民主與治理 第 12 卷第 2 期
2025 年 8 月，頁 1~44

公共管理實證研究中的同源誤差再思考：以臺灣公務人員考績資料為例

林俞君[1]、陳敦源[2]

摘要

自從同源誤差（Common Method Bias, CMB）遭到 International Public Management Journal（IPMJ）全面封殺，瞭解並面對 CMB 可能產生的問題已經是公共行政學者不能忽視的議題。然而學界對於 CMB 可能對模型估計產生的影響程度仍未有定論，台灣公共行政學界也尚未開始重視。本文為了增加公共行政學界對於 CMB 的認識，瞭解 CMB 對實證研究的影響、提升調查研究的資料水平，本文重現 Meier 與 O'Toole（2013）之研究設計，以臺灣公務人員考績為分析標的，比較典藏資料（archive data）與自評資料（self-report）兩種來源之差異。研究資料來自 2012 年考試院委託之「公部門陞遷制度之研究」，結合問卷與政府人事資料庫，有效樣本數 1,567。本研究分別以「歷年考績甲等比例」、「自評績效」及「2012 年考績分數」為依變數，比較三種模型估計結果。結果顯示，自評績效之錯估率高達 82%，顯示 CMB 對模型估計具有顯著影響。然而，典藏資料亦非解決估計誤差之萬靈丹。本文建議未來研究應從問題釐清、研究設計與統計修正

[1] 國立臺北大學公共行政暨政策學系助理教授

[2] 國立政治大學公共行政學系教授

投稿日期：2025 年 11 月 17 日　採用日期：2026 年 5 月 4 日
doi:10.30274/JDG.202508_12(2).0001

三個階段降低 CMB 影響並提升模型解釋力，同時不宜將 CMB 作為判斷研究優劣之唯一標準。

關鍵詞：同源資料誤差、績效評估、自評資料、典藏資料、研究設計

Journal of Democracy and Governance
August, 2025 , Vol. 12 No. 2, pp.1~44

Rethinking Common Method Bias in Public Management Empirical Research: Evidence from Civil Service Appraisal Study in Taiwan

Yu-Chun Lin[3], Don-Yun Chen[4]

Abstract

Since common method bias (CMB) has been comprehensively discouraged by the International Public Management Journal (IPMJ), understanding and addressing the potential problems caused by CMB has become an issue that public administration scholars can no longer ignore. However, the extent to which CMB affects model estimation remains inconclusive, and the issue has received limited attention in Taiwan's public administration community. To enhance scholarly understanding of CMB and examine its impact on empirical research, this study replicates the research design of Meier and O'Toole (2013) by analyzing civil servants' performance appraisal data in Taiwan and comparing differences between archival data and self-reported data. The dataset is drawn from the 2012 Examination Yuan–commissioned project, *An empirical study of promotion systems in the public sector: A justice-based examination,* which integrates survey responses with governmental personnel records, yielding 1,567 valid observations. Three dependent variables are examined, including the longitudinal proportion of top-tier appraisal ratings, self-reported performance, and the 2012 annual appraisal score. The results indicate that the estimation error rate of self-reported performance reaches 82%, suggesting that CMB substantially affects model estimation.

[3] Assistant Professor , Department of Public Administration and Policy, National Taipei University.

[4] Professor, Department of Public Administration, National Chengchi University
doi:10.30274/JDG.202508_12(2).0001

Nevertheless, archival data are not a panacea for resolving estimation errors. This study suggests that future research should mitigate CMB and improve model explanatory power through problem clarification, research design, and statistical adjustment, while cautioning against using CMB as the sole criterion for evaluating research quality.

Keywords: Common Method Bias (CMB), performance measurements, self-reported organizational performance, archived organizational performance, research design

壹、前言

公共行政領域的實證研究，無論是在台灣或是國外，長期以來大量仰賴問卷調查做為主要的研究資料來源。雖然這類的調查資料已經為公共行政學界累積了不少極具價值的研究成果，近年來使用單一調查做為研究資料卻引發不少爭議。原因在於，調查資料的變數測量多來自受訪者的對於自身感受的自我描述（self-report）；若自變數與依變數來自同一份調查問卷，可能研究結果會受到這些自我描述背後的隱性動機所干擾。此即同源資料誤差（Common Method Bias, CMB）。雖然 CMB 對於模型估計的干擾程度未有定論，但國際公共行政學界已有不少學者視之為嚴重問題。部分享有國際聲譽的期刊主編，例如 International Public Management Journal（IPMJ）的主編 Steven Kelman 即於 2015 年聲明，該期刊不接受疑似有同源資料誤差之文章。2016 年時任公共管理學會（PMRA）會長的 Don Moynihna 也曾在社群網站中直言「燒了那些有同源誤差的文章」（George & Pandey, 2017）。顯見國際公共行政學群對於 CMB 的重視。

主張封殺的一派認為，未來能稱得上「有效」（valid）的研究必須同時使用自評資料（self-report data）與非自評資料，例如典藏資料（archive data）。然而，全面性地封殺可能有同源誤差的文章，真的合理嗎？先不談同時獲取自評資料與非自評資料的成本有多高，更重要的是，很多時候非自評資料（例如典藏資料）未必就是正確的。例如政府機關的人事或組織資料庫，常常有未能即時更新或是缺漏的問題。要從這些典藏資料庫獲得有意義、可以進行分析的數據，往往比調查問卷更加困難。因此從理論上來說，這些統計工具、調查研究方法雖然不完整、有偏誤，並不代表不能夠回應研究現場的需要。基於這個理由，許多量化公共行政學者紛紛站出來發聲，其中以 George 和 Pandey 於 2017 年發表的文章最具代表性。George 和 Pandey 認為，自從同源誤差的問題被過度放大，對這個問題的焦慮已經讓許多人「只用膝蓋而不用頭腦思考」（axiomatic and knee-jerk）、見到黑影就開槍，甚至導致另一次的質量路線之爭（George & Pandey, 2017）。

截至目前為止，公共行政學界對於同源誤差的辯論依然沒有決定性的結論。我們只能透過更多的實證研究資料，比對自評資料與非自評資料在模型估計的差異，用更多的研究成果說服更多人究竟我們該何去何從。基於這個理由，本研究認為有三個問題相當關鍵：第一，瞭解同源誤差的來源、程度和影響為何；第二，從實證的角度檢視自評資料與非自評資料是否真的會導致不同的模型估計結果；

第三，若同源誤差的影響嚴重，未來該如何避免，若無法避免該如何補救、提升公共行政的研究品質。第一個問題本文透過文獻檢閱，回顧管理、心理以及公共行政等對同源誤差的研究論文，瞭解造成同源誤差的原因以及影響程度，以及學界目前的辯論。第二個問題，本研究以台灣公務人員的考績分數為例，比對自評資料庫（「公部門陞遷制度之研究」之調查問卷）與典藏資料庫（政府人事資料庫）的差異，以及在模型估計上是否有顯著不同。第三個研究問題則依照問題二的結論，與過去研究對話，討論未來研究應該進行的方向。本文期待能夠過這樣的討論讓同源誤差的問題獲得台灣公共行政學界的重視，並且能進一步提升未來的研究品質。

貳、文獻回顧

一、何謂同源誤差？

同源誤差（common source bias 或稱 common method variance）是一種系統性的測量誤差（Doty & Glick, 1998）。Richardson, Simmering & Sturman（2009, pp. 763）將同源誤差定義為「變數間存在系統性的誤差變異，而這個誤差是來自於變數被測量的方法或是資料的來源」；意指依變數和自變數之間的共變是來自於資料收集的過程，或受其他來源相同的變數的影響。換句話說，自變數與依變數之間的關係會受到第三個變數的影響，造成變數的測量誤差並非隨機，而是有系統性的（Meier & O'Toole, 2013; Richardson et al., 2009）；而這個影響可能放大或低估自變數與依變數之間的關係，導致模型估計錯誤。

更詳細地來說，同源誤差通常來自兩個主要的因素（Jakobsen & Jensen, 2015）。第一，自變數與依變數的資料來源相同，例如來自同一份問卷或相同的調查對象。相同的資料來源意味著若受訪者的作答方式有隱性的規律，這個規律將影響分析的正確性。舉例而言，受訪者可能在作答的時候選擇他認為別人比較能認同、認可他這個人的答案，而不是選擇他真實的感受（Podsakoff, MacKenzie, Lee, & Podsakoff, 2003）。此即「社會期望」（social desirability）（Crowne & Marlowe, 1960）的影響。受訪者也可能不管問題為何，皆傾向選擇中間的答案，不選擇最極端的答案（例如非常不同意或非常同意）（Bachman & O'Malley,

1984）。第二，變數測量的方式如果過於抽象、複雜或是模糊，亦可能導致同源誤差的產生。原因在於，受訪者面對語意模糊的問卷題目，容易帶入自己的主觀理解。這樣的下場有兩種：受訪者採用隨機的方式作答，或是受訪者的答題偏好（例如傾向依照社會期望作答，或避免選擇極端的答案）的影響將更為明顯（Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P., 2003, pp. 883）。

上述兩種情境都指出，受訪者在回答涉及態度、價值或自我評價的題項時，可能傾向提供符合社會期待或較容易被他人接受的答案，而非其真實感受，此即所謂「社會讚許」（social desirability）或「社會期望」偏誤（Crowne & Marlowe, 1960）。後續研究進一步指出，此類回應傾向可能同時包含印象管理（impression management）與自我欺騙（self-deception）兩種機制（Paulhus, 2002），進而使不同變數之間產生系統性的共變。

公共行政學界所討論的同源誤差，多屬上述第一種（c.f. Andersen, Heinesen, & Pedersen, 2016; Favero & Bullock, 2015; George & Pandey, 2017; Jakobsen & Jensen, 2015; Meier & O'Toole, 2013）：當自變數與依變數的來源相同，且變數為自我評估資料（self-report data）時，這樣的自評變數是否對資料分析產生影響。簡言之，當自變數與依變數皆來自同一份問卷，受訪者填問卷的模式，無論是受到問卷設計本身或是受訪者個人特質的影響、社會讚許或社會期望偏誤的影響，皆會使得問卷答案無法反映出受訪者真實的感受與態度，進而產生分析誤差。

二、同源誤差的影響顯著嗎？

過去同源誤差的研究主要集中在管理學界與心理學界，並且已經累積了 50 年左右的研究成果（Podsakoff et al., 2003）。然而，一般同意同源誤差開始真正受到學界重視並引起討論，是因為 Podsakoff 和 Organ 在 1986 所發表的文章"Self-Reports in Organizational Research: Problems and Prospects"（George & Pandey, 2017）。這篇文章討論在管理和組織研究經常使用的「自我描述」（self-report）的資料可能出現的問題，以及建議的解決方式。這篇文章雖然並非實證研究，但作者提供了可以檢視以及修正同源誤差的統計方法，包括哈門氏單因子測試法（Harman's one-factor test）、部分相關（partial correlation procedure）、去除社會期望因素（Elimination of social desirability）、 表題項修整（scale item trimming）等。然而 Podsakoff 和 Organ 也提到，統計方法無法完全的消除同源誤差的影響，因此最好的方式還是透過研究設計直接排除或降低同源誤差產生的可

能性。

受到 Podsakoff 和 Organ（1986）的影響，管理以及心理學界的學者開始透過實證研究檢視同源誤差的影響程度。其中有三篇研究分別用了同一組資料，但是卻做出了不同的結論。第一篇文章為 Spector（1987），該文針對 10 篇已發表的文章所使用的自我描述的資料做了，並做出了同源誤差的影響其實不大的結論（Spector, 1987）。第二篇文章為 Williams, Buckley & Cote（1989），該文用了多特質多方法模式分析（multitrait-multimethod, MTMM）同樣的資料來源，分析方法換成驗證性因素分析（confirmatory factor analysis, CFA），卻做出了同源誤差不容忽視、確實會影響分析結果的結論（Williams et al., 1989）。第三篇是 Bagozzi 和 Yi（1990），該文像是要為前兩篇文章的辯論做個裁判一般，再度分析了同樣一組資料，做出了同源誤差有時候有影響、有時候沒有影響的結論。除了這三篇研究的辯論，Keeping & Levy（2000）和 Williams, Gavin & Williams（1996）則是使用了同樣的研究方法驗證同源誤差的影響，卻得出截然相反的兩種結論：前者發現同源誤差沒有產生估計誤差，後者卻認為同源誤差個不容忽視的問題（Williams et al., 1996）。

不止不同的資料來源、不同的檢測方法，持續獲得不同的結論；相同的方法、相同的發現，不同研究者也做出了不同的解讀。Doty 和 Glick 於 1998 年檢視了六個社會科學期刊 12 年來的相關研究並使用 MTMM 進行後設分析（meta-analysis）。該研究發現同源誤差會導致變數之間的相關估計產生 26% 的誤差，而這個程度的誤差還不足以讓多數研究的結論失效（Doty & Glick, 1998; Meade, Watson, & Kroustalis, 2007）。有趣的是，Cote 和 Buckley 同樣使用 MTMM 分析 70 篇領域橫跨社會心理、行銷、企管和教育等領域的文章，同樣發現同源誤差會導致變數之間的相關估計產生大約 26% 的誤差。這個發現與 Doty 和 Glick（1998）一致，但是 Cote 和 Buckley 卻認為這影響巨大、不可忽視（Cote & Buckley, 1987）。Podsakoff 等人雖然同意 Cote 和 Buckley（1987）的判斷（Podsakoff et al., 2003），認為同源誤差的問題需要更深入的研究，但是不需要否定所有單一來源、自我描述性質的資料（Brannick, Chan, Conway, Lance, & Spector, 2010; Lindell & Whitney, 2001; Pace, 2010; Podsakoff et al., 2003）。

三、公共行政學者如何看待同源誤差

國際公共行政學界對同源誤差問題日益重視。原因在於公共行政的量化研究

大量仰賴自評資料的問卷來源，因此相當容易出現自變數與依變數來自於同一份問卷資料的情況。雖然公共行政國際重要期刊的審查已經相當重視同源誤差問題，但是相對於心理、管理領域針對同源誤差的影響以及修正有一定程度的討論，公共行政領域所累積的研究成果並不多，但所關注的焦點皆是使用單一資料庫的自評變數（self-reported data）是否對模型估計產生顯著的影響。目前已發表的研究文章主要有五篇（c.f. Andersen, Heinesen, & Pedersen, 2016; Favero & Bullock, 2015; George & Pandey, 2017; Jakobsen & Jensen, 2015; Meier & O'Toole, 2013）。有趣的是，不同於心理、管理領域對於同源誤差的態度正反兼具，而且能夠接受事後的修正，公共行政學者中，僅 George 和 Pandey（2017）認為學界過早將同源誤差妖魔化；其餘出來發聲的公共行政學者的立場幾乎一面倒，認為同源誤差應當要從研究設計就開始設法避免，且不認同以統計方法做事後修正。

Meier 和 O'Toole 在 2013 年發表的文章可以說是公共行政領域檢視同源誤差的先驅。他們從績效評估的效度談起，以美國德州高中學科能力測驗成績（Performance on the Texas Assessment of Knowledge, TAKS）的實際成績（典藏資料），比對學校主管自我評估的績效表現（自評資料）。比較的績效主要有三項：整體 TAKS 的表現、獲得大學入學許可學生的成績表現，以及學校在學區內的相對表現。研究發現在整體 TAKS 表現的比對所使用的 84 個指標中，自評績效評估分數有 23 個指標（27%）估計錯誤，包含 20 個呈現假性的正相關，3 個呈現假性的負相關。另外在以獲得大學入學許可學生的成績表現方現，自評績效評估分數有 42 個估計錯誤，錯誤率達 50%。基於這個發現，Meier 和 O'Toole 認為如果研究資料完全倚賴管理者自我評估的績效分數，對於研究結果將會造成非常嚴重的偏誤（Meier & O'Toole, 2013）。

Favero 和 Bullock 跟著 Meier 和 O'Toole 的腳步在 2015 年發表了公共行政領域第二篇有關同源誤差的研究文章。該研究分析了六篇發表於 Journal of Public Administration Research and Theory 並宣稱為了避開同源誤差問題而採用了一些補救措施的文章，以及這些補救措施的有效程度。他們發現公共行政學者面對潛在的同源誤差問題，通常有七種面對的方式：忽略問題（ignoring the problem）、調整解釋變數的方式（adjusting interpretation of variables）、哈門氏單因子測試法（Harman's one-factor test）、Brewer 所提出的樣本分離法（Brewer's split sample method）、控制問卷隱藏因子法（marker variables）、時間區別法（differencing），以及獨立資料來源補充法（independent source of data）。Favero 和 Bullock 發現，

除了獨立資料來源補充法（independent source of data）可以讓研究者從研究設計階段就直接將同源誤差的影響減到最低之外，其他的方式都各有限制而且成效有限。Favero 和 Bullock 另外建議使用結構方程式（SEM）來解決同源誤差的問題，因為結構方程式可以直接估計同源誤差（Favero & Bullock, 2015）。

Jakobsen 和 Jensen（2015）幾乎在同時與 Favero 和 Bnullock（2015）發表了類似的研究，同樣針對幾種常見的修正同源誤差的統計方式進行評估，並獲得相同的結論。然而 Jakobsen 和 Jensen 的這篇“Common Method Bias in Public Management Studies”卻掀起了相當大的波瀾。因為這篇文章，International Public Management Journal（IPMJ）的主編 Steve Kelman 特別在當期的期刊中發表聲明：由於沒有任何統計方法可以有效排除同源誤差的影響（Jakobsen & Jensen, 2015），IPMJ 將不再接受自變數和依變數來源相同的研究文章（Kelman, 2015）。

IPMJ 主編的重大決定引起了不同主張的學者的反擊，其中最具代表性的就是 George 和 Pandey 於 2017 發表的“We Know the Yin-But Where Is the Yang? Toward a Balanced Approach on Common Source Bias in Public Administration Scholarship”。George 和 Pandey 用新的資料重新複製 Spector（2006）的研究。他們假設，若存在第三變數同時影響自變數和依變數，必定會影響整份調查的所有變數，而且會與所有變數呈現顯著相關（Spector, 2006）。George 和 Pandey 依照這個定義，以公共服務動機為主要變數，搜尋 2007 年到 2015 年發表在 Review of Public Personnel Administration（ROPPA）與公共服務動機相關、且有呈現相關係數表格的文章。他們發現在 36 個相關係數當中，只有 25 個顯著，表示超過三成的相關係數不顯著；而且顯著的相關係數平均不超過 0.2，表示同源誤差的影響相當有限。因此，同源誤差不過就是一個都市傳說（urban legend）（George & Pandey, 2017; Spector, 2006）。

George 和 Pandey（2017）同時指出，公共行政領域所做的同源誤差研究，資料的同質性非常高：只有來自美國和丹麥的研究資料，而且全部使用學生的測驗成績當作比較基準。而他們重新以一般的組織行為資料分析，結果就有不同的研究發現。基於這個理由，George 和 Pandey（2017）對於公共行政學界將這四篇文章奉為圭臬的態度非常不以為然，認為許多學者過早將同源誤差妖魔化。而且，許多組織行為、人力資源管理的變數，凡舉個人的感受、觀點、判斷、本身就具備自我描述（self-report）的特質，這是無法避免，也無法以典藏資料取代的。因

此若能透過統計或研究設計盡量縮小同源誤差的影響，不應該因為使用了有同源誤差疑慮的資料來源就無條件的否定該研究的貢獻和效度。

目前台灣的公共行政學界尚無任何研究討論同源誤差的問題。本研究使用 google 學術與華藝線上圖書館搜尋引擎，以同源誤差、共同方法變異、common measurement bias、common measurement variance 等關鍵字搜尋，發現公共行政學界對於同源誤差的相關研究篇數為零。非公行領域的研究包括心理學界 1 篇（吳志文、葉光輝，2011），管理學界 1 篇（彭台光、高月慈、林鉦棽，2006），教育學界 1 篇（蕭佳純、涂志賢，2012），體育學界 1 篇（呂宛蓁、蕭嘉惠、許振明、曹校章、王學中，2012）。

心理學界的研究以強化個人的雙元自主性為研究主旨，認為透過提升個人的自由意志的展現，可以有效區辨哪些是屬於個人的獨立意見，哪些是屬於為了迎合群體才表達的和諧意見，藉此減少同源誤差的產生（吳志文、葉光輝，2011）。

管理學界的相關文章，作者為彭台光等人，發表於 2006 年管理學報。該文為一篇文獻分析文章，詳細的回顧了國際間同源誤差相關的論文，並討論了同源誤差問題的本質、影響與補救方法（彭台光等，2006），是國內被引用次數最多的相關論文。

教育學界的研究探究教師的創意教學衡量指標的同源誤差問題。該文採用 MTMM 與 CFA-MTMM 兩種分析方法檢測，發現教師創意教學衡量指標的同源誤差問題雖然存在，但不嚴重；原因可能在於創意教學的衡量指標較為具體的關係。這也意指不同的測量構面或不同領域的測量問題可能會有不同程度的同源誤差問題（蕭佳純、涂志，2012）。體育學界的兩篇文章主要進行統計方法的文獻回顧，並提出應該以隔離式資料蒐集法，例如時間 / 場域隔離法、方法隔離法等減少同源誤差的產生（呂宛蓁等，2012）。

從文獻上來看，心理、管理領域對於同源誤差的爭論雖然依舊持續，但大多數人依舊採納 Podsakoff et al.（2003）的觀點，認為不需全盤否定，並且可以透過適當的方法修正。公共行政領域相對之下對於同源誤差則是採取了較為嚴格的標準，認為所有建議的修正方式皆非完美，最好的方式就是一開始從研究設計上就避免使用有疑慮的資料來源。唯有 George 和 Pandey（2017）提出了公共行政學界在同源誤差的研究截至目前為止所使用的資料來源性質太過單一，不見得能夠推論至所有組織行為研究，更不能以此直接判斷所有使用自我描述資料的研究全無

貢獻和效度。至此，公共行政學界需要更多不同文化（非美國、丹麥）、不同主題（非以學生測驗成績為基準）的資料來源，重新檢視同源誤差的影響。

四、同源誤差對於「績效研究」的意義

進一步言之，在公共行政研究中，「績效」向來不是一個單一且穩定的概念。相較於企業部門多能以利潤或市場表現作為相對明確的績效指標，公共部門績效往往同時涉及效率、效果、回應性、合法性以及公共價值等不同面向，而不同治理脈絡下對績效的理解與強調亦可能有所差異（Bouckaert & Halligan, 2008; Talbot, 2010）。換言之，「績效」這個概念在政府機關可能不存在客觀事實，而更貼近是制度所建構出來的結果。Van Dooren、Bouckaert 與 Halligan（2015）即指出，績效指標的形成與選擇，本質上反映的是治理體系中的制度邏輯與管理需求，而非單純的技術性測量問題。

此外，績效衡量本身亦可能服務不同的治理目的。Behn（2003）指出，政府績效測量至少同時具有評估、控制、預算分配、激勵與組織學習等多重功能，而不同功能導向往往導致不同績效指標的選擇與使用方式。後續研究亦指出，績效管理制度並非中性的管理工具，而是政治與行政互動過程的一部分，其運作方式將進一步影響組織行為與決策模式（Moynihan, 2008; Pollitt, 2013）。

在此脈絡下，「客觀績效」與「主觀績效」之區分亦未必如方法論討論中所假定的那樣清楚。既有研究指出，主觀績效評估在某些情境下確實能反映組織或個體的實際表現，但同時也可能受到回應風格或認知偏誤影響（Bommer et al., 1995; Wall et al., 2004）。公共行政領域的研究亦發現，主觀與典藏績效指標之間經常存在顯著落差（Andersen et al., 2016）。因此，績效測量來源之差異，未必完全代表測量錯誤，也可能反映不同制度脈絡下對績效概念的理解差異。

基於上述討論，本文認為，同源誤差在公共行政研究中的意義，並不僅止於資料來源相同所造成的統計估計偏誤，而是與公共行政學界如何界定與操作化「績效」概念密切相關。當績效本身具有制度依賴性與多重面向特質時，不同資料來源所產生的模型估計差異，既可能來自方法偏誤，也可能反映績效概念建構上的差異。從此角度出發，重新檢視同源誤差問題，亦有助於釐清公共行政實證研究中績效測量的理論基礎。

參、研究設計與分析方法

使用同一資料庫的自評變數造成的同源誤差，需要更多來自不同文化、不同主題的資料來驗證其影響程度。基於這個目的，本研究重現 Meier 和 O'Toole（2013）的研究方法，改以台灣公務人員考績為例，蒐集「典藏考績」（真實考績）與「自評考績」（自評績效）兩種來源的績效評估資料，並以此回答三個研究問題：（一）典藏資料庫中的考績分數與自評的考績分數是否有差別？（二）在迴歸模型中分別使用典藏考績與自評考績做為依變數，估計台灣公務人員工作績效的影響因素，對於模型的估計有多大影響？（三）未來公共行政的量化研究該如何處理同源誤差的影響？

一、選擇「績效」做為研究同源誤差問題的原因

只要是自變數與依變數皆來自同一份自我評估的調查資料，研究的統計分析便有同源誤差的疑慮，因此同源誤差的問題可能發生在任何自我評估的變數。本研究選擇「績效」做為研究同源誤差的標的，是因為目前公共行政學界其他研究同源誤差多以績效為研究主軸（c.f. Andersen, Heinesen, & Pedersen, 2016; Favero & Bullock, 2015; Meier & O'Toole, 2013），且尚未獲得決定性的結論。

公共行政學界過去之所以以績效為探討同源誤差的主軸，最主要的原因是，公共行政學界一直以來缺乏一個公認客觀可信任的績效衡量標準，偏偏績效的衡量卻是領域中數一數二重要的議題。績效的測量一直是「研究者希望盡量趨近於客觀，且所有人都還在尋找最能接近客觀的那個指標」的狀態。因為客觀績效指標難尋，公共行政學界長年來多使用自評績效當作績效評估指標，做為客觀績效不可得時的替代品，並依此結果建立的許多理論模型。然而這樣的資料來源正是可能產生同源誤差的原因；因為自評績效通常做為依變數使用，而自變數得測量當然也來自於同一份調查，可能許多內衍的因素會影響分析的結果。學界對於同源誤差的批評，直接威脅到的就是這些以自評績效當作主要參考指標的研究結果。因此比對客觀績效指標與主觀自評績效對於模型估計的結果是絕對必要的。

過去這些以績效做為同源誤差研究標的的文章，其研究對象皆為公立學校的教師，不包含其他公務人員。教師與其他政府公務人員的績效測量其實有本質上的差異。首先，教師對學生的影響力超過一般公務人員對一般民眾的影響力。教

師的服務對象為學生，教師與學生的互動頻繁，與一般公務人員對其服務對象多為一次性的接觸完全不同。再者，教師與學生（尤其是高中以下的學生）之間的權力關係並不對等，因此教師與學生的關係無法和公務人員與民眾進行類比。另外，學生的考試分數最多也只能是間接測量指標，而且尚有學生個別的因素可能影響測量（例如學生本身的資質、努力程度、家庭因素等），因此學生的測驗成績與其說是教師的績效指標，更應該是學生本身的績效指標；更遑論教師的工作性質與一般公務員殊異。自然，使用學生成績做為教師績效評量的非自評指標，並以此測量同源誤差的例子，無法理所當然地直接套用到所有公務人員的績效評量問題上。一般公務人員績效評量所面臨的同源誤差問題，仍需要直接蒐集一般公務機關的資料進行研究。正因為如此，George 和 Pandey（2017）認為若要理解同源誤差在公共行政學界的影響，除了學校教師的績效之外，尚須研究其他身份的公務人員。

值得進一步指出的是，在公共行政領域中，「績效」本身並非一個穩定、單一且純技術性的概念，而是一個深受制度安排、評估機制與組織文化所形塑的建構性概念。不同於企業績效可能以利潤或市場指標為核心，公共部門績效往往同時涵蓋法規遵循、程序正當性、政治責任、公共價值與服務成果等多重面向。績效評估因此不僅是一種技術測量問題，更涉及公共組織的運作邏輯。

在此脈絡下，同源誤差問題便不再僅僅是資料來源是否相同的技術性爭議，而是觸及公共行政研究如何界定與操作化績效概念的根本問題。若績效本身具有多重面向與制度脈絡依賴性，那麼不同資料來源（自評、典藏、升遷結果）之間的差異，未必完全可歸因於測量誤差，也可能反映不同制度邏輯下的績效定義差異。因此，檢視同源誤差的影響，同時也是對公共行政績效測量概念進行反思的過程。

二、自評考績與典藏考績分數測量與資料來源

本研究以「歷年考績甲等比例」做為典藏考績分數。「歷年考績甲等比例」代表每個公務人員在政府的人事資料庫[5]中所記載的歷年考績資料中的甲等比例。

[5] 這個資料庫來自於政治大學陳敦源教授 2012 年接受考試院委託執行之「公部門陞遷制度之研究」。該資料庫以到 2012 年底為止，登錄在全國人力資料庫和銓審資料庫中的公務人員以匿名代碼的方式為研究母群，簡單隨機抽樣 1970 份樣本；請考試院依照樣本名單寄出調查問卷，回收之問卷與人力資料庫以代碼的方式對應整合。過程中研究團隊並無接觸人力資料庫的任何具名資料。總計回收 1,567 份有效樣本，回收率 79.54%（陳敦源等，2013；張鎧如等，

舉例而言，A 公務人員在 19 年公職生涯中，有 18 年考績甲等，則其歷年考績甲等比例為 94.7%；B 公務人員在 19 年公職生涯中，有 15 年考績甲等，則其歷年考績甲等比例為 78.9%；以此類推。雖然考績分數一般認為受許多潛規則（來自於制度設計、長期習慣等非正式制度運作機制）的影響，例如性別、年紀，或以輪流甲等的方式進行（陳敦源，2010）；然而此為橫跨 19 年的考績資料，在本文資料脈絡下，可以將其影響視為相對穩定的背景因素（張鎧如、陳敦源、簡鈺珒、李仲彬，2015）。意思是，雖然有潛規則的干擾，但依照真實績效給予考績，依然是主要的考量[6]。換言之，若真的是績效良好的員工，儘管部分時候因為潛規則的干擾未能拿到甲等，但時間拉長之後拿到甲等的機率依然會比其他人高，因此長期累積的考績結果尚能代表員工的真實績效表現。因此將歷年考績做為最接近真實績效的非自評績效變數，即便依然與真實績效有落差，也足以做為真實績效的替代變數（proxy）（張鎧如等，2015）。此為歷年考績分數做為典藏考績分數的原因。

自評考績分數為「自評績效」，來自於 2012 年「公部門陞遷制度之研究」之調查問卷。自評績效的測量由三個題項組成：我的工作品質能獲得主管的肯定、我的工作品質能獲得同仁的肯定、我的整體工作表現十分優異。三個題項的測量方式皆為李克特式量表六點尺度，1 分代表非常不同意，6 分則是非常同意。三個題項的分數加總後，為公務人員自我評估的績效。

由於自評考績分數應是針對資料蒐集當年（2012）的績效做為判斷，可能與「歷年考績甲等比例」的成績有時間單位差異的問題（單一年度的考績無法與歷年考績做比較），特別再加入「2012 年考績分數」做為參照的標準。資料來源為 2012 年的政府人事資料庫，屬於典藏資料。

雖然這三種績效測量方式（歷年績效、2012 考績、2012 自評績效）的內涵各有著重（例如歷年績效是長時間的績效觀察，2012 單一年度的績效可能會受當其

2015）。雖然個別受訪者的政府人事資料庫資料與問卷調查結果可以對應，但因為調查之初就以匿名化處理，因此研究人員並無法辨識個別樣本。

[6] 根據 2019 台灣文官調查（TGBS，2019），在「對組織的貢獻、目標的完成度、出勤狀況、官職等、年資、考績輪流、個人陞遷或獎金需求」等數個考績評比的可能指標中，無論是員工或是主管排序，對組織的貢獻程度、目標的完成度等兩個與真實績效相關的指標，皆為偏好程度最高的兩個選項。顯見儘管有潛規則干擾，真實的績效依然是主管、以及員工認為主管心目中最重要的標準。基於這個理由，在未有其他更能代表真實績效的指標之前，本研究認為歷年考績甲等比例是可以暫時被視為最接近真實績效的指標。

因素左右，而自評績效則可能反映受評者的個人特質），但若我們先建立一個前提：「歷年績效應該是最接近真實績效的指標」，那麼 2012 考績的功能在於提供單年度的參照，避免 2012 的表現相較於歷年的績效來說距離平均值太遠。因此，若 2012 考績的結果較接近歷年績效，則我們可以肯定兩個典藏資料的績效評估結果是比較接近的，增加以歷年績效做為衡量指標的可信程度，也提升了典藏資料做為客觀績效指標的可信程度。在這個情況下，自評績效與 2012 考績、歷年績效的落差，就可以歸因於自評績效本身的測量誤差。反之，若 2012 考績與自評績效較為接近，則說明 2012 這年的考績是變異比較大的一年，無法和歷年績效進行比較，自然也無法證實同源誤差可能產生的影響。

三、自評考績與典藏考績比較方法

為了能夠判斷這三項績效指標的相關程度，本研究先行以 Pearson 相關係數評估歷年考績甲等比例、2012 年考績分數、與 2012 年自評績效分數三者的相關程度。雖然這三個指標的單位不同，但三者皆可視為連續變項，因此以 Pearson 相關係數進行初步的分析。

四、自評考績與典藏考績分數對模型估計的影響

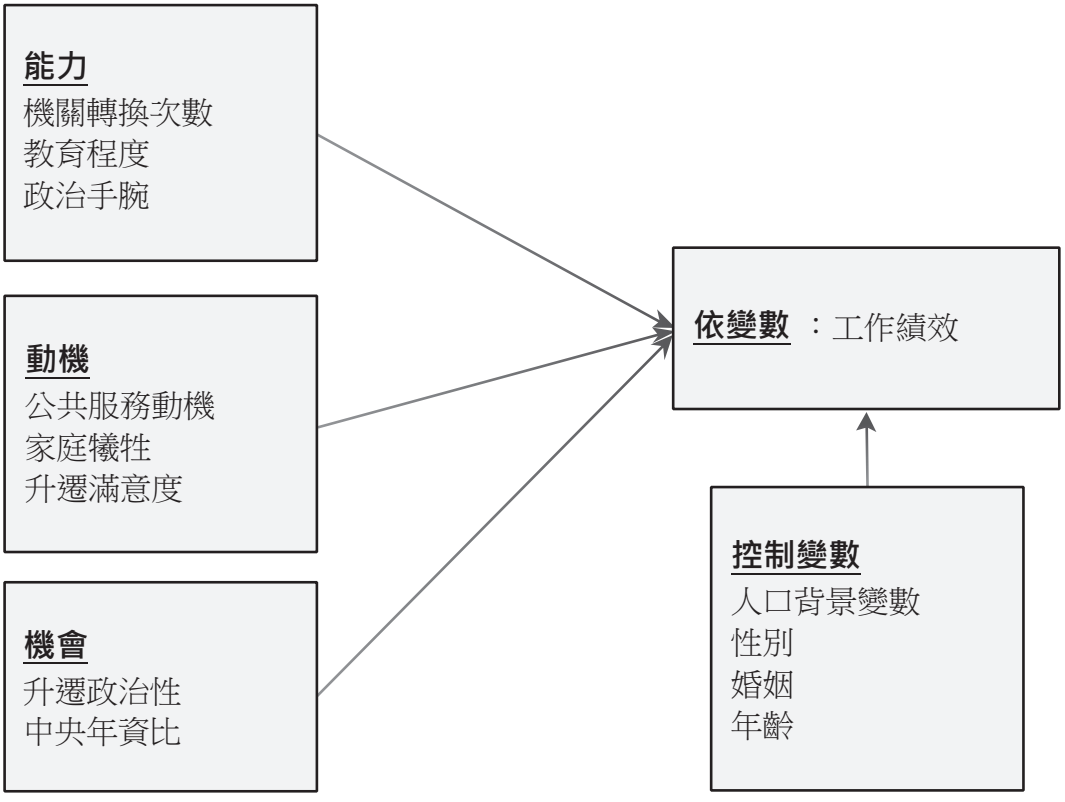
為了瞭解同源誤差對模型估計的影響，本研究複製張鎧如等人於 2015 年發表於東吳政治學報的「探索台灣公務人員工作績效之影響因素：能力、動機與機會的整合觀點」之中的研究模型。之所以以這篇文章的模型為基準進行比較，主要原因有二。其一，這篇文章為了避免同源誤差產生模型估計偏誤，使用不同來源的自變數與依變數。自變數為自評資料，來自於 2012 年考試院委託國立政治大學公共行政學系陳敦源教授所主持的「公部門陞遷制度之研究」之調查問卷（陳敦源，2013）；依變數則為典藏資料，來自於 1988 年以後的政府人事資料庫。因為這樣的研究設計，本研究得以在張鎧如等（2015）的模型基礎上，繼續探討同源誤差的影響，同時討論張鎧如等（2015）所使用的避開同源誤差的方法是否有效。

其二，本研究的主旨在於探討同源誤差對於模型估計的影響，績效評估僅是為了測量估計誤差所選用的主題，自變數與依變數如何相互影響並非本研究探究的重點。然而，一個建構完整的模型，能幫助本研究排除因為模型建構不

佳而引起的估計誤差。張鎧如等（2015）所驗證的模型，由 Blumberg 和 Pringle 於 1982 年發表，描述個人工作績效的差異，取決於個人能力（Capacity）、動機（Motivation）與機會（Opportunity）的綜合影響（Blumberg & Pringle, 1982）。這個模型自 1982 年發表以來已經經過多許多的驗證與討論，其解釋力已被廣泛地 接 受（Appelbaum, Bailey, Berg, & Kalleberg, 2000; Bayo-Moriones & Galdon-Sanchez, 2010; Boselie, 2010; Boxall & Purcell, 2003; Knies & Leisink, 2014）。張鎧如等（2015）的研究也證明了這個模型能夠解釋台灣的公務人員績效。

圖 1

研究模型



資料來源：探索台灣公務人員工作績效之影響因素：能力、動機與機會的整合觀點。張鎧如等，2015，東吳政治學報，33(4)，頁 24。

模型的依變數為公務人員考績，自變數則包含個人能力（Capacity）、動機（Motivation）與機會（Opportunity）等三個面向，並控制人口背景變數。張鎧如等（2015）依照 Blumberg & Pringle（1982）的定義，使用「歷年考績甲等比例」以及「2012 職等」兩個變數做為模型的依變數，分別代表不同的工作績效測量方式；並將自變數的三個面向操作化。以下針對模型各變數間的關係稍做說明。

依照張鎧如（2015）的詮釋，能力變數包含機關轉換次數、教育程度與政治手腕。機關轉換次數高，代表該名公務人員有能力在不同機關之間獲得商調的機會，這是對能力的認可，因此機關轉換次數多，有機會正向影響其工作績效。學歷（教育程度）通常在職場上也是常見的能力面向指標。受過高等教育者，通常被假設有較多的學識能力，能夠處理較為複雜的業務內容，因此教育程度所代表的能力面向，也被預期將正向影響工作績效。政治手腕則強調政治技能對公務人員在組織當中的重要程度，政治手腕意味著在組織中的溝通與影響能力，亦對工作績效有正向的影響（張鎧如等，2015）。

動機變數包含公共服務動機、家庭犧牲、升遷滿意度。公共服務動機為常見的公務人員動機變數，過去研究認為，公共服務動機高者，通常有較高的工作投入以及較高的主動程度，並且能有較好的工作績效。家庭犧牲意味著公務人員以犧牲家庭時間換取更多的工作時間，因此家庭犧牲的程度亦可能對工作績效有正面影響。升遷滿意度意指公務人員是否將組織內的升遷制度或文化視為合理的、有激勵效果的；升遷滿意度高，意味著公務人員滿意目前的升遷環境，並比較容易受到這個制度的激勵，進而提高工作績效（張鎧如等，2015）。

機會變數則包含升遷政治性以及中央年資比。升遷政治性代表公務人員所感受到的組織內的政治氛圍是否可能影響其升遷的機會；通常升遷政治性對工作績效的影響是負面的，因為政治力量的影響愈多，實際工作績效表現對升遷的影響愈小，降低人員提升績效的動力。中央年資比反應公務人員在中央機關任職的時間長度；通常在中央機關有較多的升遷機會，因此有較高的機會激勵人員提昇工作績效（張鎧如，2015）。控制的人口背景變數包含性別、婚姻以及年齡（見圖 1）（變數的操作化請見附錄一）。

本文延續張鎧如等（2015）對於變數的定義與操作化方式，僅針對依變項的部分依照本研究的需要調整。為了比較自評資料與典藏資料在模型估計上的差異，本研究依照依變數的來源分為三個統計模型進行比較。第一個模型依變數為「歷年考績甲等比例」，為典藏資料。第二個模型依變數為「自評績效」，為自評考

績分數。第三個模型依變數為「2012 年考績分數」，為典藏資料；第三個模型是為了避免單一年度的考績無法與歷年考績做比較，特此入「2012 年考績分數」做為參照的標準。

五、分析方法

本研究採用 Meier 和 O'Toole（2013）的方法評估自評資料與典藏資料在模型估計上的落差。Meier 和 O'Toole（2013）是少數從模型估計進行同源誤差研究的文章。多數研究直接從變數的效度探討同源誤差的問題，使用的方法包括哈門氏單因子測試法（Harman's one-factor test）、部分相關（partial correlation procedure）、多特質多方法模式分析（MTMM）等。然而這些統計方法其實都屬於變數的效度分析，沒有針對同源誤差的偏誤是否會在模型估計中造成整體模型的適合度和迴歸係數的估計問題進行討論。此外，Meier 和 O'Toole（2013）所使用研究對象為公立學校教師，其績效測量的指標為學生的考試成績；正如同本文前述所言，以學生考試成績做為老師的績效指標與其他政府機關公務人員的績效測量方式差異甚大，無法直接推論，因此本研究認為直接以公務人員的績效作為直接研究資料實屬必要。

Meier 和 O'Toole（2013）以測量理論（measurement theory）為基礎，發展出一套透過比較迴歸係數的 t 值評估同源誤差影響的方法。研究者必須建立兩個以上的迴歸模型，模型結構相同，只有變數的測量方式不同。Meier 和 O'Toole（2013）以教師的績效評量做為依變數，並依照不同的依變數來源建立了不同的迴歸模型；典藏資料模型的依變數為學生的考試成績，自評資料模型的依變數則是教師的績效自我評估。比較自評資料模型與典藏資料模型自變數的迴歸係數 t 值[7]，即可得知資料的測量方式是否會影響模型估計。舉例而言，假設典藏

[7] 值得一提的是，除了同源誤差可能導致 t 值的估計差異，尚有其他因素也可能讓 t 值產生變化。舉例來說，黃旻華（2006）指出，即便針對同一組問卷題目，採取不同的量度法（scaling methods），並控制其他模型變數不變，t 值依然會產生變化，因此 t 值的變化有可能只是因為採取不同測量尺度所造成的結果。黃旻華（2006）一文建議，正因為不同的測量尺度可能對模型估計的結果有根本性的影響，選擇最適合的測量尺度是唯一解決辦法。唯有確定所使用的尺度是最合適的，才能確保 t 值的估計沒有因為研究設計而產生的偏誤。為了避免上述的問題影響本研究的估計結果，本研究在尺度選擇上有特別留意。歷年考績與 2012 考績兩個變數是分數，本身為連續變項，沒有尺度選擇的問題。自評績效的測量由三個題項組成：我的工作品質能獲得主管的肯定、我的工作品質能獲得同仁的肯定、我的整體工作表現十分優異。

資料模型中的自變數 A 的迴歸係數所對應的 t 值為 3.69，而自變數 A 在自評資料模型中的迴歸係數所對應的 t 值為 0.94，兩個 t 值相減後的絕對值為 2.75（|3.69-0.94|=2.75）。由於 0.05 的信賴水準所對應的 t 值大約為 2（更精確來說是 1.96），只要是兩者 t 值的絕對值差超過 2，即可判定兩個模型在自變數 A 的迴歸係數估計上有顯著差異。

顯著的迴歸係數估計誤差可以分為兩種。一種為假性的正相關，表示自評資料庫估計出來的自變數與依變數有正向的顯著關係，然而典藏資料庫的估計結果卻顯示這層關係不存在。另一種則為假性負相關，表示自評資料庫估計自變數與依變數關係不顯著，然而典藏資料庫卻顯示顯著的負向關係。依此原則，本研究可以統計自評與典藏資料庫在迴歸係數的估計上產生了多少假性正相關與假性負相關，並統計整個模型估計錯誤的比例。

採用與 Meier 和 O'Toole（2013）相同的分析方法，除了在方法上呼應之外，更重要的是藉由方法的控制，比較同源誤差的影響對於研究學校教師與一般公務人員的績效是否有差別；更能呼應 George 和 Pandey（2017）的建議，增加同源誤差研究的多樣性。

肆、研究發現

一、自評考績與典藏考績分數比較

三種績效評估方式的相關分析呈現如表 1（變數的描述性統計請見附錄二）。歷年考績分數與 2012 典藏考績的相關係數為 0.48，為中度相關。而 2012 自評績效則是與歷年考績以及 2012 典藏考績的相關係數皆只有 0.1，為低度相關（見表 1）。換句話說，若我們同意歷年考績分數是最能表現員工真實績效的參考值，2012 典藏考績的測量是較接近歷年考績分數的測量結果的。

三個題項的測量方式皆為李克特式量表六點尺度；三個題項的分數加總後，為公務人員自我評估的績效。之所以採用加總而非因素分析方式產生自評績效，是因為這三題所代表的績效分別為代表從主管、同仁與自我的眼光所判斷的績效，是三個不同的構面。本研究亦嘗試透過因素分析估計因素分數（factor scores），並用因素分數當作依變數進行迴歸分析，其結果與直接加總的結果 t 值無顯著差異，與另兩組模型的 t 值比較結果更是完全一致。由此判斷，自評績效這個變數的測量尺度在本研究中不是造成 t 值變化的原因。

表 1

歷年考績、2012 典藏考績、2012 自評績效相關分析

	歷年考績	2012 典藏考績	2012 自評績效
歷年考績	1		
2012 典藏考績	0.48**	1	
2012 自評績效	0.10**	0.11**	1

資料來源：本研究

二、自評考績與典藏考績分數對模型估計的影響

同時使用典藏分數與自評分數進行模型估計，是評估同源誤差是否存在的檢測方法之一（Meier & O'Toole，2013）。本研究以歷年累積的甲等比例當作典藏分數，並以此為比較基準，評估 2012 的自評績效分數在模型估計中是否與典藏分數有所差距。2012 的典藏考績分數則做為另一項參考值，避免歷年資料無法與單一年度資料類比的疑慮。由於三個模型的依變數的單位不同（歷年考績甲等比例為百分比例、2012 典藏考績為考績分數、2012 自評績效分數為三題李客特式量表答題總和），因此表 2 的迴歸係數大小無法直接進行比較。然而，迴歸係數的方向性與是否能夠拒絕迴歸係數等於零的虛無假設，則依然具有可比性。迴歸分析結果請見表 2。

表 2 三個迴歸模型的分析結果，很明顯的呈現出些許的落差。本研究以「歷年考績甲等比例」做為基準，每一個變數的結果可能為「顯著正」、「顯著負」、「不相關」；依此比對自評資料的估計結果，共有 3 種結果配對：典藏資料為「負、不」而自評資料為「正」，此為「假性正相關」；典藏資料為「正、不」而自評資料為「負」，此為「假性負相關」；最後一種為典藏資料為「正、負」，而自評資料為「不」，此為「假性無相關」。舉例而言，與歷年考績甲等比例模型相比，2012 自評績效分數在「性別」這個變數的估計上即呈現了假性的負相關。亦即性別這個自變數在 2012 自評績效分數模型中，與依變數呈現顯著負相關，但是在歷年考績甲等比例模型中卻是顯著正相關；假定歷年考績甲等比例才是與真實的績效接近的參考值，則 2012 自評績效分數模型所估計出來的迴歸係數含有偏誤。同樣的問題也出現在 2012 自評績效分數模型中，「機關轉換」這個自變數的迴歸係

數估計。「機關轉換」在 2012 自評績效分數模型中與依變數的關係方向為顯著正相關，在歷年考績甲等比例模型中卻為顯著負相關，因此在 2012 自評績效分數模型中為假性正相關。2012 自評績效分數模型另在「性別」這個變數上出現了「假性無相關」的估計；亦即典藏模型在性別變數中為顯著正相關，但在自評績效模型則為無顯著相關的結果 > 經統計，2012 自評績效分數模型與歷年考績甲等比例模型相比，共計有 2 個假性正相關、4 個假性負相關，以及 3 個假性無相關。在 11 個自變數中有 9 個迴歸係數產生估計偏誤，錯估率為 82%（見表 3）。[8]

歷年考績甲等比例模型與 2012 典藏考績分數模型則沒有這麼大的估計差異。同樣是典藏考績，2012 單一年度的模型估計只有「升遷政治性」產生了假性負相關，另有 3 處為假性無相關，錯估率僅佔了 4 / 11，約 36%。[9] 換言之，2012 典藏考績這個變數與歷年考績的測量結果是比較接近的，符合本研究在一開始的前提假定：「歷年績效應該是最接近真實績效的指標」，且肯定兩個典藏資料的績效評估結果是比較接近的，增加以歷年績效做為衡量指標的可信程度，也提升了典藏資料做為客觀績效指標的可信程度。

[8] 本文錯估率的計算可能無法與 Meier & O'Toole（2013）的結果進行比較。原因在於，Meier & O'Toole（2013）並未統計「假性無相關」。若本文扣掉「假性無相關」的結果估計，錯估率則為 6/11=55%。

[9] 若依照 Meier & O'Toole（2013）不計算「假性無相關」，則錯估率為 1/11=9%。

表 2

迴歸結果比較分析

	歷年考績甲等比例			2012 典藏考績分數			2012 自評績效分數		
	β	t		β	t		β	t	
人口變項									
性別（男）	2.13	2.66	***	-0.14	-0.98		-0.22	-2.18	**
已婚	2.62	2.12	**	0.69	3.36	***	0.17	1.15	
1993（初任）年齡									
能力									
機關轉換	-13.14	-3.72	***	-0.23	-0.38		0.86	2.01	**
教育程度	1.50	3.02	***	0.30	3.47	***	0.09	1.45	
政治技能	0.05	0.74		0.02	1.43		0.09	9.25	***
動機									
公共服務動機	0.14	2.35	**	0.01	1.1		0.07	9.63	***
家庭犧牲	0.25	2.61	***	0.05	3.22	***	-0.05	-3.56	***
升遷滿意度	0.73	6.42	***	0.12	5.97	***	-0.08	-5.24	***
機會									
升遷政治性	-0.09	-0.95		-0.04	-2.31	**	-0.03	-2	**
中央年資比	-0.03	-3.28	***	0.00	1.07		0.00	1.25	
_cons	48.34	6.61	***	78.09	63.48	***	8.59	9.38	***
N	1469			1396			1471		
F	13.42	***		11.29	***		28.3	***	
R2	0.09			0.08			0.21		
假性正相關								2	
假性負相關					1			4	
假性無相關					3			3	

***p<0.01; **p<0.05

資料來源：本研究

表 3

與歷年典藏資料相比的估計誤差

	2012 典藏	2012 自評
T 值誤差平均	1.33	2.48
T 值誤差標準差	0.86	2.72
T 值誤差最大值	3.34	8.51
T 值誤差最小值	0.45	0.48
假性正相關		2
假性負相關	1	4
假性無相關	3	3
錯估率	36%	82%

資料來源：本研究

迴歸係數的 t 值的差異可以判斷這些錯估的係數是否有達到統計上的顯著水準，結果呈現於表 4。表 4 中的各項數值為迴歸係數 t 值的差的絕對值。舉例而言，性別在歷年考績甲等比例模型中的 t 值為 2.66，在 2012 典藏考績分數模型中的 t 值為 -0.98，則兩者差的絕對值為 |2.66-（-0.98）|=3.64。由於 t 值達到 95% 的顯著水準的值大約為 2（更精確為 1.96），t 值差的絕對值若大於 2 則可以判定兩個模型在同一個變數的迴歸係數的估計上有顯著落差。本研究發現，2012 自評考績分數模型與歷年考績甲等比例模型共有 3 個自變數的迴歸係數差達到統計上的顯著水準；2012 自評考績分數模型則計有 7 個自變數的迴歸係數差達到統計上的顯著水準。2012 自評績效、典藏考績兩個模型也有 8 個自變數達到顯著水準。顯見 2012 自評資料無論是與當年度的典藏資料相比，或是與歷年累積的典藏資料相比，都有超過 50％以上變數的 t 值估計具有顯著誤差；同樣是典藏資料，歷年累積的考績與 2012 單一年度的考績則只有 3 個自變數的 t 值估計差顯著。表 4 另外統計了 2012 自評與典藏模型和歷年考績甲等比例模型的 t 值差。可以發現 2012 自評考績模型無論是在 t 值的平均數、標準差、最大值、最小值上，都比 2012 典藏考績模型的數值來得高。可見自評資料在模型估計的誤差確實比典藏模型大，也證明了同源誤差確實不容忽視。

表 4

自評資料與典藏資料的 t 值差異

	歷年～ 2012 典	歷年～ 2012 自	2012 典～ 2012 自
性別（男）	3.64	4.84	1.2
已婚	1.24	0.97	2.21
1993 年齡	1.36	1.58	0.22
機關轉換	3.34	5.73	2.39
教育程度	0.45	1.57	2.02
政治技能	0.69	8.51	7.82
公共服務動機	1.25	7.28	8.53
家庭犧牲	0.61	6.17	6.78
升遷滿意度	0.45	11.66	11.21
升遷政治性	1.36	1.05	0.31
中央地方年資比	2.21	4.53	2.32
顯著估計錯誤	3 ／ 11	7 ／ 11	8 ／ 11

資料來源：本研究

三、延伸討論：典藏資料來源與模型估計的解釋力問題

從研究結果來看，另一個可以討論的問題是典藏資料對於模型估計的解釋力問題。從表 3 的結果來看，典藏資料模型的 R2 只有 0.09，表示這個模型只能解釋 9% 的變異。R2 不高的原因可能有二。其一，這個模型本身的解釋力不夠好。其二，依變數（績效）的測量效度不夠。過去的研究使用 Blumberg & Pringle（1982）的模型，其 R2 多能達到 0.2 至 0.3 的程度（c.f. Boselie, 2010; Knies & Leisink, 2014; Reinholt, M., Pedersen T., Foss, N. J., 2011），因此模型解釋力不夠好的這個原因應該可以排除，依變數（績效）的測量效度不夠的可能性較高。張鎧如等（2015）使用相同的變數與模型亦使用「考績甲等比例」做為依變數，獲得的 R2 為 0.11，與本研究的 0.09 相去不遠（請見表 5）。然而，張鎧如等（2015）所使用的另一

個績效指標「2012 職等」做為依變數，R2 上升至 0.34。顯見縱使「考績甲等比例」與「2012 職等」皆可視為典藏資料，對模型估計的影響效果卻有很大的差異。這個差異也凸顯了使用典藏資料變數的選擇與測量問題所帶來的模型估計問題，可能不比使用自評資料來得小。

表 5

本研究與張鎧如等（2015）模型結果比較

	本研究整體模型檢定								張鎧如等（2015）整體模型檢定	
	歷年考績甲等比例		2012 典藏考績分數		2012 自評績效分數		考績甲等比例[10]		2012 職等	
機關轉換	−				+		−		+	
教育程度	+		+				+		+	
政治技能					+					
公共服務動機	+		+		+		+			
家庭犧牲	+		+		−		+		+	
升遷滿意度	+		+		−		+		+	
升遷政治性			−							
中央地方年資比	−		−				−		+	
N	1469		1396		1471		1420		1420	
F	13.42	***	11.29	***	28.3	***	16.97	***	76.52	***
R^2	0.09		0.08		0.21		0.11		0.34	

資料來源：本研究整理、張鎧如等（2015）

[10] 張鎧如等（2015）使用「考績甲等比例」詞，與本研究使用「歷年考績甲等比例」稍有不同，但變數計算的方式相同。

典藏資料的變數測量問題以及變數的選擇問題，將成為研究者未來使用典藏資料最關鍵的挑戰。以張鎧如等（2015）為例，該研究將「考績甲等比例」與「2012職等」同樣定義為績效指標，並分別做為兩個迴歸模型的依變數；這兩個變項皆非受訪者主觀評估，而是客觀存在的事實，因此理論上該研究設計已經避開了同源誤差的問題。然而，兩個績效變數所估計的兩個迴歸模型依然出現了迴歸係數方向不一致的情形：機關轉換次數在「考績甲等比例」模型中為負，在「2012職等」模型中為正；中央地方年資比在「考績甲等比例」模型中為負，在「2012職等」模型中為正（見表5）。

張鎧如等（2015）將這樣的差異歸因於「考績甲等比例」與「2012職等」本質上的差異，因此兩個模型估計雖然在迴歸係數方向與R2皆不相同，卻不難理解背後的原因。然而，這樣的解釋卻恰好顯示，無論資料來源是典藏或自評，問題的關鍵依然是研究者對變數的定義以及研究問題的旨趣。

張鎧如等（2015）一開始將「考績甲等比例」與「2012職等」同樣定義為公務人員績效的典藏評估指標，卻在研究結果呈現出現差異之後同意這兩個績效指標內涵不同，這樣的解讀方式其實與使用自我評估的績效指標並無差異。以本研究表3為例，同樣出現機關轉換次數在典藏模型中為負、在自評模型中為正的情況，研究者大可將這樣的差異解釋為「當機關轉換次數較少，受訪者容易因為對環境、工作熟悉而增加對自身工作績效的認同」，相反的「當機關轉換次數多，依照新進人員容易拿乙等的潛規則，考績容易較低」。若「公務人員績效」這個內涵可以有數種定義，使用不同的定義進行模型估計自然會出現不同統計分析結果，對於模型的解釋也會不同。若將這樣差異全部歸因於「相同資料來源所造成的估計誤差」，並否定這個差異的理論內涵，研究者很難針對「不同來源的典藏資料所造成的模型估計差異」自圓其說。因此比堅持選擇典藏資料進行研究更重要的，是研究者是否能給予重要的研究變數一個不模糊的定義，並且依照這個定義選擇最合適的資料測量方式或資料來源。

伍、結論與建議：未來該怎麼做？

一、主要研究結論

（一）同源誤差確實與典藏資料的估計結果有顯著誤差

同源誤差可以說是使用資料庫進行研究最大的潛在問題。本研究也發現自評資料庫所測量的考績分數無論是在分數本身，或是使用迴歸模型進行分析，皆與典藏資料庫的統計結果有顯著誤差，且錯估率皆超過 50%。這個結果甚至比 Meier & O'Toole（2013）研究的錯估率更高。這樣的誤差讓我們不得不正視同源誤差的存在，同時也必須探討未來如何使用資料庫進行量化研究的問題。

（二）資料來源與研究模型的搭配相當關鍵

除了自評資料的錯估率問題，本研究也進一步發現，僅僅是堅持使用典藏資料庫進行研究依然無法解決模型估計誤差的問題。是否能依照研究問題給予變數明確、不含糊的定義，才是決定何種測量方式、何種資料來源最合適用於模型估計的關鍵。而這也意味著同源誤差的問題並不是什麼太過高深、需要高級分析方法才能排除的技術問題，而是最根本的定義與研究設計問題。只是過去研究者在做績效測量變數的選擇的時候，容易認為不同的績效相關變數是可以相互替代的，或是認為當客觀變數不存在的時候，可以逕行以主觀變數代之。然而本研究的發現顯示，不但主觀變數與客觀變數在模型估計的係數不同，所謂的客觀變數之間也會有不同的估計係數以及 R2，可見未來必須用更嚴謹的態度對待變數選擇以及替代的問題。這也是本研究為什麼必須強調，解決同源誤差的方法其實是回到最根本的定義與研究設計。

（三）同源誤差同時包含「測量誤差」與「概念建構差異」

本文並非僅將同源誤差視為一項測量技術問題。既有公共行政文獻多將同源誤差理解為資料來源一致所導致之統計偏誤，並主張透過研究設計避免或以多來源資料加以修正。然而，本文主張，在以「績效」為依變數的研究脈絡下，不同資料來源之差異，未必僅反映測量誤差，亦可能反映績效概念本身的制度建構差異。同源誤差問題因此不僅涉及估計偏誤，更涉及研究者如何界定與操作化績效概念。

此一觀點亦意味著，單純以多來源資料取代自評資料，未必能完全解決模型估計偏誤，反而可能引入不同績效概念之間的不一致性，進而影響研究結論的解釋。基於此，本文將同源誤差重新定位為一個同時包含「測量誤差」與「概念建構」的分析問題，並據此重新詮釋不同資料來源所產生之模型估計差異。

從此觀點出發，未來研究在處理同源誤差問題時，除了關注測量技術與統計修正之外，亦應當檢視概念建構是否因為測量問題而被扭曲。

（四）不能以同源誤差通盤否定研究價值

除了發現同源誤差確實可能影響研究結論之外，本文另外想強調的是：並非只要有疑似同源誤差的問題，就應該全盤否定該研究的價值。首先，本研究所呈現之錯估比例（如自評模型錯估率達 82%），並不意味著所有使用自評資料之研究皆存在相同程度之估計偏誤。此一結果主要出現在特定研究設計條件下：其一，自變數與依變數同時來自同一份問卷資料；其二，依變數為具評價性質之自我績效判斷；其三，自變數包含態度、動機或認知類型變數，可能與依變數在心理層面產生一致性回應傾向。在此情境下，同源誤差較可能透過回應風格或角色期待機制，系統性影響變數間之相關結構。

再者，若研究關心之本即為個人主觀感受（例如自我效能、滿意度或主觀認知），且理論上預期變數間存在心理層面的內在連結，則自評資料未必構成方法論問題。此外，若研究設計採用時間區隔、來源區隔或多方法測量等方式，自評資料所產生之偏誤程度亦可能降低。因此，本文之發現較適用於依變數屬於評價性績效指標，且研究設計未區隔資料來源之情境，而非對所有自評研究之一般性否定。

二、研究建議

基於上述發現，本研究提出幾個從研究方法上減少同源誤差影響的建議，同時，也再次針對同源誤差問題，提出學術界實務上應該如何理解同源誤差的建議。

（一）從方法上減少同源誤差的三階段建議

本研究認為未來的研究可從三個階段減少同源誤差的影響，並且提升模型解

釋力。

第一階段：釐清研究問題、清楚變數定義。許多名詞在社會科學研究中的定義有多重面向。舉例而言，「績效」這個概念就分為個人真實的績效、個人自我評估的績效以及從他人角度認知的績效。由於真實的績效相當難測量，過去的研究經常使用自我評估的績效做為替代，也因此產生同源誤差的問題。然而從他人角度認知的績效也不是一個單一的面向，例如考績成績和升遷都是他人角度認知績效的衍生，理論上兩者雖然有正相關，但若說兩者可以相互替代則太過武斷。因此研究者應當在研究進行之前就先釐清所使用的績效概念，並清楚定義。若需要能接近真實績效的測量變數，必須敘明所選擇的變數與真實績效的差距，以及因為這個差距可能造成的偏誤。若研究問題關心的並非真實的績效，而是自我評估的績效，亦不能說該研究存在同源誤差的問題。

第二階段：研究設計。延續上例，若真實的績效為研究所關心的議題，研究者可透過研究設計以及資料庫的選擇，控制、縮減誤差的產生。最好的方法當然是避免自變數與依變數的來源相同，然而因為社會科學、行為科學資料性質的限制，這個選擇並非永遠可行。如果無法避免使用相同資料來源，可以使用不同的測量工具盡可能的減少變數間的連結（Podsakoff et al., 2003）。例如在資料收集的時候創造一些時間的空檔，或是在回答某些問題之前閱讀些情境，盡量讓答題者不會認為某些題項之間具有關連。此外，研究者也可使用不同的測量方式，例如交叉使用李克特式量表（Likert scales）、開放式問題（open-ended questions）、臉譜量表（faces scales）等，或是透過不同的媒介收集資訊（例如網路問卷、紙本問卷、面對面訪談）、更換訪談地點等等（Podsakoff et al., 2003）。除了答題方式與情境的變化，對答題者強調研究的匿名性以及答案沒有對錯，也可以降低同源誤差的產生。另外，研究者亦需避免使用模糊的概念、語句，盡量將問題簡化、使用易懂的語言、避免使用專有名詞，並特別針對容易混淆、或不清楚的概念加以說明。這些方式都有助於答題者依照他們最真實的感受進行答題，減少同源誤差（Podsakoff et al., 2003）。

第三階段：使用統計方法做事後檢驗。雖然透過研究設計、問卷設計能夠避免或減少同源誤差的產生，然而這些方式卻不利於必須使用二手資料的研究者，此時就需要透過統計方法進行檢驗或修正。以下介紹幾種常用的檢驗或修正方法。

1、哈門氏單因子測試法（Harman's one-factor test）

這個方法的原理是，如果透過因素分析，未轉軸的第一個主成分同時解釋了自變數和依變數的主要變異量，表示這個主成分是一個綜合因子（general factor）；綜合因子的出現表示變數間存在同源誤差（彭台光、高月慈、林鉦棼，2006；Podsakoff et al., 2003）。研究者可以將與這個因子相關的題項刪除（partial out），排除同源誤差的影響。然而這個做法可能有相當嚴重的後遺症。被排除的題項除了被同源誤差影響之外，也可能含有理論上與實務上有意義的變異；一旦完全排除，可能從原本由同源誤差所造成的型一錯誤，掉入型二錯誤的陷阱（彭台光等，2006）

2、潛在因素控制（Controlling for the effects of a directly measured latent methods factor）

這個方法係透過控制一些常見的造成同源誤差的變項，例如社會認同（social desirability）、正面情緒（positive affectivity）或負面情緒（negative affectivity）等，來排除潛在的同源誤差問題。這個方法需要研究者在問卷中加入測量社會認同、正面情緒或負面情緒的變數，並分析這些變數與研究的依變數、自變數的關係，然後排除自變數與依變數受到這些潛在誤差變數的影響。然而這麼做的最大問題是，研究者必須確定這些變數是該研究最主要的同源誤差來源，否則可能表面上控制了同源誤差，實際上卻沒有真正的排除。不過在大部分的情況，社會認同、正面情緒或負面情緒已經被認定為主導同源誤差的主要變數（Brief et al., 1988; Burke et al., 1993; Chen & Spector, 1991; Jex & Spector, 1996），因此若真的無法避免，控制這些變數的影響不失唯一個可行且有效的辦法（彭台光等，2006）。

3、多特質多方法模式分析（Multitrait-Multimethod, MTMM）、驗證因素分析（confirmatory factor analysis, CFA）與結構方程式模型（Structural Equation Modeling, SEM）

這三個方法經常在檢驗同源誤差時經常合併使用。多特質多方法模式分析是一組分析架構，其方式是製作多特質多方法相關係數矩陣，使研究者能夠分析不同變數相同來源、相同變數不同來源的相關係數，藉此瞭解變數的信度、效度、以及因為測量方法而產生的誤差。而這個方法可以與 CFA 共同使用，分析探討變數來自於特質（traits）、方法（methods）、以及隨機（random）上的誤差。甚至進一步搭配 SEM 模型，直接探討變數之間的因果關係。透過這一系列的方法的

優點在於特質（traits）、方法（methods）、以及隨機（random）的變異皆可以被計算出來，研究者若要控制同源誤差，只要在後續的分析中排除來自於方法的誤差即可。這個方法是目前學界最普遍使用的方法。

（二）對公共行政研究社群該如何看待同源誤差的實務建議

1、區分變數來源且清楚說明

本研究結果顯示，當依變數屬於評價性自我判斷，且自變數同樣來自相同資料來源時，模型估計可能出現系統性偏誤。因此，未來公共行政實證研究在研究設計階段，應更加明確區分變數來源與測量性質，並在可行範圍內採取來源區隔或時間區隔策略。此外，研究者亦應在方法章節中清楚說明資料來源一致可能產生之風險，而非僅於事後進行統計修正。

2、研究資料選擇應重視與理論建構的對應

本研究亦顯示，典藏資料雖可降低同源誤差風險，但並非自然等同於「較佳資料」。不同典藏指標之間仍可能反映不同制度邏輯與評量內涵。因此，未來研究在選擇績效或結果變數時，應更加重視變數與理論之對應關係，而非單純依據資料取得便利性。資料選擇不僅是技術決定，更是理論判斷。

3、同源誤差問題並非研究品質的關鍵判斷

在學術社群層面，本文認為，對同源誤差問題的重視應建立在審慎判斷之上，而非採取一概否定自評資料之立場。若僅以資料來源是否相同作為研究品質的判準，可能忽略不同研究問題與理論構面性質之差異。未來在審稿與研究評估過程中，應更強調研究者是否清楚界定變數內涵、說明資料來源限制，以及是否提出合理的研究設計或敏感度分析，而非僅依資料形式進行價值判斷。如此方能在維持方法論嚴謹性的同時，避免對公共行政研究議題產生不必要的限制。因此，公共行政期刊在審稿指引中，建議可鼓勵作者揭露資料來源潛在偏誤與處理方式，而非將資料來源一致視為絕對拒稿條件。

三、研究限制

本研究有兩個研究限制必須提及：

（一）本研究假定「典藏資料為相對客觀的」績效變數，若這個假設前提不存在，那麼本研究的結論也會受到影響。然而究竟哪種績效測量方式最能接近真實客觀的績效，目前學界依然沒有肯定的答案。因此本研究只能假設，在一個長期制度化下的績效評估機制，以及這個機制所累積下來的歷年評量結果，能夠帶來相對可信的績效訊號。然而此一假設仍受制度設計與組織文化條件限制。例如，若考績制度具有高度比例管制、輪流分配或人際協調機制，則長期累積未必能完全消除制度偏誤。此外，不同機關之評量文化可能存在差異，亦可能影響甲等比例之分布。若制度存在重大改革或評量標準明顯轉變，則長期累積假設亦可能不成立。因此，本文之操作化方式較適用於制度運作相對穩定、評量邏輯未劇烈改變之情境，未來研究若處於制度快速變動時期，應審慎評估替代指標之適用性。

（二）本研究旨在檢驗同源誤差問題是否存在，以及問題的嚴重程度，並依照研究發現給予未來相關研究建議。不同統計技術對同源誤差的排除效果則不在本研究的討論範圍之內，因此無法針對統計方法的使用做進一步的建議。

參考書目

吳志文、葉光輝（2011）。雙元自主性的共存與範疇優勢性運作機制：以訊息區辨表現降低共同方法變異。**中華心理學刊，53**（1），59-77。[Wu, C.-W., & Yeh, K.-H. (2011). The Coexistence and Domain Superiority Processing of Dual Autonomy: Decreasing Common Method Variance with Message Discernment. *Chinese Journal of Psychology, 53(*1), 59~77.]

呂宛蓁、蕭嘉惠、許振明、曹校章、王學中（2012）。臺灣體育運動學術研究的共同方法變異。**大專體育學刊，14**（4），419~427。[Lu, W.-C., Hsiao, C.-H., Hsu, C.-M., Tsao, H.-C., & Wang, H.-C. (2012). The Common Method Variance of Physical Education and Sport Research in Taiwan. *Sports and Exercise Research, 14*(4), 419~427.]

張鎧如、陳敦源、簡鈺珒、李仲彬（2015）。探索台灣公務人員工作績效之影響因素：能力、動機與機會的整合觀點。**東吳政治學報，33**（4），1~71。[Chang, K.-J., Chen, D.-Y., Chien, Y.-C., & Lee, C.-P. (2015). Exploring Work Performance of Public Employees in Taiwan: A Compound Perspective of Capacity, Willingness, and Opportunity. *Soochow Journal of Political Science, 33*(4), 1~71.]

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**。考試院委託研究報告。[Chen, D.-Y. (2010). A review of Taiwan's performance appraisal system: Evidence-based theory, justice perspective,and participatory management approach. Examination Yuan.]

陳敦源（2013）。**公部門陞遷制度之實證研究：以正義觀點為核心的檢驗**。考試院委託研究報告。[Chen, D.-Y. (2013). An empirical study of promotion systems in the public sector: A justice-based examination. Examination Yuan.]

彭台光、高月慈、林鉦棽（2006）。管理研究中的共同方法變異：問題本質、影響、測試和補救。**管理學報，23**（1），77~98。[Peng, T.-K., Kao, Y.-T., & Lin, C.-C. (2006). Common Method Variance in Management Research: Its Nature, Effects, Detection, and Remedies. *Journal of Management and Business Research, 23*(1), 77~98.]

蕭佳純、涂志賢（2012）。教師創意教學衡量中共同方法變異問題之探討。**測驗學刊，59**（4），609~639。[Hsiao, C.-C., & Tu, C.-H. (2012). Common Method

Variance in the Measurement of Teachers' Creative Teaching. *Psychological Testing, 59*(4), 609~639.]

Andersen, L. B., Heinesen, E., & Pedersen, L. H. (2016). "Individual Performance: From Common Source Bias to Institutionalized Assessment." *Journal of Public Administration Research and Theory, 26*(1), 63~78.

Appelbaum, E., Bailey, T., Berg, P., & Kalleberg, A. L. (2000). Manufacturing Advantage: Why High-Performance Work Systems Pay Off. Ithaca, NY: Cornell University Press.

Bachman, J. G., & O'Malley, P. M. (1984). "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly, 48*(2), 491~509.

Bayo-Moriones, A., & Galdon-Sanchez, J. E. (2010). "Multinational Companies and High-Performance Work Practices in the Spanish Manufacturing Industry." The International *Journal of Human Resource Management, 21*(8), 1248~1271.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review, 63*(5), 586~606.

Blumberg, M., & Pringle, C. D. (1982). "The Missing Opportunity in Organizational Research: Some Implications for a Theory of Work Performance." *Academy of Management Review,7*(4), 560~569.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance. *Personnel Psychology, 48*(3), 587~605.

Boselie, P. (2010). "High Performance Work Practices in the Health Care Sector: A Dutch Case Study." *International Journal of Manpower, 31*(1), 42~58.

Bouckaert, G., & Halligan, J. (2008). Managing performance: International comparisons. Routledge.

Boxall, P., & Purcell, J. (2003). Strategy and Human Resource Management. New York: Palgrave Macmillan.

Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. (2010). "What Is Method Variance and How Can We Cope With It? A Panel Discussion." *Organizational Research Methods, 13*(3), 407~420.

Brief, A. P., Burke, M. J., George, J. M., Robinson, B. S., & Webster, J. (1988). "Should negative affectivity remain an unmeasured variable in the study of job stress?" *Journal of Applied Psychology,* (73), 191~198.

Burke, M. J., Brief, A. P., & George, J. M. (1993). "The role of negative affectivity in understanding relations between self-reports of stressors and strains: A comment on the applied psychology literature." *Journal of Applied Psychology,* (78), 402~412.

Chen, P. Y., & Spector, P. E. (1991). "Negative affectivity as the underlying cause of correlations between stressors and strains." *Journal of Applied Psychology,* (76), 398~407.

Cote, J. A., & Buckley, M. R. (1987). "Estimating Trait, Method, and Error Variance: Generalizing across 70 Construct Validation Studies." *Journal of Marketing Research,* (3), 315.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349~354.

Doty, D. H., & Glick, W. H. (1998). "Common Methods Bias: Does Common Methods Variance Really Bias Results?" *Organizational Research Methods, 1*(4), 374~406.

Favero, N., & Bullock, J. B. (2015). "How (Not) to Solve the Problem: An Evaluation of Scholarly Responses to Common Source Bias." *Journal of Public Administration Research and Theory, 25*(1), 285~308.

George, B., & Pandey, S. K. (2017). "We Know the Yin-But Where Is the Yang? Toward a Balanced Approach on Common Source Bias in Public Administration Scholarship." *Review of Public Personnel Administration, 37*(2), 245~270.

Jakobsen, M., & Jensen, R. (2015). "Common Method Bias in Public Management Studies." *International Public Management Journal, 18*(1), 3~30.

Jex, S. M., & Spector, P. E. (1996). "The Impact of Negative Affectivity on Stressor Strain Relations: A Replication and Extension." *Work and Stress,* 10, 36~45.

Keeping, L. M., & Levy, P. E. (2000). "Performance Appraisal Reactions: Measurement, Modeling, and Method Bias." *Journal of Applied Psychology, 85*(5), 708~723.

Kelman, S. 2015. "Letter from the Editor." *International Public Management Journal, 18*(1), 1~2.

Knies, E., & Leisink, P. (2014). "Linking People Management and Extra-Role Behaviour: Results of a Longitudinal Study." *Human Resource Management Journal, 24*(1), 57~76.

Lindell, M. K., & Whitney, D. J. (2001). "Accounting for Common Method Variance in Cross-sectional Research Designs." *Journal of Applied Psychology, 86*(1), 114~121.

Meade, A. W., Watson, A. M., & Kroustalis, C. M. (2007). "Assessing Common Methods Bias in Organizational Research." Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology, April 27~29, New York.

Meier, K. J., & O'Toole, J. L. J. (2013). "Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships." *Journal of Public Administration Research and Theory, 23*(2), 429~456.

Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Georgetown University Press.

Pace, V. L. (2010). "Method Variance From the Perspectives of Reviewers: Poorly Understood Problem or Overemphasized Complaint?" *Organizational Research Methods, 13*(3), 421~434.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun et al. (Eds.), The role of constructs in psychological and educational measurement (pp. 49~69). Lawrence Erlbaum.

Podsakoff, P. M. and D. W. Organ. (1986). "Self-Reports in Organizational Research: Problems and Prospects." *Journal of Management, 12*(4), 531~544.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology, 88*(5), 879~903.

Pollitt, C. (2013). The logics of performance management. Evaluation, 19(4), 346~363.

Reinholt, M., Pedersen T., Foss, N. J. (2011). "Why a Central Network Position Isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks." *Academy of Management Journal, 54*(6), 1277~1297

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). "A Tale of Three

Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance." *Organizational Research Methods, 12*(4), 762~800.

Spector, P. E. (1987). "Method Variance as an Artifact in Self-Reported Affect and Perceptions at Work: Myth or Significant Problem?" *Journal of Applied Psychology, 72*(3), 438~443.

Spector, P. E. (2006). "Method Variance in Organizational Research: Truth or Urban Legend?" *Organizational Research Methods, 9*(2), 221~232.

Talbot, C. (2010). *Theories of performance: Organizational and service improvement in the public domain.* Oxford University Press.

Van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance management in the public sector* (2nd ed.). Routledge.

Williams, L. J., Buckley, M. R., & Cote, J. A. (1989). "Lack of Method Variance in Self-Reported Affect and Perceptions at Work: Reality or Artifact?" *Journal of Applied Psychology, 74*(3), 462.

Williams, L. J., Gavin, M. B., & Williams, M. L. (1996). "Measurement and Nonmeasurement Processes With Negative Affectivity and Employee Attitudes." *Journal of Applied Psychology, 81*(1), 88~101.

附錄一：變數操作化方式

變數名稱	操作化定義	編　　碼	Cronbach's Alpha 值	資料來源／屬性
歷累積甲等比例	個人從 1993 年至 2012 年生涯考績甲等次數／考績總次數	甲等比例（0~100 的百分比）	N／A	人事資料庫／客觀
2012 典藏考績分數	2012 年人事資料庫考績分數	考績分數	N／A	人事資料庫／客觀
2012 自評績效分數	61. 我的工作品質能獲得主管的肯定 62. 我的工作品質能獲得同仁的肯定 63. 我的整體工作表現十分優異	Likert 6 點尺度，3 題分數加總	.843	問卷／主觀
機關轉換	進入公職後之機關總轉換次數除以年資	進入公職後之機關總轉換次數除以年資	N／A	人事資料庫／客觀
教育程度	最高教育程度	高中職＝ 1； 專科＝ 2； 大學＝ 3； 碩士＝ 4； 博士＝ 5	N／A	問卷／主觀
政治技能	72. 我很瞭解人性，沒有什麼人可以騙得了我。 73. 在社交場合中，我總是知道如何透過言行舉止來影響他人。 74. 與別人互動時，我懂得察言觀色。 75. 我能夠讓我身旁多數人對我沒有戒心。	Likert 6 點尺度，9 題分數加總	.832	問卷／主觀

變數名稱	操作化定義	編　　碼	Cronbach's Alpha 值	資料來源／屬性
	76. 我與他人的溝通協調技巧很好。 77. 對我而言，與他人發展友好關係是輕而易舉的事。 78. 我在工作上花大量的時間與精力與他人建立網絡關係。 79. 我在工作上善於和有影響力的人建立關係。 80. 我善於使用我的人脈關係來處理工作上面臨的任何事情。			
公　共服務動機	91.「政治」對我來說是一個骯髒的詞彙。(R) 92. 我不關心政治人物的一舉一動。(R) 93. 政策過程中的意見交換與相互妥協對我來說是沒意義的。(R) 94. 我對我所服務的社區與人民不感興趣。(R) 95. 我對我所服務的社區與人民作了無私的奉獻。 96. 執行公共服務對我來說是一種公民責任。 97. 當我看到弱勢的人受害時，總是感到於心不忍。	Likert 6 點尺度，14 題分數加總	.820	問卷／主觀

變數名稱	操作化定義	編　　碼	Cronbach's Alpha 值	資料來源／屬性
	98. 我經常提醒自己，人們總是互相依賴的。 99. 許多社會福利與社會救助計畫是不可或缺的。 100. 對人民、社會作出貢獻，比我個人的生涯發展更有意義。 101. 只要我認為對大眾有利，我可以冒著失去陞遷機會的風險去爭取。 102. 為社會奉獻本來就是我應該做的。 103. 我認為當今人們該學、該做的，不是「取之於社會」而是「用之於社會」。 104. 只要政府官員作出對人民有益的事，即使傷害了我個人的利益，我也能夠接受。			
家庭犧牲	68. 我的工作使得我沒有時間陪家人。 69. 工作上所花費的時間使我不能分擔家庭的責任與事務。 70. 下班後，我常有被榨乾的感覺，而無法為家庭盡心盡力。 71. 我在工作上所用來解決問題的方法，	Likert 6 點尺度，4 題分數加總	.868	問卷／主觀

變數名稱	操作化定義	編　　碼	Cronbach's Alpha 值	資料來源／屬性
	無法有效解決家庭生活中所遇到的問題。			
升遷滿意度	23. 與資歷相當的同事相較，我相當滿意自己的陞遷情形。 24. 以自己付出的努力與工作表現而言，我應該有更好的陞遷狀況。(R) 25. 本機關陞遷決定的結果讓人覺得工作表現好就能獲得陞遷 26. 我不認為同仁對本機關的陞遷結果都滿意（R）。 30. 我滿意擔任公務人員以來的陞遷狀況。	Likert 6 點尺度，5 題分數加總	.779	問卷／主觀
升遷政治性	17. 本機關的陞遷決定經常受本機關以外人士影響。 18. 本機關中對長官的逢迎奉承比起工作表現，對陞遷更有幫助。 19. 本機關中決策者人馬或某些特定的人較容易獲得陞遷。 20. 本機關的陞遷決定，常是反映決策者的個人偏好。 21. 本機關陞遷決定主	Likert 6 點尺度，6 題分數加總	.722	問卷／主觀

變數名稱	操作化定義	編　　碼	Cronbach's Alpha 值	資料來源／屬性
	要是以所需人力為考量。 22. 本機關陞遷決定，未必以工作表現為最主要因素。			
中央地方年資比	中央機關年資／總年資	中央機關年資／總年資	N / A	人事資料庫／客觀
性別（男）	性別	男性＝ 1；女性＝ 0	N / A	問卷／主觀
已婚	婚姻狀態	已婚＝ 1；未婚＝ 0	N / A	問卷／主觀
初任年齡	1993 年進入公職時之年齡	1993 年齡數字	N / A	人事資料庫／客觀

資料來源：本研究、張鎧如等（2015）

附錄二：變數描述統計

變數	N.	平均數	標準差	最小值	最大值
歷年考績甲等比例	1559	82.32	14.23	16.67	100
2012 典藏考績分數	1480	83.21	2.53	76	94
2012 自評績效分數	1560	13.94	1.97	3	18
機關轉換	1567	0.16	0.11	0	0.6
教育程度	1563	3.34	0.79	1	5
政治技能	1553	33.41	5.86	9	54
公共服務動機	1539	60.37	7.20	34	84
家庭犧牲	1558	13.14	4.08	4	24
升遷滿意度	1550	15.58	4.06	5	25
升遷政治性	1540	22.37	4.88	6	36
中央地方年資比	1567	57.20	42.65	0	100
性別（男）	1567	0.50	0.50	0	1
已婚	1561	0.88	0.32	0	1
初任年齡	1567	27.50	4.10	20	44

資料來源：本研究。