的反應？而受測者中加入老師，是考量該考核工作涉及寫作測驗，同時也觀察我國公務人員考績的「乙等偏見」是否爲地域現象，起碼非侷限在文官體系內。具體而言，若眞的存在「乙等偏見」，臺灣對照組的整體評等應高於其他兩組。需再強調的是，這實驗是完全沒有涉入任何利害關係因素之考慮，就是對作文的評核。所以如果仍出現乙等偏見，這就不可能是其他利益考慮的干擾。

三個評核組各約 30 人，各組文官與老師占比約 1 比 2。臺灣文官樣本來自政府機關內曾擔任主管職的公務人員（即有打考績經驗者），老師樣本來自北部兩所高中教師；香港文官樣本則來自當地政府機關（制度上未區分主管／非主管職），老師樣本來自當地中學教師。兩類樣本都是經研究者個人聯絡、自願參與試驗，且考量職業、性別與年齡等屬性，以準配對設計方式分派至實驗組或對照組，並透過事後的統計確認，在控制干擾因素與內部效度方面進行補強。

伍、研究發現(一)：乙等就是不好

第一階段實驗問卷在 2022 年 5 月於臺北市政府施測。調查問卷在正式發放前先有預試，因應回饋意見修正個別題項的文字。正式問卷的發放則是經由市府人事人員協助，以網絡問卷方式發送給機關內公務人員，以不具名方式調查。本調查是實驗性質，毋須著重樣本的母體代表性，惟仍儘可能兼顧受測者背景的多樣性，在抽樣上選取具不同職務變項的人員填答問卷，包括主管／非主管、行政職系／技術職系／警消人員、不同性質機關（一般行政、工程、警消）等。最終，調查回收 420 份問卷，樣本的人口背景分布如表 1 及表 2 所示。

臺灣大學學術期刊資料庫

表 1

第一階段實驗中問卷調查回收樣本之各類職務背景的人數與占比統計

變　項	問卷數	變　項	問卷數	變　項	問卷數	變　項	問卷數
主　管	162 (38.6%)	簡　任 (警監)	33 (7.9%)	行政職系	94 (22.4%)	一般行政機　關	82 (19.5%)
						工程機關	12 (2.9%)
		薦　任 (警正)	126 (30.0%)	技術職系	30 (7.1%)	一般行政機　關	19 (4.5%)
		委　任 (警佐)	3 (0.7%)			工程機關	11 (2.6%)
				警消人員	38 (9.0%)	警消機關	38 (9.0%)
非主管	258 (61.4%)	簡　任 (警監)	11 (2.6%)	行政職系	113 (26.9%)	一般行政機　關	97 (23.1%)
						工程機關	16 (3.8%)
		薦　任 (警正)	183 (43.6%)	技術職系	45 (10.7%)	一般行政機　關	25 (6.0%)
		委　任 (警佐)	64 (15.2%)			工程機關	20 (4.8%)
				警消人員	100 (23.8%)	警消機關	100 (23.8%)
合　計	420						

資料來源：本研究自行整理

表 2

第一階段實驗中問卷調查回收樣本之其他個人背景統計

類　別	項　目	人　數	百分比 (%)
性　別	男	223	53.1
	女	197	46.9
年　齡	30 歲以下	32	7.6
	31-40 歲	118	28.1
	41-50 歲	149	35.5
	51-60 歲	102	24.3
	61 歲以上	19	4.5
最高學歷	高中 / 高職以下	6	1.4
	專科	64	15.2
	大學	196	46.7
	研究所以上	154	36.7
婚姻狀況	未婚	114	27.1
	已婚	306	72.9
任公務人員年資	1-10 年	124	29.5
	11-20 年	122	29.0
	21-30 年	120	28.6
	31-40 年	49	11.7
	41-45 年	5	1.2

資料來源：本研究自行整理

臺灣大學學術期刊資料庫

一、尖峰分析

樣本對這 11 個題項的反應，在 10 點量表的分布上皆呈現類似 W 型（參看圖 2，題項內容參看表 3），即每一題的尖峰都出現在最左（1 分）及最右（10 分），而中間（5 分）也會出現一個尖峰，但皆不及最左或最右的尖峰高。中間的尖峰可以理解爲有一定比例的受測者對選題的陳述沒有明顯的好惡或同意、不同意。值得注意的是，從題項 1 到 9，最高的尖峰都出現在最左側，而最右側的尖峰則遠低於最左側的尖峰。在題項 1、5、7、8、9，最右側的尖峰甚至低於中間的尖峰。而依著題項順序，最左側的尖峰有逐漸往下降的趨勢。這表示：

1. 就算自認工作表現不在單位內前 25%，在沒有犯任何錯誤，也沒有受到處分，而且相關工作都有如期如質圓滿完成的條件下，大部分樣本表示極不接受考列乙等（填 1 分者達 199 人 [47.4%]），完全接受的只有 14 人（3.3%）。
2. 就算改變利益條件（題項 2～6）（即甲、乙等無關獎金、陞遷條件、訓練進修機會與職務調動機會的差異）能提升接受乙等的人數，但仍改變不了最左側尖峰的顯著性，縱然其人數有所下降。
3. 若改變考列乙等的墊底狀況（題項 7～9）（乙等非少數、有丙等墊底、考列甲等沒那麼容易），最左側尖峰會進一步下降（仍是最多），但最右側尖峰也同時下降。反而中間的尖峰比最右側的尖峰高。這表示改變墊底的心理因素能進一步紓緩部分受測者對考列乙等的「極端厭惡感」，但改善接受考列乙等的程度不及調整利益條件。

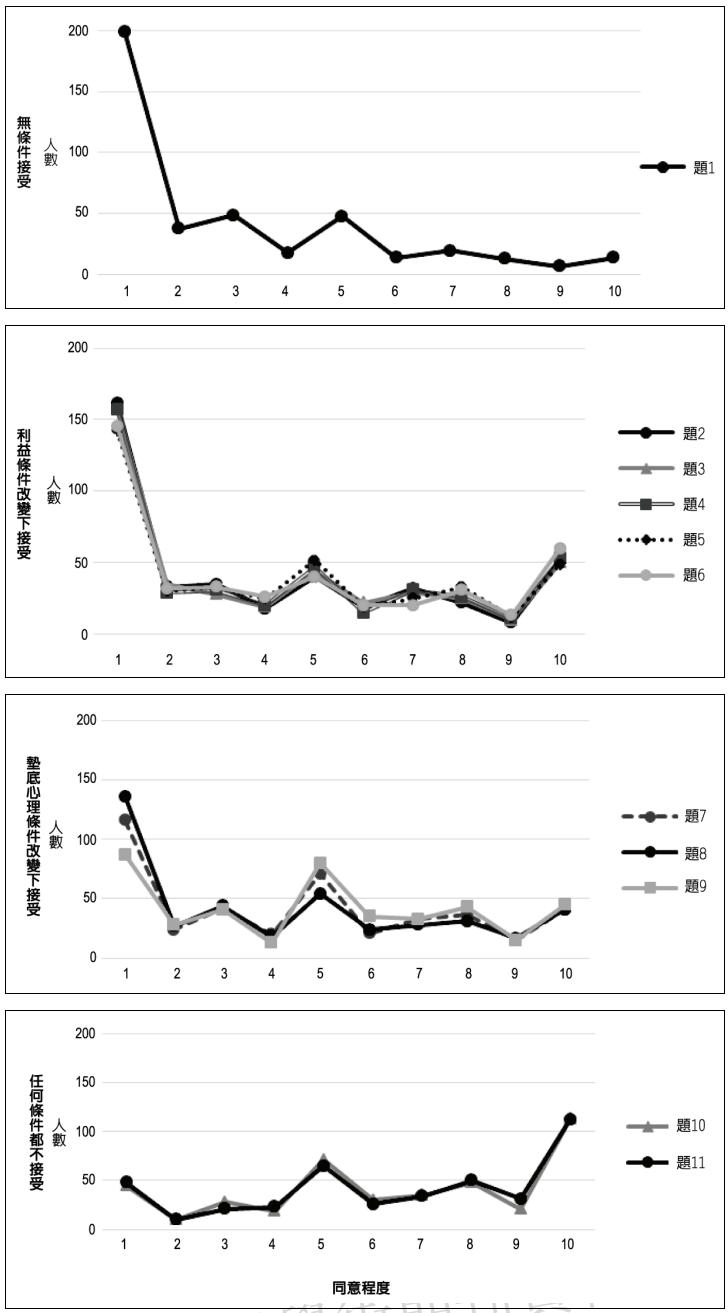
以上分析顯示，利益條件改變能微幅改善受測者對接受乙等的程度（右側尖峰上升），也微幅降低對乙等的厭惡度（左側尖峰下降），但對厭惡度之改善不及墊底心理條件之改變。然而，墊底心理條件的改變，於改善接受乙等程度上不及利益條件改變。無論如何，兩種條件之改變皆無法大幅度扭轉受測者對乙等的接受度。

而從題項 10 與 11 的填答回應看，對上述發現的理解會更爲明朗。其中最高尖峰轉至最右側，最左側的尖峰相對於題項 1 至 9 進一步下降。雖

臺灣大學學術期刊資料庫

圖 2

第一階段實驗中受測者對各題項表示同意程度的分布情形分析



資料來源：本研究自行繪製

臺灣人學學術期刊

然最右側的尖峰不及題項 1 至 8 的最左側尖峰高，但這反映了不少受測者對乙等的不接受情況已到達無法因考績制度設計調整而改變的程度，純粹因爲「乙等在同事心目中就是工作表現不好」或「我認爲乙等就代表工作表現不好」。

二、二分法分析

倘若將數據二分爲相對不同意（1～5）與相對同意（6～10）兩組，則會對上述尖峰分析發現之解讀有所調整（參看表 3）。因中間的尖峰屬於相對不同意的區間，會更凸顯利益條件與墊底心理條件的改變對提升樣本接受乙等的程度有限，相對同意度大多略高於 30%，且墊底心理條件的改變

表 3
第一階段實驗中受測者對各題項同意程度的二分法統計

干擾因子	題　　項	樣本數	平均數	標準差	相對不同意 (1-5分)	相對同意 (6-10分)
沒　有	1. 在以上情境下，我願意接受考績考列乙等。	420	3.01	2.563	352 (83.8%)	68 (16.2%)
考績獎金	2. 若甲、乙等無關考績獎金發給金額，我願意接受考績考列乙等。		4.07	3.256	287 (68.3%)	133 (31.7%)
考績升等	3. 若甲、乙等無關考績升等要件，我願意接受考績考列乙等。		4.13	3.250	284 (67.6%)	136 (32.4%)
陞遷機會	4. 若甲、乙等無關陞遷機會，我願意接受考績考列乙等。		4.21	3.287	284 (67.6%)	136 (32.4%)
訓練進修機　會	5. 若甲、乙等無關訓練進修機會，我願意接受考績考列乙等。		4.31	3.216	282 (67.1%)	138 (32.9%)
職務調動	6. 若甲、乙等無關職務調動之順利與否，我願意接受考績考列乙等。		4.37	3.345	276 (65.7%)	144 (34.3%)
我非少數	7. 若乙等不是少數人獲評之等第，我願意接受考績考列乙等。		4.58	3.065	273 (65.0%)	147 (35%)
有比我更糟的	8. 若有一定比例考列丙等的人員，我願意接受考績考列乙等。		4.32	3.140	279 (66.4%)	141 (33.6%)
優秀沒那麼容易	9. 若刪除現行公務人員考績法施行細則列舉之考列甲等一般條件規定，代表並非那麼容易符合甲等條件，我會願意接受考績乙等。		4.97	2.988	249 (59.3%)	171 (40.7%)
他人形象	10. 無論考績制度設計如何改變，我都不願意接受考績考列乙等，因為乙等在同事心目中就是工作表現不好。		6.45	3.021	173 (41.2%)	247 (58.8%)
自我形象	11. 無論考績制度設計如何改變，我都不願意接受考績考列乙等，因為我認為乙等就代表工作表現不好。		6.53	3.049	167 (39.8%)	253 (60.2%)

資料來源：本研究自行整理

臺灣大學學術期刊資料庫

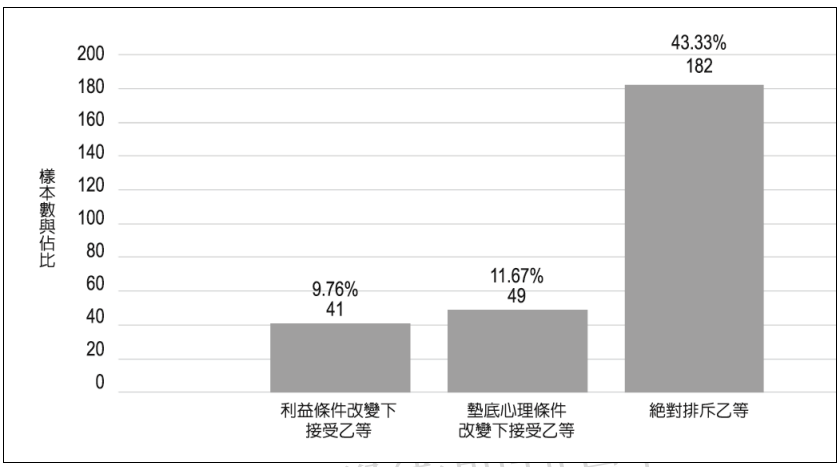
對提升接受乙等的程度稍高於利益條件的改變，題項 9 更達 40.7%。當然，這反映墊底心理條件改變的刺激性不至於使較多受測者填答 10 分，但仍可改善整體受測者接受乙等的幅度，且微幅高於利益條件改變。

上述分析是兩個條件個別產生的作用，本研究不排除兩者合併後，受測者有可能再提升接受度。然而從題項 10 與 11 可反過來推論，即便兩者合併也不太可能有過半受測者相對接受乙等。因爲有 58.8%（題項 10）或 60.2%（題項 11）的受測者相對同意「無論考績制度設計如何改變，我都不願意接受考績考列乙等」。

三、獨鍾單一條件類型的受測者

此外，本實驗發現有六成五比例受測者的相對同意填答（即 6 分以上）僅出現在上述某一條件類型的題項。這代表有不少比例的受測者剛好與本實驗的價值分類一致，且有獨鍾單一價值取向，他們不接受乙等的根本原因非常明確。除了極少數 5 個樣本只獨鍾題項 1 外（無條件接受乙等），有 272 個樣本獨鍾其餘三類題項。其中有 41 個樣本偏向只對利益條件的題項相對同意（題項 2～6），占整體樣本的 9.76%；49 個偏向只對墊底心理條件的題項相對同意（題項 7～9），占整體樣本的 11.67%；182 個偏向絕對排斥

圖 3
第一階段實驗中獨鍾單一條件類型群組的統計



資料來源：本研究自行繪製

臺灣大學學術期刊

乙等（題項 10～11），占整體樣本的 43.33%。堅定不接受考列乙等的一致取向者比例上雖不過半，但仍占有最高的比例，且在這些有明確取向者中占壓倒性比例。

四、受測者的背景變項分析

本實驗針對受測者的背景變項進行 t 檢定與單因子變異數分析，發現一些背景因素在某些題項的反應有顯著差異。受限於篇幅，本文不詳列內容，僅列出有顯著差異者。

1. 性別：在無條件接受考列乙等（題項 1）及考績獎金的條件設定下（題項 2），男性的同意程度顯著高於女性。
2. 年齡：在考績升等（題項 3）、陞遷機會（題項 4）、訓練進修機會（題項 5）、職務調動（題 6）、有一定比例考列丙等（題 8）、我認爲乙等就代表工作表現不好（題項 11）等條件設定下，不同年齡層有顯著差異。經多重比較分析（Scheffe 法），其中針對題項 3、4、5、6，受測樣本中 30 歲以下者的同意程度顯著高於 41 至 50 歲者。
3. 教育程度：僅在「我認爲乙等就代表工作表現不好」（題項 11）有顯著差異，惟多重比較分析結果顯示，各教育程度別之間的差異未達顯著。
4. 婚姻狀況：針對題項 3、4、5、6，未婚者的同意程度顯著較已婚者高。
5. 現任職務：僅在題項 6 有顯著差異，委任（警佐）及薦任（警正）受測者較在意職務調動順利與否。
6. 主管與非主管：僅在題項 6 有顯著差異，非主管樣本較在意職務調動順利與否。
7. 職系：在題項 1、2、5、6 有顯著差異。在沒有條件改變下，技術職系人員較願意接受考列乙等；在考績獎金、訓練進修及職務調動機會沒有差異的情況下，技術職系及警消人員較願意接受考列乙等。

從上述發現可以推論，男性較有可能接受乙等，也較在意獎金；未婚者較重視利益層次；中低階及非主管人員較重視申請調動的可能性，理由

臺灣大學學術期刊資料庫

可能是外補缺大多是針對中低階；技術職系與警消人員也較重視利益條件，這可能與兩者在專業加給標準較高有關，故考績獎金給與的差異較行政職系人員明顯，且其訓練進修名額（如升官等訓練）較少，另技術職系職缺較少、異動不易等也可能是原因。

不過，上段分析題項的同意程度平均值都低於 5，所以就算人員在這些變項上有顯著差異，也不代表改變相關條件（如利益條件）就能大幅提升乙等的接受程度。年齡因素則不同，30 歲以下的受測者對物質條件改變的反應最大，平均值都在 5 以上，最高達 6.06（題項 6：職務調動）。這反映年輕公務人員（可能包含不少未婚者）的物質或利益需求較明顯。

五、小結

經過上述第一階段實驗的多重分析，本研究有理由相信，大部分公務人員不接受考列乙等並非條件性反應，根本原因就是對「乙等」這符號本身已有所定見。只有極少數人員表示對乙等毫無條件地接受。部分人員可在物質條件改變或改變乙等墊底的格局之下改變態度。一些年輕公務人員或個別職種人員條件性地接受乙等，並非絕對否定。但難保這些年輕公務人員變成資深後不會進入「主流」。因此，在「甲乙丙丁」評等框架下，任何制度的調整，都不能逆轉整體公務人員不接受乙等的執著。

陸、研究發現(二)：揚甲貶乙

第二階段實驗在 2022 年 10 月至 2023 年 4 月期間實施。所有受測者各自獨立在網絡上完成前述研究設計提及十二篇作文的評核。其所看到的網頁資訊差異，僅在於香港及臺灣對照組內的受測者，與臺灣實驗組內的受測者所用的評核等次不一樣而已；前者爲甲、乙、丙、丁，後者爲傑出、優良、尚可、待改進。

受測者共 99 人，香港實驗組 30 人，臺灣對照組 35 人，臺灣實驗組 34 人，背景資料統計如表 4。整體參與的文官共 34 人、教師共 65 人，臺港兩地教師樣本皆有八成以上教授國文科。女性比率略高於男性，教育程度皆以大學／專與碩士居多。受限於研究者個人的關係網絡，香港實驗組

臺灣大學學術期刊資料庫

受試者的年齡與從事公／教職的時間略高於臺灣的其他兩組，且有兩人已退休（一位公務員與一位教師）。後續將透過統計方法釐析可能的影響，惟各組樣本數不高，仍難以排除因此導致變項差異的檢驗較容易達顯著的潛在問題，相關數據的解釋宜更謹慎。

表 4
第二階段實驗受測者的背景資料統計

背景資料		香港實驗組（甲乙丙丁）		臺灣對照組（甲乙丙丁）		臺灣實驗組（傑出、優良、尚可、待改進）	
		人數	百分比(%)	人數	百分比(%)	人數	百分比(%)
整體		30	100.0	35	100.0	34	100.0
性別	男性	11	36.7	13	37.1	15	44.1
	女性	19	63.3	22	62.9	19	55.9
年齡	未滿 20 歲	0	0.0	0	0.0	0	0.0
	20-24 歲	0	0.0	1	2.9	0	0.0
	25-29 歲	0	0.0	3	8.6	4	11.8
	30-34 歲	1	3.3	5	14.3	4	11.8
	35-39 歲	4	13.3	13	37.1	7	20.6
	40-44 歲	2	6.7	5	14.3	8	23.5
	45-49 歲	5	16.7	5	14.3	7	20.6
	50 歲以上	19	60.0	3	8.6	4	11.8
教育程度	大學／大專	16	53.3	12	34.3	8	23.5
	碩士	13	43.3	18	51.4	25	73.5
	博士	1	3.3	5	14.3	1	2.9
職業	公務員	9	30.0	13	37.1	12	35.3
	老師	21	70.0	22	62.9	22	64.7
從事公／教職時間	不到 1 年	0	0.0	1	2.9	1	2.9
	1 年以上，未滿 2 年	0	0.0	0	0.0	0	0.0
	2 年以上，未滿 3 年	0	0.0	3	8.6	0	0.0
	3 年以上，未滿 5 年	0	0.0	1	2.9	5	14.7
	5 年以上，未滿 10 年	2	6.7	9	25.7	5	14.7
	10 年以上，未滿 15 年	0	0.0	10	28.6	9	26.5
	15 年以上，未滿 20 年	7	23.3	5	14.3	7	20.6
	20 年以上	21	70.0	6	17.1	7	20.6

資料來源：本研究自行整理

一、三組針對各級別作文樣卷的整體評等分析與比較

本研究首先針對四個級別作文樣卷的評等結果做分析。六、五、四級分的九篇文章被評爲最高等次者，最高比率皆在臺灣對照組；而三級分的三篇文章被評爲最高等次者，最高比率則在香港實驗組。臺灣實驗組的評

臺灣大學學術期刊資料庫

核在三組中最爲嚴厲，在所有級分文章被評爲最高等次（傑出）的比率最低，且三級分文章更是唯一掛零。在四級分文章中，臺灣實驗組出現最多的評等爲「尚可」，其餘兩組出現最多爲「乙」。而三級分文章方面，臺灣實驗組出現最多的評等爲「待改進」（兩篇）與「尚可」（一篇），其餘兩組出現最多的爲「丙」（兩篇）與「乙」（一篇）。從以上分析可見，相對於其他兩組，臺灣對照組的受測者較傾向使用甲等評核最高三個級分的文章；香港實驗組僅針對三級分文章評核最寬鬆；臺灣實驗組則傾向以較低評等去評核各級別的文章（參看表 5）。

表 5
第二階段實驗中三個評核組對各作文樣卷的評等分布

作文卷樣與評核等次		香港實驗組（甲乙丙丁）		臺灣對照組（甲乙丙丁）		臺灣實驗組（傑出、優良、尚可、待改進）	
		人數	百分比(%)	人數	百分比(%)	人數	百分比(%)
原標準評級：六級分（評等標準中的最佳等次）							
6-1	傑出/甲	14	46.7	28	80.0	16	47.1
	優良/乙	12	40.0	6	17.1	16	47.1
	尚可/丙	4	13.3	1	2.9	2	5.9
	待改進/丁	0	0.0	0	0.0	0	0.0
6-2	傑出/甲	21	70.0	32	91.4	20	58.8
	優良/乙	9	30.0	2	5.7	11	32.4
	尚可/丙	0	0.0	1	2.9	3	8.8
	待改進/丁	0	0.0	0	0.0	0	0.0
6-3	傑出/甲	13	43.3	22	62.9	4	11.8
	優良/乙	12	40.0	11	31.4	27	79.4
	尚可/丙	5	16.7	2	5.7	2	5.9
	待改進/丁	0	0.0	0	0.0	1	2.9
原標準評級：五級分（評等標準中的次佳等次）							
5-1	傑出/甲	7	23.3	13	37.1	3	8.8
	優良/乙	20	66.7	20	57.1	21	61.8
	尚可/丙	3	10.0	2	5.7	10	29.4
	待改進/丁	0	0.0	0	0.0	0	0.0
5-2	傑出/甲	16	53.3	19	54.3	5	14.7
	優良/乙	12	40.0	14	40.0	23	67.6
	尚可/丙	2	6.7	2	5.7	6	17.6
	待改進/丁	0	0.0	0	0.0	0	0.0
5-3	傑出/甲	6	20.0	13	37.1	5	14.7
	優良/乙	18	60.0	18	51.4	20	58.8
	尚可/丙	6	20.0	4	11.4	7	20.6
	待改進/丁	0	0.0	0	0.0	2	5.9

臺灣大學學術期刊資料庫

表 5（續）

作文卷樣與評核等次		香港實驗組（甲乙丙丁）		臺灣對照組（甲乙丙丁）		臺灣實驗組（傑出、優良、尚可、待改進）	
		人　數	百分比(%)	人　數	百分比(%)	人　數	百分比(%)
原標準評級：四級分（評等標準中的第三等次）							
4-1	傑　出／甲	2	6.7	5	14.3	1	2.9
	優　良／乙	20	66.7	20	57.1	7	20.6
	尚　可／丙	8	26.7	9	25.7	24	70.6
	待改進／丁	0	0.0	1	2.9	2	5.9
4-2	傑　出／甲	4	13.3	7	20.0	1	2.9
	優　良／乙	17	56.7	21	60.0	12	35.3
	尚　可／丙	9	30.0	7	20.0	18	52.9
	待改進／丁	0	0.0	0	0.0	3	8.8
4-3	傑　出／甲	7	23.3	8	22.9	1	2.9
	優　良／乙	16	53.3	17	48.6	9	26.5
	尚　可／丙	6	20.0	9	25.7	23	67.6
	待改進／丁	1	3.3	1	2.9	1	2.9
原標準評級：三級分（評等標準中的第四等次）							
3-1	傑　出／甲	2	6.7	0	0.0	0	0.0
	優　良／乙	4	13.3	9	25.7	3	8.8
	尚　可／丙	17	56.7	18	51.4	10	29.4
	待改進／丁	7	23.3	8	22.9	21	61.8
3-2	傑　出／甲	8	26.7	5	14.3	0	0.0
	優　良／乙	11	36.7	15	42.9	7	20.6
	尚　可／丙	9	30.0	12	34.3	15	44.1
	待改進／丁	2	6.7	3	8.6	12	35.3
3-3	傑　出／甲	2	6.7	1	2.9	0	0.0
	優　良／乙	9	30.0	8	22.9	2	5.9
	尚　可／丙	13	43.3	21	60.0	14	41.2
	待改進／丁	6	20.0	5	14.3	18	52.9

資料來源：本研究自行整理

二、三組針對各級別作文樣卷評等的平均分與乙等／優良分析與比較

若以樣卷所屬四個級別對應三組所用的四個評等，將其分數化（最高評等爲 4 分，最低爲 1 分），並計算其平均，請參看圖 4。使用「甲乙丙丁」爲等次的兩組受測者，整體評分高於臺灣實驗組；其中臺灣對照組的評等整體又高於香港實驗組。評分結果由高至低順序爲：臺灣對照組、香港實驗組、臺灣實驗組。

三組當中，臺灣實驗組的評核結果與原評級最爲一致，但其針對六級分文章的評分低於原評級，四級及三級分文章的評分則略高於原評級。其

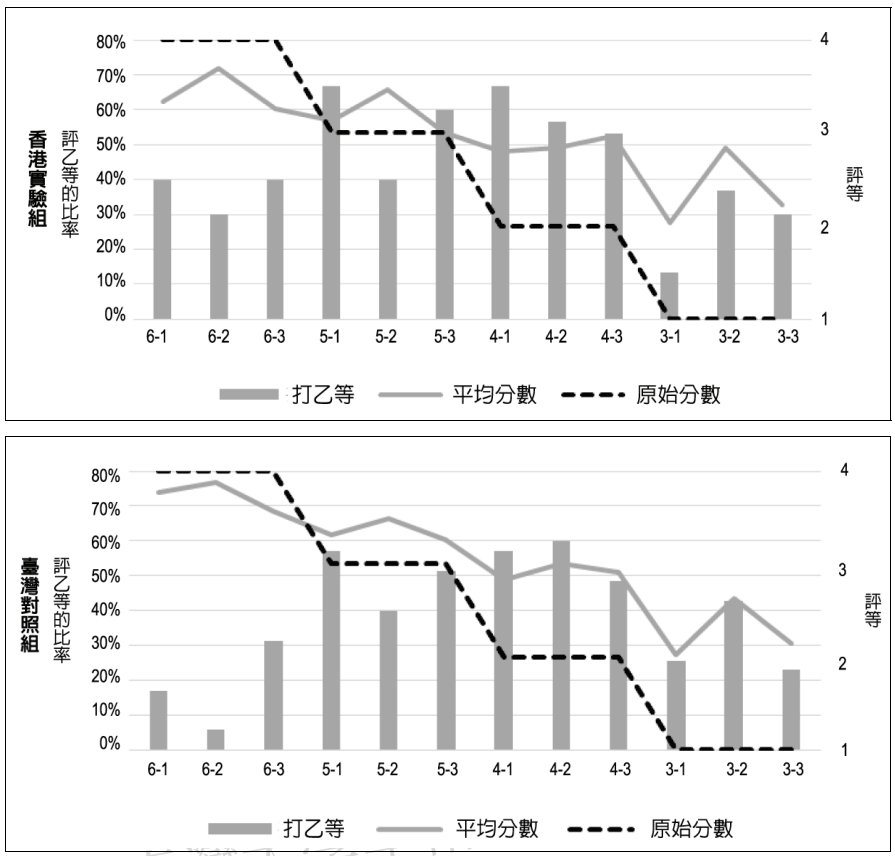
臺灣人學學術期刊資料庫

餘兩組在六級分文章的評分也略低於原評級，但落差不大，針對其餘三個級分文章的評分則比臺灣實驗組更高。其中臺灣對照組在五、四、三級分文章的評分明顯高於原評級，在五及四級分文章也略高於香港實驗組。

至於受測者評核文章給予乙等／優良評等的情形，若僅聚焦在六、五級分文章，香港實驗組較臺灣對照組會評文章爲乙等，僅文章 5-2 兩者比例相同，而臺灣對照組顯然較避免評六級分文章爲乙等。針對四級分文章，仍是香港實驗組較臺灣對照組會評文章爲乙等，其中在兩篇文章中前者比率高於後者，一篇則相反。到三級分文章，情況才逆轉，臺灣對照組有兩篇較多評乙等，一篇較少。相反地，就六、五級分文章，臺灣實驗組評「優良」（相當於乙等）的比例就較顯著，特別是相對於臺灣對照組（參看圖 4）。

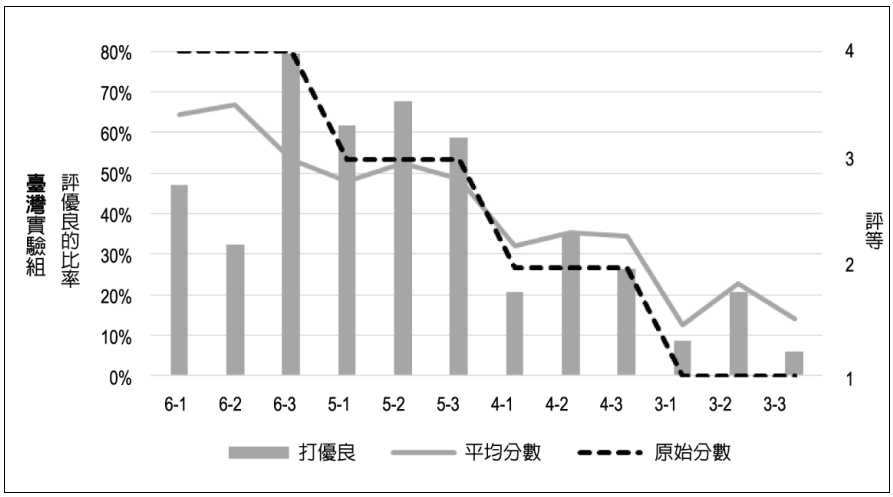
圖 4

第二階段實驗中三個評核組在各作文樣卷評乙等／優良及平均分數之比較



臺灣大學

圖 4（續）



資料來源：本研究自行繪製

備註：左側座標軸為評分乙等的比率，右側座標軸為評分等第，由 4 至 1 依序分別表示：甲 / 傑出、乙 / 優良、丙 / 尚可、丁 / 待改進的評等。

三、各組運用四個評核等次的盒鬚圖分析

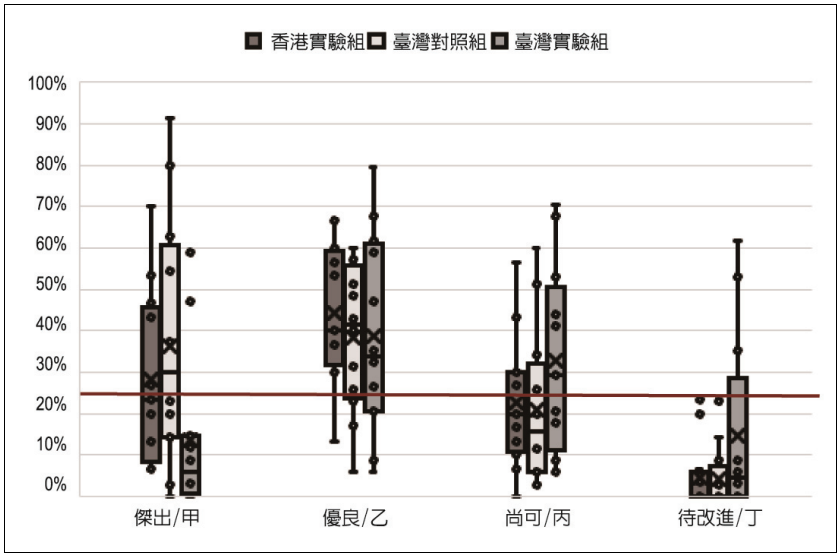
統計三組針對各文章評核四個等次的比率，以盒鬚圖呈現（參看圖 5），可更清楚反映四個等次在各組評核上的分布情況。原評級中，各評等分布皆爲 25%，以此爲期望值。針對評等標準中的最佳等次（評甲等 / 傑出），香港實驗組的中位數爲 23.3%，四分位距爲 8.3%～45.8%；臺灣對照組的中位數爲 30.0%，四分位距爲 14.3%～60.7%；臺灣實驗組的中位數爲 5.9%，四分位距爲 0.7%～14.7%。顯示臺灣對照組的主要分布顯著高於期望值，也高過其他兩組，臺灣實驗組則完全低於期望值，差別懸殊。

針對評等標準中的次佳等次（評乙等 / 優良），香港實驗組的中位數爲 40%，四分位距爲 31.7%～59.2%；臺灣對照組的中位數爲 41.4%，四分位距爲 23.6%～55.7%；臺灣實驗組的中位數爲 33.8%，四分位距爲 20.6%～61%。其中香港實驗組與臺灣實驗組的最大值皆明顯高於臺灣對照組的最大值，後者的最大值僅相若於其他兩組的上四分位數。

針對評等標準中的第三等次（評丙等 / 尚可），香港實驗組的中位數爲 20%，四分位距爲 10.8%～30%；臺灣對照組的中位數爲 15.7%，四分位距

臺灣大學圖書館期刊資料庫

圖 5
第二階段實驗中三組評核等次比率之盒鬚圖比較



資料來源：本研究自行繪製

爲 5.7%～32.1%；臺灣實驗組的中位數爲 29.4%，四分位距爲 11%～50.7%。顯見香港實驗組與臺灣對照組的主要分布都低於期望值，臺灣實驗組的主要分布則高於期望值，其上四分位數明顯高於另兩組的上四分位數。針對評等標準中的第四等次（評丁等／待改進），三組的主要分布皆低於期望值，其中香港實驗組與臺灣對照組的主要分布遠低於臺灣實驗組，上四分位數後者達 28.7%，其他兩組分別爲 5.8%（香港實驗組）及 7.1%（臺灣對照組）。

若將盒鬚圖中四個評核等次的主要分布連成一個圖像，會發現香港實驗組頗類似圖 1 中一般國家文官考績評等的左側較右側高的鐘型分布：乙等最高、甲等次之、丙等再次之、丁等最低；而臺灣對照組則類似圖 1 中的 L 型分布：甲最高、乙等次之、丙等再次之、丁等最低。臺灣實驗組也是鐘型分布，但是左側比右側低的嚴格評等。

綜上分析可見，臺灣對照組有明顯「揚甲」的傾向。雖然評乙等的比率並不低，但不要忘記，本實驗的作文樣卷包含較差的四與三級分文章，十二篇文章品質是有明顯差異。然而，該組高比率的乙等評等卻集中在 4 級

臺灣人學學術期刊資料庫

分文章，且比率最高不過 60%，甚至還有不低比率評甲等，要到三級分文章該組才主要以乙、丙等評核。

四、各作文樣卷題目的迴歸分析

最後，爲釐清在控制相關影響因素的情況下，各評核組的評核結果差異是否達顯著，本研究針對受評的十二篇文章進行迴歸分析（參看表 6），依變項爲四個評核等次。爲利理解與解釋，將其重新編碼爲 1 至 4，數字越高，評等越高。放入迴歸模型的變項包括年齡、性別、職業、教育程度及實驗組別。結果顯示多數文章的迴歸模型具有預測力，尤其在較低兩個級分所有文章迴歸分析結果皆達顯著。而模型的變項中，臺灣實驗組對比臺灣對照組在每篇文章的評等都顯著較低；香港實驗組在 6-3 及 5-3 兩篇的評等顯著低於臺灣對照組；女性在 5-1 篇評等顯著較男性低。

表 6
樣卷迴歸分析結果（依變項為四評等）

題　目	6-1	6-2	6-3	5-1	5-2	5-3	4-1	4-2	4-3	3-1	3-2	3-3
年　齡	-.224*	-.043	.067	-.166	.143	.220	.108	.067	.164	.155	-.012	-.016
性　別	.084	-.049	.065	-.192*	-.124	.001	-.032	.125	-.016	.006	-.088	-.079
職　業	.176	.028	-.002	.054	.102	.078	.038	.024	.094	.183	-.064	-.081
教育程度	-.100	.042	-.109	-.123	-.054	-.089	-.012	.005	-.020	-.006	-.158	-.059
G1	-.238	-.127	-.271**	-.079	-.109	-.312*	-.082	-.145	-.067	-.094	.076	.054
G2	-.247*	-.338**	-.140***	-.399***	-.406***	-.318**	-.450***	-.451***	-.405***	-.367***	-.404***	-.347***
樣本數	99	99	99	99	99	99	99	99	99	99	99	99
p	.001	.154	.022	.002	.005	.107	.003	.004	.004	.011	.000	.006
F	4.190	1.600	2.610	3.820	3.310	1.800	3.640	3.500	3.430	2.940	4.800	3.270
R^2	.215	.095	.146	.200	.178	.105	.192	.186	.183	.161	.238	.176

資料來源：本研究自行整理
註 1：G1 代表香港實驗組對比臺灣對照組；G2 代表臺灣實驗組對比臺灣對照組
註 2：p<.05*；p<.01**；p<.001***

就本研究所聚焦的乙等，我們再以「是否評爲乙等／優良」爲依變項進行邏輯迴歸分析（參看表 7），若評等爲第二等次，則依變項重新編碼爲 1；若非第二等次，則編碼爲 0。放入迴歸模型的變項同樣包括年齡、性別、職業、教育程度及實驗組別。結果顯示雖有較少模型達顯著，但在最高級分三篇文章的迴歸模型皆具預測力。模型的變項中，臺灣實驗組對比臺灣對照組在較高級分（五級分以上）評優良／乙等的情形顯著較高，在較低

臺灣大學學術期刊資料庫

級分（四級分以下）評優良／乙等的情形較低；香港實驗組與臺灣對照組評乙等的資料趨勢無明顯差異。這顯示在同樣以甲乙丙丁爲評等的情況下，地域差異對於各題目打乙等的影響未達顯著；而在臺灣脈絡中，使用「傑出、優良、尚可、待改進」評等者的評核結果，則多顯著較趨嚴格。此外，教師在 6-1 及 3-3 題評乙等的情形顯著低於公務員，年齡較長者在 6-2 題評乙等的情況較顯著。

表 7
樣卷迴歸分析結果（依變項為是否乙等）

題　目	6-1	6-2	6-3	5-1	5-2	5-3	4-1	4-2	4-3	3-1	3-2	3-3
年　齡	.157	.275*	.075	.065	.107	-.156	.204	-.002	-.044	-.048	-.078	-.062
性　別	-.117	.083	.027	.096	-.044	.092	-.002	.167	.157	.035	.098	-.154
職　業	-.207*	-.119	.071	.051	.022	.083	.030	.123	.210	.126*	-.104	-.208*
教育程度	.032	-.111	.027	.022	-.129	-.018	-.122	-.064	-.049	.118	-.100	-.011
G1	.160	.100	.040	.057	-.058	.121	-.052	-.054	.041	.053	-.035	.129
G2	.279*	.311**	.448***	.044	.247*	.092	-.370***	-.226*	-.201	-.165	-.208	-.205
樣本數	99	99	99	99	99	99	99	99	99	99	99	99
p	.006	.004	.003	.940	.223	.578	.002	.121	.578	.440	.280	.046
F	3.250	3.520	3.540	.291	1.400	.792	3.700	1.730	.792	.984	1.260	2.240
R^2	.175	.187	.188	.019	.084	.049	.195	.102	.049	.060	.076	.128

資料來源：本研究自行整理
註 1：G1 代表香港實驗組對比臺灣對照組、G2 代表臺灣實驗組對比臺灣對照組
註 2：p<.05*；p<.01**；p<.001***

五、小結

第二階段實驗並非直接與工作考績相關，而以評核者的角度評作文，且不涉及任何利害考慮的情況下，仍可見在「甲乙丙丁」格式下，臺灣對照組有評等膨漲的情形，相對其他兩組傾向給予較高評等（除了三級分文章低於香港實驗組），尤其在與臺灣實驗組的比較具有統計上的顯著差異。雖然香港實驗組與臺灣對照組的差異大多未有統計上的顯著性，但依其他方面的比較分析顯示，香港實驗組在評核上較趨中（以乙等爲準），而臺灣對照組則趨頂（甲等）。誠然，香港的受測者始終不太熟悉臺灣的國文寫作，在評核上可能因此趨於寬鬆，故在三級分文章上將評等拉高，但也不會因此給予太多甲等。以此爲據，對香港受測者而言，乙等應非不良。

反觀我國的受測者則有避開乙等之虞，對僅有一定水準，甚至水準平

臺灣大學學術期刊資料庫

平的文章，不少受測者都會給予甲等。然而，當受測者面對不同的評核等次──「傑出、優良、尚可、待改進」，我國的受測者就好像得到解放，評核趨於嚴格，且較接近原評核結果。如以文章原分級作爲客觀基準，則在新評等選項之下，我國的受測者會有較少評核謬誤，而用「甲乙丙丁」則有「趨頂寬濫謬誤」；香港的受測者則有「趨中謬誤」。基此，本研究有理由認爲，對乙等的偏見可能是一個地域現象。

柒、結論與政策建議

一、結論

設計思維在公部門之應用大多聚焦在爲民服務或對民眾有影響的措施上。本研究罕有地運用其於政府機關內部的人事管理制度上，因爲「以人爲本」的設計不應止於一般民眾，也可以用來透視機關的工作人員。而本研究首度嘗試以設計思維途徑去解構我國長期以來無法有效處理的公務人員考績問題。透過回溯性推理與反思式框架重建，挑戰了傳統通則性的理性受限與考績謬誤的思維框架，針對我國特殊的情境與脈絡，重新審視當事人的心理與行爲狀態，透過實驗做問題溯源的診斷。這種設計思維的邏輯可給予一些長期無法解決之棘手問題（可能是受到通則性理論所束縛）新的曙光，並因其給予問題不同的議題框架而產生創新性的解決方案。縱然這方案不一定能適用於其他國家，但這正是設計思維解決問題的特色與優勢。

本個案研究將公務人員考績問題焦點從過去一直執著的甲等轉移到一直被忽視的乙等，並批判地點出過去研究的盲點：爲何我國公務人員不接受乙等？透過兩階段的問卷調查實驗，謹愼剖析所得出的結論就是，我國公務人員，甚至是文官體制以外人員，也存在對乙等的固有偏見。亦即，乙等就是一個代表「不良」的符號。這偏見未必是全民性的，但具有一定的普遍性，且不會因利益或其他外在條件的改變而改變。

透過第一階段的情境性問卷調查實驗，初步發現過半公務人員無論如何都不接受考績乙等。爲進一步驗證這初步發現，本研究利用香港華人與改變評核選項的方式進行第二階段的問卷調查實驗，研究發現基本與第一

臺灣大學學術期刊資料庫

階段實驗的結果一致。因此，可以合理認爲，「乙等」的存在會對公務人員考績改革不利。而透過臺灣實驗組的反證，我們不僅可指出在「甲乙丙丁」格式下，考核工作會出現寬濫問題，同時也揭露出，僅改變考績評等選項，就有可能大大改善問題。

雖然本研究爲實務導向，不一定能將研究發現直接變成可廣泛適用的通則性理論，因乙等偏見是地域特殊現象，只適用於臺灣，不見得適用其他華人地區；惟回溯性推理也無意建立通則性理論，且本個案正好展示了設計思維在處理特殊個案的優勢，特別針對一些理性假設無法有效處理的棘手問題，可突破一般性理論或經驗的盲點，並促進政策創新。本研究突破理性受限的盲點，也就是理性受限可以普遍解釋考績謬誤，但其無法解釋爲何甲等是唯一理性選擇的疑惑。若香港受測者因有限理性而在評核上出現謬誤，其並不會否定乙等，但我國的則會，爲何如此？如果這也是謬誤，那麼這並非源於有限理性，而是受非理性偏見所影響。這是一般人力資源管理理論或國外經驗無法解釋的現象。本個案填補傳統考績謬誤理論不足之處，就是在考績中的認知偏誤不僅可導因於意欲理性（intended rationality）下的有限理性，也可在於固有價值造成的認知偏誤。

二、政策建議

根據以上研究發現，我們順理成章地建議我國公務人員考績應改變「甲乙丙丁」格式及任何在這格式上添加額外選項的做法（例如過去曾提出在甲等之上加上「優等」）。因爲依實驗結果發現，乙等早就被「污名化」了。而第二階段實驗中採用「傑出、優良、尚可、待改進」四個評等則能促使完全不同的行爲反應，可見僅操作推力式的選擇工程（choice architecture），即能對考績行爲產生槓桿效應（Hansen, 2016, pp.155-174）。

誠然，該調查僅對評核者一方進行了實驗，且非直接測試員工考績，故受評者對新選項的反應尚有未知之數。而當本文撰寫時，考試院正好提出公務人員考績法修正草案（2024 年 2 月 5 日），將考績評等改爲：傑出、優良、良好、待改善、汰除。這方案巧合地與本研究中實驗組的選項相近，而考試院在相關等次結構改變的說明上，也指出「表現稱職人員，[已] 將

臺灣大學學術期刊資料庫

考列乙等解讀爲機關認其表現較多數人差」，因此提出上述變革。[7] 本研究之發現可謂替上述的官方變革方案提供有力的佐證。

不過，修正草案並非預設第二等次（即優良）占多數比例，因爲草案對考列「傑出」與「優良」分別設上限爲前者 5% 及兩者累計 15%，而沒有對考列「良好」設比例限制。明顯地，草案預設第三等次（即良好）占多數比例的。如此的比例設限已引起公平性問題的質疑。[8]

因此，縱然本研究認爲上述推力式的選項設計改變是關鍵的一步，但並無意否定其他因素之影響力或干擾性，包括上述的比例設限及其他利益因素。其他的相關設計仍然是不可忽略的魔鬼細節。以現行制度而言，考列乙等的確對公務人員的生涯發展造成一定障礙，因爲要達到考績升等、升官等訓練及選送進修的門檻，皆要有甲等考績之紀錄。因此，如果在改革方案的設計中，僅有考列「良好」的稱職人員仍不能跨過一些生涯發展的基本門檻，那麼考列「良好」有可能變成現在乙等之翻版。

無論如何，本研究及建議並非是要徹底排除所有公務人員考績制度出現的問題，包括公務人員最在意的公平性問題，亦沒有現存研究能提供完全排除考核者偏誤的方案。不過，以本研究的焦點而言，現制實質上的甲／乙二分法，非好即差，不能容納中間評價的設計，是十分異常，也不合理，有必要進行調整。

參考書目

余致力等人（2003）。**公務人員考績制度改進之研究**。考試院委託研究案報告。[Yu, C.-L. et al. (2003). *Gongwu renyuan kaoji zhidu gaijin zhi yanjiu*. Kaoshiyuan weituo yanjiuan baogao].

林水波（1990）。**考績制度：理論研析與經驗印證**。行政院國家科學委員會資料中心。[Lin, S.-P. (1990). *Kaoji zhidu: lilun yanxi yu jingyan yinzheng*. Xingzheng yuan guojia kexue weiyuanhui ziliaozhongxin].

林立曼（2009）。98 年行政院人事主管會報考試院關院長中致詞全文。**人事月刊**，

[7] 草案條文內容，參看銓敘部（2024）。

[8] 參看台灣公務革新力量聯盟對草案的評論，曾執轡（2024）。

臺灣人事學術期刊資料庫

48(4)，9-13。[Lin, L.-M. (2009). 98 nian xingzhengyuan renshi zhuguan huibao Kaoshiyuan Guan yuanzhang Zhong zhici quanwen. *Personnel Monthly*, *48*(4), 9-13.]

施能傑（1992a）。考績謬誤的類型與原因：理論闡述與經驗分析（上）。**人事月刊，14**(3)，4-12。[Shih, N.-J. (1992a). Kaoji miuwu de leixing yu yuanyin: Lilun chanshu yu jingyan fenxi (shang). *Personnel Monthly*, *14*(3), 4-12.]

施能傑（1992b）。考績謬誤的類型與原因：理論闡述與經驗分析（下）。**人事月刊，14**(4)，4-16。[Shih, N.-J. (1992b). Kaoji miuwu de leixing yu yuanyin: Lilun chanshu yu jingyan fenxi (xia). *Personnel Monthly*, *14*(4), 4-16.]

施能傑（2024）。**公共人力資源管理制度的實踐：數據趨勢**。雙葉書廊。[Shih, N.-J. (2024). *Gonggong renli ziyuan guanli zhidu di shijian: Shuju qushi*. Yeh Yeh Book Gallery Publishing].

徐有守（1996）。**考銓新論**。臺灣商務印書館。[Hsu, Y.-S. (1996). *Kaoquan xinlun*. The Commercial Press Publishing].

徐有守（2006）。**官職併立制度的理論與結構：現行公務人事制度析論**。臺灣商務印書館。[Hsu, Y.-S. (2006). *Guanzhi bingli zhidu de lilun yu jiegou: Xianxing gongwu renshi zhidu xilun*. The Commercial Press Publishing].

徐有守、郭世良（2019）。**考銓制度（增修四版）**。五南。[Hsu, Y.-S., & Kuo, S.-L. (2019). *Kaoquan zhidu (zeng xiu si ban)*. Wu-Nan Book Inc.].

張瓊玲（2009）。改進公務人員考績制度的議題探討：公平原則觀點。**文官制度季刊，1**(2)，171-197。[Chang, C.-L. (2009). Current issues on civil service's performance appraisal system: Equal principles perspective. *Journal of Civil Service, 1*(2), 171-197.]

許道然、林文燦（2015）。**考銓制度**。國立空中大學。[Hsu, T.-J., & Lin, W.-T. (2015). *Kaoquan zhidu*. National Open University Publishing.]

陳敦源（2010）。**我國考績制度之檢視：從循證理論、正義觀點及參與管理角度檢視**。考試院委託研究案報告。[Chen, D.-Y. (2010). *Woguo kaoji zhidu zhi jianzhi: Cong xunzheng lilun, zhengyi guandian ji canyu guanli jiaodu jianzhi*. Kaoshiyuan weituo yanjiuan baogao].

陳敦源、林靜美（2005）。有限理性下的不完全契約：公部門績效管理制度的反思，**考銓季刊，43**，96-121。[Chen, D.-Y., & Lin, J.-M. (2005). Incomplete contract under bounded rationality: A reflection on the performance management system in the public sector. *Examination & Personnel Quarterly*, *43*, 96-121].

臺灣大學學術期刊資料庫

陳敦源、蘇孔志、簡鈺珒、陳序廷（2011）。論資排輩還是工作表現？年資因素對於我國公務人員績效考評影響之研究。**文官制度季刊，3**(1)，53-91。[Chen, D.-Y., Su, K.-C., Chien, Y.-C., & Chen, S.-T. (2011). Norm of seniority or performance appraisal? An analysis of the performance appraisal system (PAS) in Taiwanese public sector. *Journal of Civil Service*, *3*(1), 53-91.]

曾玟儀、蘇偉業（2017）。必然的謬誤與潛規則的合理性：制度約束下的地方主管年終考績行爲。**文官制度季刊，9**(2)，79-107。[Tseng, W.-Y., & So, B. W.-Y. (2017). Inevitability of errors, justifiability of hidden rules: Behavior of performance appraisal under institutional constraints. *Journal of Civil Service, 9*(2), 79-107.]

曾執轡（2024）。投書：公務員考績法重磅改革 解決不了 4 個舊問題。**上報**，3 月 12 日。https://www.upmedia.mg/news_info.php?Type=2&SerialNo=196544&fbclid=IwZXh0bgNhZW0CMTEAAR0P2ZPAVa9i5BY7lzXwRw2VO8wt2-l4etWMzpw5TiegrRimDoTw6Sl4uXQ_aem_AdnHBYWXneEqtEnA2jC1afrWdSIUd6BoOjtUZMZbhErpt46Y3At9RHHJi5d8cvYmJZVJH3DQWiDfZ9N-qVClU6iu [Tseng, Z.-P. (2024). Toushu: Gongwuyuan Kaojifa zhongbang gaige jiejue buliao 4 ge jiuwenti. *Up Media*, March 12.]

臺北市政府教育局終身教育科（2023）。**「2023 台灣燈會在台北」全國花燈競賽得獎名單**。臺北市政府教育局，2 月 4 日。https://www.doe.gov.taipei/News_Content.aspx?n=0CA2D00FCC591C52&sms=78D644F2755ACCAA&s=BE05411BB328B3F6 [The Continuing Education Division, Department of Education, Taipei City Government (2023). 2023 *Taiwan denghui zai Taipei quanguo huadeng jingsai dejiang mingdan*. Department of Education, Taipei City Government, February 4.]

銓敘部（2024）。**公務人員考績法修正草案再修正部分條文意見表**。嘉義縣政府 C.C.S. 人事服務平台，2 月 7 日。https://ccs.cyc.edu.tw/modules/tadnews/index.php?nsn=7135 [The Ministry of Civil Service, Republic of China. (2024). *Gongwu renyuan Kaojifa xiuzheng caoan zai xiu zheng bufen tiaowen yijian*. Chiayi County Government, C.C.S., February 7.]

銓敘部（未註明）。**銓敘部統計資料**。無日期。https://www.mocs.gov.tw/ [The Ministry of Civil Service, Republic of China (n.d.). *Main Civil Service Statistical Indicators*, n.d.]

Alexander, E. R. (1982). Design in the decision-making process. *Policy Sciences*, *14*(3), 279-292.

Bason, C. (Ed.). (2014). *Design for policy*. Gower Publishing.

臺灣大學學術期刊資料庫

Bason, C. (2018). *Leading public sector innovation* (2nd ed.). Policy Press.

Bekkers, V., Edelenbos, J., & Steijn, B. (Eds.) (2011). *Innovation in the public sector: Linking capacity and leadership*. Palgrave Macmillan.

Belle, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive biases in performance appraisal: Experimental evidence on anchoring and halo effects with public sector managers and employees. *Review of Public Personnel Administration*, *37*(3), 275-294.

Berger, J., Harbring, C., & Sliwka, D. (2013). Performance appraisal and the impact of forced distribution: An experimental investigation. *Management Science*, *59*(1), 54-68.

Berry, F. S., & Berry, W. D. (1999). Innovation and diffusion models in policy research. In P. A. Sabatier (Eds.), *Theories of the policy process* (pp. 169-200). Westview.

Brinkman, G., van Buuren, A., Voorberg, W., & van der Bijl-Brouwer, M. (2023). Making way for design thinking in the public sector: A taxonomy of strategies. *Policy Design and Practice*, *6*(3), 241-265.

Brown, T. (2019). *Change by design, revised and updated: How design thinking transforms organizations and inspires innovation.* Harper Collins.

Brown, T., & Wyatt, J. (2010). Design thinking for social innovation. *Stanford Social Innovation Review*, *8*(1), 31-35.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues*, *8*(2), 5-21.

Clarke, A., & Craft, J. (2017). The twin faces of public sector design. *Governance*, *32*(1), 5-21.

Collings, D. G. (2017). Workforce differentiation. In D. G. Collings, K. Mellahi, & W. F. Cascio (Eds.), *The Oxford handbook of talent management*, (pp. 299-317). Oxford University Press.

Dorst, K. (2011). The core of 'design thinking' and its application. *Design Studies*, *32*(6), 521-532.

Dorst, K. (2019). Co-evolution and emergence in design. *Design Studies*, *65*(C), 60-77.

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, *7*(1), 155-174.

Hassi, E. L. M., & Laakso, M. S. (2011). *Conceptions of design thinking in the*

臺灣大學學術期刊資料庫

management discourse [Conference presentation]. 9th European Academy of Design Conference, May 4-7, Porto, Portugal.

Hermus, M., van Buuren, A., & Bekkers, V. (2020). Applying design in public administration: a literature review to explore the state of the art. *Policy & Politics*, *48*(1), 21-48.

Howlett, M. (2020). Challenges in applying design thinking to public policy: Dealing with the varieties of policy formulation and their vicissitudes. *Policy & Politics*, *48*(1), 49-65.

Hunnes, M. G. (2016). Nudging: How human behavior is affected by design. *Annual Review of Policy Design*, *4*(1), 1-10.

Kong, D., Kim, S., & Yang, S.-B. (2013). *Individual performance appraisal in the government of Korea*. Ministry of Strategy and Finance.

Lewis, J. M., McGann, M., & Blomkamp, E. (2020). When design meets power: Design thinking, public sector innovation and the politics of policymaking. *Policy & Politics*, *48*(1), 111-130.

Lin Y.-C., & Kellough, J. E. (2019). Performance appraisal problems in the public sector: Examining supervisors' perceptions. *Public Personnel Management*, *48*(2), 179-202.

McGann, M., Blomkamp, E., & Lewis, J. M. (2018). The rise of public sector innovation labs: Experiments in design thinking for policy. *Policy Sciences*, *51*(3), 249-267.

Mintrom, M., & Luetjens, J. (2016). Design thinking in policymaking processes: Opportunities and challenges. *Australian Journal of Public Administration*, *75*(3), 391-402.

Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widła, A., & Park, Y. (2020). Policy labs: The next frontier of policy design and evaluation? *Policy & Politics*, *48*(1), 89-110.

Otani, M. (2023). Japan. In C.-M. Park, Y. Han, & Y. Chang (Eds.), *Civil service systems in East and Southeast Asia* (pp. 27-44). Routledge.

Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, *15*(3), 302-316.

Rubaltelli, E., Manicardi, D., Orsini, F., Mulatti, C., Rossi, R., & Lotto, L. (2021). How

臺灣大學學術期刊資料庫

to nudge drivers to reduce speed: The case of the left-digit effect. *Transportation Research Part F*, *78*, 259-266.

Schleicher, D. J., Bull, R. A., & Green, S. G. (2009). Rater reaction to forced distribution rating systems. *Journal of Management*, *35*(4), 899-927.

Serrat, O. (2017). The five whys technique. In O. Serrat (Ed.), *Knowledge solutions: Tools, methods, and approaches to drive organizational performance* (pp. 307-310). Springer.

Staroňová, K. (2017). *Performance appraisal in the EU member states and the European Commission*. Government Office of Slovakia.

van Buuren, A., Lewis, J. M., Peters, B. G., & Voorberg, W. (2020). Improving public policy and administration: Exploring the potential of design. *Policy & Politics*, *48*(1), 3-19.

臺灣大學學術期刊資料庫

Why Do Civil Servants not Accept Grade *Yi* Ratings in Performance Appraisals?

Issue Reframing and Innovation under Design Thinking*

*Bennis Wai-yip So***, *Yi-hua Lai****,
*& Yung-han Chang*******

Abstract

In Taiwan, there has long been a dilemma between the inflation of *Jia* grades in civil service appraisals, and civil servants' resistance to being rated grade *Yi*. The government has expediently imposed a cap on the proportion of *Jia* grades, but this has not resolved the problem. This study adopts a design thinking approach to reframe the issue, focusing on: Why do civil servants not accept grade *Yi* ratings? Through the administration of a situational survey for abductive reasoning and an essay-rating experiment involving civil servants and teachers in Taiwan and Hong Kong, this study finds that a bias possibly exists in Taiwan against grade *Yi* ratings, as it is considered to be a symbol of poor performance. This phenomenon cannot be explained by assumptions of rationality or any theories of human resource management. Accordingly, the government should redesign the grading categories instead of adjusting the existing designs of "*Jia*, *Yi*, *Bing* and *Ding*."

Keywords: Civil Service Appraisal, Design Thinking, Survey Experiment, Irrationality, Nudging

---

* DOI:10.6166/TJPS.202509_(105).0001

** Professor, Department of Public Administration, National Chengchi University.
E-mail: bennisso@nccu.edu.tw

*** Assistant Professor, Department of Political Science, Soochow University.
E-mail: evalai920@gmail.com

**** Master's degree holder of Eminent Public Administrators, College of Social Sciences, National Chengchi University.
E-mail: annie8062@gmail.com

臺灣人學學術期刊資料庫