



Delphi分析發展、 應用問題與改進(上)



葉連祺/國立嘉義大學教育系教育行政與政策發展碩士班教授

摘要

Delphi分析廣受教育研究採用，屢見創新分析實務做法，但存在一些應用問題影響其分析效益。對此，本研究說明有關Delphi分析理念、分析取向、運作流程、應用實務等發展成果，進而討論涉及專業評估不足、分析嚴謹性偏弱、新法應用落後、採用係數簡略、運作生態薄弱等八項問題，並提出強化學理和方法研究、建立品管機制、結合多元統計、設計專用程式等改善做法。另設計適用傳統Delphi分析的程式適合SPSS作業環境，其操作簡易，提供豐富分析資訊，能大幅降低分析成本。

關鍵字：德懷術、共識評估、變異係數

投稿日期：2026年1月5日

完成修稿日期：2026年1月28日

DOI: 10.6423/HHHC.202603_(162).0007

Development, Application Problems and Improvement of Delphi Analysis (I)

LAIN-CHYI YEH¹

Abstract

Delphi method are used prevail among education study. Various innovative analytic practices about Delphi analysis have been proposed, but on the other hand several application problems also impacted its benefits. The study illustrated analytic ideas, analysis approaches, operational procedures, and practices of application on Delphi method. Moreover using Delphi method encountered with eight problems that related with insufficient professional assessment, weak rigidity in analysis, lag of adopting new methods, using oversimple consensus measurement coefficient, and frail operational ecology. Therefore, the study proposed some resolution strategies: enforce research on theory and method for Delphi analysis, build quality control mechanism, use various statistical methods, and design specific programs. Finally several specific programs for classic Delphi analysis worked on SPSS platform that conduct easy and rapidly, provide rich analytic information, as well as reduce most cost of analysis.

Keywords: Delphi method, consensus assessment, coefficient of variation, CV

¹ Master's Program of Educational Administration and Policy Development, Department of Education, National Chiayi University

壹、前言

Delphi分析(Delphi method, Delphi technique, Delphi analysis, 譯為德懷術、大慧法等)被廣泛應用於決策分析、民意調查、指標研製等方面,主要藉助專家意見共識,以確認最佳決策、提出最適方案、或建立指標和項目。多年來臺灣地區教育研究者也廣泛使用Delphi分析於政策和方案制定、項目和指標編製,然而觀察應用成果極大多數採用簡單的描述統計如平均數、標準差、變異係數(coefficient of variation, CV)等評估共識性(consensus),使用共識性差異指數(Consensus Deviation Index, CDI)(鄧振源,2012)但缺乏統計考驗提供客觀評判證據,也很少推介關於Delphi的新方法如RAND/UCLA適當法(RAND/UCLA Appropriateness Method, RAM)(Fitch, *et al.*, 2000)、RAND/USC OPTIC政策專家評定法(RAND/USC OPTIC method for policy expert ratings, ROMPER)(Grant & Smart, 2024),這些都有礙於Delphi分析的推廣和應用品質,有待討論和謀求改善對策。至於,模糊Delphi分析(fuzzy Delphi method, fuzzy Delphi)被不少研究者採用,所累積研究成果頗豐,但是顯現眾聲喧嘩的局面,已知分析做法甚多,考量有限篇幅,將另行撰文討論。

考量既有教育領域研究成果透露出對於Delphi分析的了解和應用似有不足,其影響甚巨。對此,本研究先大致檢視Delphi分析的學理和方法發展、相關研究、應用等概況,以建立基本的了解,接續針對檢視既有研究文獻和論述,指出若干值得關注的問題,加以討論,再配合論理推導和實證資料考驗,據以提出對應的可行解決對策和做法。具體言之,研究目的包括:1.檢視Delphi分析的學理和發展概況;2.探討研究應用Delphi分析值得關注問題和遭遇問題;3.探討應用Delphi分析的改善思維和解決做法。

又需要界定本文所指的“問題”(problem),是指討論事項符合應用Delphi分析時採用做法似乎違背該法倡議的理念、採用做法未達預期彙總意見和建立共識的成效、未持續創新或更新既有做法以致於未能增益使用效益等項規準,若能洞視問題發生原因,進行針對性的根本改善措施,將使得分析效益提升和擴大。而此處“問題”係符合前述三項規準之一,如果沒有了解和改善,最終會危及Delphi分析品質和應用效益,必須重視。

貳、Delphi分析發展概述

一、Delphi分析學理和運作概述

基本上，Delphi分析係一種循環（iteration）、匿名、結構性、團體本位引導看法的技術，乃多次詢問多位專家相同問題，並分享其他專家看法，以協助發展出共識（Khodyakov, Grant, Kroger, & Bauman, 2023）。為RAND公司在1940年代晚期至1950年代早期針對空軍研究專案所創，爾後廣受各領域研究者採用，並引發眾多擴展性和結合性研究，被視為彙總眾議共識的良好方法。後續，在1980年代中期RAND公司另與加州大學研究者共同倡議RAND/UCLA適當法（RAM）（Fitch, *et al.*, 2000），也引起一些論者進行擴展和考驗，後續又發展出ROMPER法，這些可視為修正的Delphi（modified Delphi）。

Delphi分析理念係基於俗諺“三個臭皮匠勝過一個諸葛亮”（two heads are better than one）（Dalkey, 1969），藉助眾人集思廣益，獲得最佳決策，其學理基礎可參見Linstone和Turoff（2002）、Hanafin（2004）、Hasson、Keeney和McKenna（2025）等討論，包括實證主義典範（positivist paradigm）、詮釋典範（interpretative paradigm）、社會建構論（social constructivism）等。關於本方法的特性，Dalkey（1969）、Rowe和Wright（1999）等指出匿名、循環、控制回饋（controlled feedback）和統計總合團體反應為Delphi分析的特性。此四者係可顯見的結果；歸根究底，本研究認為為確保Delphi分析合理運作和產出，其必須植基於幾個假設：包括平等參與、專業尊重、民主共識、科技協助等。平等參與指參與諮詢者在諮詢期間地位一律平等，此因匿名和無直接面對面接觸，故有效避免面對面會議造成莫名團體迷思或霍桑效應產生的偽共識弊端；專業尊重指邀請具專業素養的專家或代表性的利害關係人參加諮詢，並尊重每位專家提供的不同意見，採取匿名措施能保持諮詢專家獨立判斷的空間，避免受到他人干涉和干擾；民主共識指參與者能本諸自重和尊人的精神，基於良心和考量團體利益不吝陳述意見，並以匯聚取得絕大多數參與者共識的看法，專家可安心檢視他人提出修改意見的利弊得失，基於自己專業判斷做出評估；而科技協助意指藉助採用科技化環境或平台、及科學化分析技術（主要是描述和推論統計方法），以能有效促進共識達成或異議偵測品質及成效，避免因傳遞評估訊息的時間延宕或傳送內容失真，而有扭曲專家共識達成的品質和合宜性。因此，可確保Delphi分析結果能代表大多數參與者匯聚智慧、專業深思熟慮、達高共識、有高公信力、具適切性的客觀結論，供決策或進階應用之需。

關於 Delphi method 的英文名稱，在文獻中出現 Delphi method、Delphi analysis、Delphi study、Delphi approach、Delphi survey、Delphi technique、Delphi research、Delphi survey、Delphi-like survey、Delphi survey technique、Delphi exploration、Delphi process、Delphi methodology、Delphi consensus methods、Delphi consensus、Delphi expert panel、Delphi-based analysis、Delphi panel-based analysis 等稱呼，細視內容大致相似。至於中譯名稱，多採音譯和意譯兩種做法，前者如稱得懷術，後者如稱大慧術，可謂多如繁星，亦見保留英文名稱的變通音譯做法，如 Delphi 技術、Delphi 法等。大抵上，以 Delphi method、Delphi technique、Delphi analysis 三者較被採用；而觀察 RAND 公司在其出版的論著中係使用 Delphi method 一詞 (Khodyakov *et al.*, 2023)。檢視前述所列名稱，不難看出強調其為 1. 一種研究取向、研究法或研究技術，2. 一種探求共識的方法或歷程，3. 一種採用組成評估團隊的做法；換言之各名稱顯示應用者看待該方法的定位及側重點，本研究側重討論應用該方法及衍生問題，本質上視為一種分析，故採用 Delphi 分析 (Delphi analysis) 一詞。

再者其後續衍生頗多做法，故論者提出分類觀點。對於分類，可以探討目的為依據，Häder (2014) 根據分析功能提出四類：1. 總合理念 (Aggregation of ideas)、2. 最精準預測不確定議題 (Most precise prediction of an uncertain issue)、3. 匯集專家對混淆議題的意見 (Collecting expert opinions on a diffuse issue)、4. 獲取共識 (Consensus) (Niederberger & Spranger, 2020)。San-Jose 和 Retolaza (2016) 針對傳統性和目的 (traditional and Goals) 區分出 Delphi、倫理 Delphi (ethical Delphi) 和論理目的 Delphi (Ethical Goal Delphi)。Paré、Cameron、Poba-Nzaou 和 Templier (2013) 提出古典 Delphi (Classic Delphi)、政策 Delphi、排序型 Delphi (Ranking-Type Delphi)、決策 Delphi (Decision Delphi) 四類；Strasser (2017) 再擴展提出古典 Delphi (Classic Delphi)、政策 Delphi、排序型 Delphi (Ranking-Type Delphi)、決策 Delphi (Decision Delphi)、爭論 Delphi (Argument Delphi)、評估—回饋—對話—評估 Delphi (estimate-feedback-talk-estimate Delphi, EFTE Delphi) 和解析政策 Delphi (Disaggregative Policy Delphi) (Tapio, 2003)。Niederberger 和 Deckert (2022) 也提出即時 Delphi (Real-Time Delphi, RT Delphi)、市場 Delphi、政策 Delphi、爭論 Delphi、團體 (group) Delphi、慎思 (deliberative) Delphi 等類型。Khodyakov、Grant、Kroger 和 Bauman (2023) 歸納出四分類觀點：政策 Delphi (policy Delphi) 及有關方法與異議 Delphi (dissensus Delphi)、修正型 Delphi (modified Delphi)、即時 Delphi、專家視界 (ExpertLens) Delphi (可視為一種 E-Delphi)。其他如

Hasson和Keeney（2011）歸納出10類型，包括微型Delphi（Mini-Delphi）。總結上述諸多分類觀點顯得龐雜，實肇因於不同視角所致，有待整合。

大抵而言，採取分類視角的依據能包括實施目的、討論事項、評估人員、實施平臺、評估做法等項見表1，據以得以形成多類型的Delphi分析，如古典Delphi分析聚焦議題為A1,B1,C1,D1,E2a的組合，若聚焦編製指標則是A1,B3,C2,D2,E3的組合做法。當然也可以試著以此做為檢視已知論者提出多元和龐雜的Delphi分析類型及分析取向，或者去構思出合宜的新Delphi分析做法，或據以綜析（systematic review）既有Delphi分析採行做法的情形。

表1
Delphi分析構成內容之思考項目

A實施目的	B討論事項	C評估人員	D實施平臺	E評估做法
1獲取共識	1議題釐清	1利害關係人	1文件郵遞	1質性意見
2了解異議	2政策制定	2精熟專家	2電信通訊	2a計量評值（分析項目）
	3項目設立（工具編製）		3即時視訊	2b計量評值（分析評估者）
	4事項預測		4線上軟件	3混合做法
	5.看法回饋處理			

就運用研究法而言，已見採用質性（qualitative）、量化（quantitative）和混合方法（mixed-methods）三種取向（Sekayi & Kennedy, 2017），實際上以採用量化取向者占大多數。若就分析對象，即評估項目或者評估者的共識，能分成Q型Delphi（Q Delphi analysis）和R型Delphi（R Delphi analysis），前者著重分析各項目被不同評估者評定的共識，焦點是項目，這是一般Delphi分析的做法；後者強調檢視各評估者對於項目評定的共識，焦點是評估者，但頗少被採用，兩類分析的差異見表2說明。簡言之，若要偵查具異議的評估者（dissensusor），了解原因，則宜使用R型Delphi分析做法。

表2
計量Q型Delphi分析和R型Delphi分析之比較

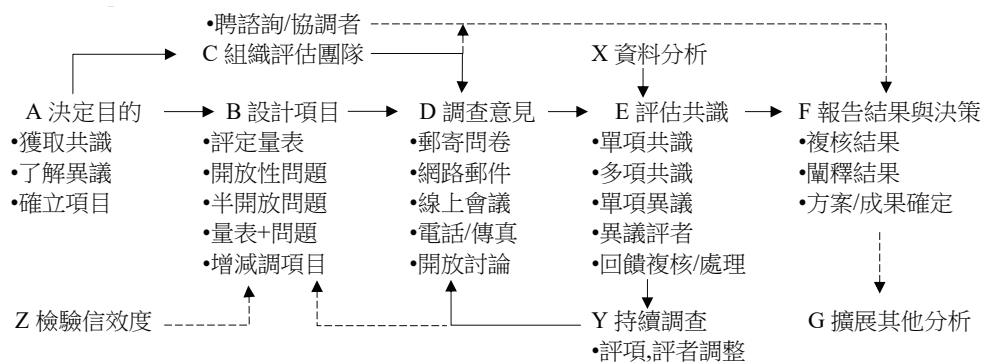
	Q型Delphi分析	R型Delphi分析
分析對象	項目	評估者
共識檢視	全部評估者對不同項目評估共識	不同評估者對全部項目評估共識(或異議)
分析假設	評估者能客觀評估項目,了解項目評估共識優先於評估者全體共識	評估者可主觀評估項目,了解評估者全體共識優先於項目評估共識
資料矩陣	評估者(列)×項目(行)	項目(列)×評估者(行)
分析效益	偵測達共識看法的項目,確認屬於異議的項目	偵測達共識看法的評估者,確認屬於異議的評估者
統計焦點	項目內部一致性信度、內容和構念效度、項目反應一致性、集中性和離散性	評估者反應一致性、集中性、離散性和影響性

至於實施流程部分，論者看法不一，Day和Bobeva（2005）指出探索（exploration）、蒸餾（distillation）和功用（utilization）三階段模式觀點，其

對應於規劃、執行和分析報告等流程，Sekayi和Kennedy（2017）提出四輪回的質性Delphi分析流程構想，Belton、MacDonald、Wright和Hamlin（2019）構思六階段做法；Khodyakov等（2023）指出設計研究、收集資料、分析資料和報告結果四階段；而一些系統性評論應用Delphi分析的研究（如Schifano & Niederberger, 2025）也顯示諸多流程。綜言之，可知論者對於其實施流程看法頗紛雜，但細究內容則大同小異，大致包括決定目的（set-up goals）、設計項目（designing items）、組織評者（organizing panel）、調查意見（survey opinion）、評估共識（assessing consensus）、報告/決策（reporting and decision making）等關鍵階段，或再加上擴展分析（expanding analysis）階段，其中調查項目和評估共識兩階段構成循環圈，通常當評估達共識後即可進入報告/決策階段，此見圖1揭示整個實施流程梗概。簡言之，一般常見流程為A決定目的→B設計項目→C組織評估團隊→D1調查意見→E1評估共識→[D2調查意見→E2評估共識→D3調查意見→E3評估共識]→F報告結果與決策，此處[]表示視情形可能會發生的運作流程。

基本上該圖是基於量化和混合方法取向的Delphi分析，其中E.評估共識階段的回饋/複核處理部分是指評估者針對前輪回回饋意見的處理結果表示適切等看法，僅少數研究採用此項做為。Y.持續調查部分，通常係指評估共識未達成通過標準，再請同一批評者持續針對同一份項目進行評估，一些另類調整做法可能會刪去不當項目、潤飾項目不佳敘述、調整項目至適合的向度，也可能排除少數異議評者（指評估意見明顯違反常理或專業倫理）參加下一輪回評估，甚至是考慮提升專業性而增加納入新評者、或因項目大幅度調整更換而重新規劃新評者團隊，另邀請新一批評者（指利害關係人，呈現另一角度看法）參加下一輪回評估，大體來說調整評者做法僅極少數研究採行。在C.組織評估團隊階段，此之前可先成立諮詢團隊或協調人（moderator），決定評者可採取方案主持人自定、諮詢者提供名單、入選評者推薦（即滾雪球抽樣）等方法，並能於後續調查項目和評估共識階段調整不適任評者（指評者評估行為違反專業倫理），新增或汰淘評者，而RAM法著重於建立一套建立評估團隊的流程和運作標準，包括選擇專家或協調人、辦理宣導RAM法會議和相關教育活動、實施評估項目和回饋意見處理做法等（Fitch, *et al.*, 2000）。

圖1
 Delphi分析之通則性實施流程



註：實線表示必要的進程，虛線表示是需要可選擇的工作事項。

圖1說明在A.決定目的部分，得確定探討焦點問題是協助決策、建構項目、或分析方案，決定目的是獲取決策共識、了解爭論異議、或者確立可用項目，進而思考使用的Delphi分析方法類型，如政策Delphi，可詳見前述討論。而B.設計項目部分，可使用質性調查即採取開放性問題徵詢看法，多數使用量化調查做法，要求參與者針對項目以3、5、6、7、9、10或11點Likert量尺表達看法（指認同、重要、可行、適切或贊成程度）（示例見附錄一），或者要求評定等級、選擇項目，另有採取混合方法，即使用開放性問題和多點量尺評定看法（Jünger *et al.*, 2017）；一些研究因為Delphi分析目的在發展指標或研訂項目，故在此階段Z.檢驗信效度部分，考驗項目的內部一致性信度（多採取Cronbach α 係數）、評分者信度、內容效度和構念效度。在C.組織評估團隊部分，多採取邀請利害關係人、有關專家等組成團隊，進行焦點評估，組成人數依需要而定，一般研究多僅成立一個評估團隊，亦可採用群體Delphi分析（group Delphi），建立多個評估小組團隊（Niederberger & Renn, 2023），甚至採取修正Delphi分析（modified Delphi），針對不同輪回實施目的，每個輪回皆重組評估團隊參與評估（Gohres & Kolip, 2023）。就實務運作來說，組成評估團隊的人數至少3人，至多則無限制，除考慮評估運作所需符合專業知能和經驗的可邀請參與人員數、評估時間、評估經費成本等，尚可考量參與評者數可減少錯誤度（error reduction）及淨錯誤減少變化（net change of error reduction）（Pazilah, Hashim, & Yunus, 2024），見表3說明；換言之，據此推知在兼顧減少錯誤度和評估成本（指時間、經費等）時，適當的採用評者數似乎至少為9人，至多不逾21人。

表3

Delphi分析評者數減少錯誤度之情形

評者數	減少錯誤度	淨錯誤度變化	評者數	減少錯誤度	淨錯誤度變化
1~5	1.20~0.70	0.50	17~21	0.50~0.48	0.02
5~9	0.70~0.58	0.12	21~25	0.48~0.46	0.02
9~13	0.58~0.54	0.04	25~29	0.46~0.44	0.02
13~17	0.54~0.50	0.04			

註：取自“Refining the TPACK framework: A fuzzy Delphi approach to 21st-century competency and self-efficacy constructs” by F. N. Pazilah, H. Hashim, & M. M. Yunus, 2024, *Cogent Education*, 11(1), 2428886, p.6。

D. 調查項目階段，乃決定 Delphi 分析成敗的關鍵部分，過去採取郵寄信件或電子郵件做法，近年來結合特定軟體或分析系統如 Welphi、edelphi、DelphiManager 等，使用線上會議或線上調查等做法，可即時了解其他評者意見整理及回饋，漸受重視；另可視需要聘請諮詢者和協調人（moderator），控管流程運作品質，提供改善思維或措施。

再者 E. 評估共識部分在檢視共識性（consensus）、適切性（appropriateness）或重要性（或必要性），通常以質性或計量方法評估單項共識和多項共識，大多數評估做法如檢視變異係數（CV）等都屬於這類評估，但少見聚焦於發現屬於異議的單項和持異議的評估者；若未通過共識評估，則再重複一輪回調查看法和共識評估，反之如果通過共識考驗，則進入 F. 報告結果與決策階段。至於 G. 擴展其他分析部分，係將 Delphi 分析作為前置研究，如定位建構項目，接續做 AHP、IPA 等進階分析，以建立項目權重體系或找出需關注項目。

至於應用質性取向的 Delphi 分析則實施流程略有不同，綜合論述（Foth, *et al.*, 2016; Hallowell & Gambatese, 2010; Sekayi & Kennedy, 2017; Klenk & Hickey, 2011），主要是不採用評定量表收集看法的計量資料，改採取腦力激盪法（brainstroming）、名義團體法（Nominal group technique, NGT）等技術，針對開放性問題（open-ended question）或封閉性問題，經過多輪回評者提出看法或回饋意見、彙總提出陳述（statement）、以及增修陳述再徵詢和確認，直至多數評者達共識，再無增修、調整陳述看法提出為止，即可形成最終具高度共識的陳述，做為決策方案、研訂項目、規劃預測等用途之需。

而採取混合方法（mixed-methods）取向時，有論者使用質性分析和量化做為的同時並行推進做法（Suominen, Hajikhani, Ahola, Kurogi, & Urashima, 2022）：亦見採用先計量分析確定項目，再多輪回質性討論以確定最終方案（Dragostinov, *et al.*, 2022）；另有先使用半結構訪談初定項目，接續兩輪回進行計量評估項目，最後採用結構訪談，針對兩輪回量化評估結果，做確定和研提最終項目（Firth, *et al.*, 2019）。換言之，進行混合方法取向 Delphi 分析可有質量並

行、先質後量、先量後質等三個基本型，質量並行型得面對如何合理統合質性討論和量化評估看法衝突的問題，另兩種類型則似無類似困擾，何者較優宜視研究對象和研究目的而定。

二、Delphi分析發展概述

觀察多年來Delphi分析發展和應用可謂蓬勃，出現不少創新構想和做法。在分析學理部分已提出政策Delphi (policy Delphi)、修正Delphi (modified Delphi approach)、ROMPER法 (the RAND/USC OPTIC Method for Policy Expert Rating) (Grant & Smart, 2024)、RAND/UCLA適當法 (RAM) (Fitch, *et al.*, 2000)、紮根Delphi (Grounded Delphi Method, GDM) (Päivärinta, Pekkola, & Moe, 2011)、真實世界Delphi (real-world Delphi)、市場Delphi (market Delphi)、即時空間Delphi (real-time spatial Delphi) (Zartha Sossa, Halal, & Hernandez Zarta, 2019) 及其他新創做法 (Khodyakov, Grant, Kroger, & Bauman, 2023)，使得Delphi分析更有效適用於彙總共識看法、發現異議意見、建立指標項目、提出預測看法等用途。在資料分析方法方面可見提出最大-最小Delphi (max-min Delphi method) (Ishikawa *et al.*, 1993)、模糊Delphi (fuzzy Delphi analysis)、機率Delphi (probabilistic Delphi method)、灰色Delphi (grey Delphi) 等，顯示已擴大傳統Delphi分析採用Likert量尺資料的收集看法做法，納入模糊集、灰集等類型資料，以更精確收集評估人員的評估意見；而質性Delphi分析亦見結合概念構圖 (concept mapping) (Klenk & Hickey, 2011)。另有論者倡議電腦化Delphi分析 (Computerized Delphi Technique)，使用電腦科技協助整個實施流程，增進運作效率和實施品質 (Seker, 2015)。而實施途徑或應用介面角度，亦有即時Delphi (Real-Time Delphi, RT Delphi) (Aengenheyster *et al.*, 2017)、雲Delphi階層分析 (Cloud Delphi hierarchical analysis)、線上Delphi (online Delphi)、E化Delphi (e-Delphi technique)、網路本位Delphi (web-based Delphi study) 等，可見以傳統採取郵寄、電話、傳真等傳送徵詢意見做法，擴大採用網際網路發展的有關通訊媒介。

又綜觀RAND公司大致發表了三個Delphi分析方法 (Dalkey, 1967, 1969; Dalkey & Helmer, 1963; Fitch, Bernstein, Aguilar, Burnand, & LaCalle, 2001; Grant & Smart, 2024)，見表4彙整和比較，其中ROMPER法增加了評者在最終輪回根據檢核項目及採取開放性問題方式，提供評論回饋意見，參考項目可見Grant和Smart (2024) 說明，甚具參考價值。又已見提出多種Delphi分析法，RAM法也有修正做法 (Kraines, *et al.*, 2020; Verburg, 2022)，值得注意。

表4

RAND公司提出三代Delphi分析方法之比較

	Delphi	RAM	ROMPER
提出時間	1953	1980年代	2024
參與機構	USAF	UCLA	USC
方法名稱	Delphi method	RAND/UCLA appropriateness method	the RAND/USC OPTIC Method for Policy Expert Rating
適用類型	無特定	無特定	政策Delphi
適用取向	量化取向為主	量化取向為主	兼採量化和質性取向
方法特點	匿名回應、循環與控制回饋、統計團體回應	同左，專業評估確保	同左
控制偏誤	降低個人獨斷偏差、避免無關的溝通、減少順從團體壓力	同左，增加評者專業評估教育、調查評者評估依據/理由、調查評估結果品質	同左
實施平臺	未限定	未限定	線上
使用工具	問卷(採用Likert 3,5,7量尺)	問卷(採用Likert 9點量尺)	問卷(採用Likert 9點量尺)
參加對象	經挑選的專家	同左	同左
實施流程	建立團隊→多輪回(調查自評看法→評估共識)→提出結論	建立團隊→教育團隊→多輪回(調查自評看法和評估依據→評估共識)→評論結果→提出結論	建立團隊→多輪回(調查自評看法→評估共識)→提出回饋評論→提出結論
項目評定	重要性或影響性	GRADE架構，評估適切性、必要性，說明評估依據(如經驗證科學證據、僅薄弱科學證據、專家意見、自身經驗或同行意見)	GRADE架構，主要評估效果(視評值6-9增效、5無效和1-4降效)，亦能評估可接受性、可行性、可負擔性、公平性等(視評值1-3低度、4-6中度、7-9高度)
評估看法	每輪回評估項目	每輪回評估項目(採用1-9量尺)，輪回1和2評估項目適切性，輪回3評估項目必要性	每輪回評估項目(採用1-9量尺)，最終輪回依據檢核項評論對整個執行做法的意見(採用1-7量尺，1極不同意，7極同意)
資料處理	簡單統計分析評值，如平均數、中數、標準差、全距、IQR等	平均數、中數、標準差、IPRAS指數和DI指數	ROMPER法：包括評值描述分析和分配情形、評值變化、相關(評者評值)、IPRAS指數和DI指數、森林圖；根據ACCOR和檢視質性意見
報告規範	眾多研究者已提出參考規範		質性結果依據SRQR規範和呈現GRADEPro證據報告表
有關發展	已提出修正Delphi法	已提出修正RAM法	

又Delphi分析可單獨分析或者與其他分析技術結合應用，檢視研究文獻，大致能歸納出立項、增益、決策和預測四類取向，其設定分析功能和定位、及運用流程可見表4示例的說明。其顯示Delphi分析能做為前導分析，提供具共識的項目做為後續分析權重、分析項目信、效度的合理基礎，或是指出具統計群聚合理證據的項目或指標，此即增益類取向(added benefits approach)，目前已見諸多研究；亦能藉助文獻評論或調查以收集可供討論的項目、方案、措施等，供

接續進行Delphi分析，獲取共識或發現異議，達成提出合理方案或措施、建立可用項目或指標，此係立項類（building items）或決策類（decision making）取向應用，為多數研究所採用；至於預測類（forecasting for planning），較為少見，其係藉助文獻評論、分析已知資料或調查來收集可供討論的項目及有關課題，提供參與評估的專家參考，以期取得具共識和合理性論述基礎的預測規劃作為或構想，供有關機構或政府部門決策參考。而ROMPER法、RAM法亦可適用表5所提的運用流程。

表5

Delphi分析結合其他分析技術之運用流程示例

取向	Delphi分析功能	定位	運用流程		
立項類	確立合適項目	總結性	系統性文獻評論	→ Delphi分析	→ 提出項目
			SWOT,ABCD分析等	→ Delphi分析	→ 提出項目
增益類	發展合適項目	前導性	Delphi分析	→ PCA, AHP, IPA等	→ 提出(權重)指標
			Delphi分析	→ 集群分析等	→ 提出合宜指標
決策類	指出合適決策	總結性	意見/做法調查	→ Delphi分析	→ 提出決策
預測類	提出共識預測	總結性	系統性文獻評論、既有資料分析、意見/做法調查	→ Delphi分析	→ 提出預測/規劃

此外涉及進行和報告Delphi分析的規範亦見討論，如Diamond等（2014）提出遍及研究目的、評估者和過程的六項規準；Jünger等（2017）提出執行與報告Delphi研究規範（Conducting and REporting DELphi Studies, CREDES），包括實施理念、規劃設計、執行、報告四部分的16項建議；Niederberger等（2024）提到ACCORD（ACcurate COnsensus Reporting Document）規範、DELPHISTAR規範（Delphi studies in social and health sciences-recommendations for an interdisciplinary standardized reporting）具參考應用價值；Hasson、Keeney和Mckenna（2000）針對研究問題、研究理念、文獻探討、研究方法、資料分析、討論與結論、附錄等部分提出指引，並提出9項查核項（checklist）供參考；Spranger等（2022）構思關於報告Delphi分析結果的綱要項目，包括標題和摘要、理念、正式內容、知識基礎、知識應備、Delphi類型和角色、樣本、調查工具、實施輪回、回饋、評鑑、結果、資料品質和詮釋、討論和發現限制、應用等方面的15個應思考項；Khodyakov等（2023）研議在設計、資料收集、資料分析、結果報告四個方面的39項建議（recommendations）做法，及包括核心和額外兩部分20個檢核項的Delphi關鍵評估工具（Delphi Critical Appraisal Tool, DCAT）檢核表（checklist），供管控執行品質；Landeta和Lertxundi（2024）指出包括專家團隊品質、專家獲取資訊品質、專家互動品質三個領域的14個評估指標；Hanafin（2004）、Naisola-Ruiter（2022）都指出進行Delphi分析應注意的研

究倫理議題 (ethical issues)。綜觀之，可見論者對 Delphi 分析規範和品質確保的看法繁簡不一，不外包括實施理念、規劃設計、執行、報告、成效評鑑五部分，其中成效評鑑部分係針對 Delphi 分析成果的應用成效做追蹤評估，性質類似後設評鑑 (meta evaluation)，並未普見於前述各論述討論，有待探討；至於比較而言，Khodyakov 等 (2023) 的觀點相對較為完整，能夠參考應用。

再者，Delphi 分析自提出以來，除了眾多關於方法、應用等方面的研究論述之外，亦不乏一些論者提出針砭之見 (Hanafin, 2004; Hasson & Keeney, 2011; Hasson, Keeney, & McKenna, 2025)，主要涉及方法嚴謹性 (rigour)、信度 (包括評估者間測量、平行測量)、效度 (包括內容效度、構念效度)、可信賴性 (trustworthiness) 等問題。其中方法嚴謹性是過去 Delphi 分析最被詬病之處，RAND 後續提出的 RAM 法和 ROMPER 法 Delphi，嘗試採用 GRADE 架構、ACCOR 規範、GRADEPro 證據報告表等做法，以增進執行的嚴謹性；但是有關於選擇評者造成評定偏誤 (bias)，即評者評定不夠客觀、評者依據有限的專業自信、評者評定依據參與評者的權威名聲等問題則仍亟待克服。又信、效度考驗其實涉及提供客觀可信的證據，去支持 Delphi 分析的項目是有合理的構念效度、內容效度、項目評分內部一致性信度、評分者信度，前三者對於目的是構建指標 (indicator) 時係特別重要的證據，仍少見應用，而評分者信度則關涉評估者看法的共識性或一致性達成與否，提出客觀佐證，此已見到頗多研究提出 Kendall w 係數做為說明；另一方面，Delphi 分析結果的穩定性和可重製性 (stability or replicability) 則涉及跨團隊、跨時間、跨地域等條件時，能夠比較和驗證其分析結果交叉效度 (cross-validity) 的證據，是有必要，但是基於絕大多數 Delphi 分析都屬於一次性研究 (one-time study)，這部分的應用實例和探討十分稀少。綜合言之，論者對於 Delphi 分析所提批評，有助於增益其分析品質，已見不少改善做法，仍有持續努力之處，如尋求檢測可重製性的測量係數，並考驗其效力。

參、Delphi 分析應用問題探討

觀察臺灣地區 Delphi 分析的應用頗為眾多，但存在不少“問題”或“困境”值得討論，以利於後續進行改善，提升 Delphi 分析品質和效益，以下簡述之：

一、重視和堅信專家能做下合乎專業評估論理的共識評估， 忽略人性評估受外部環境影響的問題

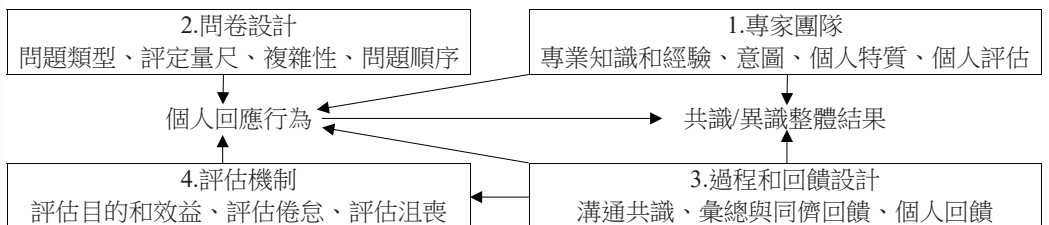
Delphi分析需要注重可信度（credibility），而關於評估者的選擇組成、實施評估實為影響關鍵。一方面評估者提出看法是基於自我認知，有時未必正確和合理，會遭遇專業陷阱（expertise trap）或認知固著（cognitive entrenchment）問題，而產生看法偏執或存在盲點（Sohst, Acostamadiedo, & Tjaden, 2023），使得評估意見的專業性降低。另一方面，Delphi分析也具有類似集體決策的性質，當專家團體對項目評估看法達可接受的共識性檢視，就形成一個類決策的結果，據此成員在多輪迴評估時可參考其他評者回饋意見，決定修改或保留原評估看法，這些修改意圖就可能受到團體迷思（groupthink）或其他來自團體隱性壓力的影響（Jorm, 2015）、權威選擇盲從（指盲從成員聲望權威）（Hanafin, 2004）。又Delphi分析經多輪回反覆評估，耗時且需費心檢閱，可能造成評者產生評估疲倦感（assessment burnout），不願意繼續堅持己見，選擇屈從提供的已知團體多數評者看法回饋資訊，進而改變自己的評值，論者（Barrios, Guilera, Nuño, & Gómez-Benito, 2021）就指出頗高比例（75%）應用研究論述顯示評者在第三輪回改變看法，遵從其他人看法，顯然Delphi分析本身的機制設計也可能影響評估意見轉向，這表明要求評者說明改變看法的原因似有必要。而Hallowell和Gambatese（2010）提出集體無共識（collective unconscious）、對比效應（contrast effect）、機遇忽視（neglect of probability）、von Restorff效應（von Restorff effect, 或稱isolation effect, novelty effect）、我方偏誤（myside bias）、新近效應（recency effect）、首序效應（primacy effect）、獨斷（dominance）等影響Delphi分析的偏誤，其中集體無共識、獨斷等與評者專業有關，其提出的對應改善做法值得參考。

通常Delphi分析假設參與評估的人員不僅具有評估項目的專業知能，也能做下合乎專業倫理的判斷，故能形成合理的評估共識，即預設成立“專精專業知能→專業倫理判斷→合理評估共識”的假設。此假設要成立，必須提供足夠良好的證據。所以，若干研究文獻就報告各輪回（round）參與評估專家的特性資訊，包括地區（或國家）、年齡、性別、人員專業類別、職務（或職位）、相關專業經驗（如專業服務年資）、教育程度等（如Guthrie *et al.*, 2022; van Grootven *et al.*, 2018），供判斷評估結果的專業性和合理性；另見論者（陳青山等，2004）建議分析評估者的積極係數和權威係數，以說明評估人員的專業性。然而不少臺灣地區研究缺少報告這類參考資訊，只籠統說明參與評估人員人數、幾句話簡略道出

參與人員皆具專業知能和經驗，這實在不利於檢視評估人員是否適當、人員評估結果是否可信，有待補強有效措施。

共識性是Delphi分析的核心，宏觀而言，整個Delphi分析係受到諸多因素的交互影響，Niederberger和Sonnberger（2025）談到執行Delphi分析時影響判斷資訊存在四向度因素，包括脈絡（contextual）、個人、認識論（epistemic）和方法學（methodological）向度，脈絡向度有實施機構、贊助關係、研究時間、主題等，個人向度有社經背景、個人特質、社會溝通技巧、參與和形成判斷動機，認識論向度亦有特定團體集體想法和理解文化、特定團體溝通型態，至於方法學向度則有調查設計、資料蒐集模式、意見回饋設計、干擾因素（moderation）。而Schifano和Niederberger（2025）另指出類影響因素：專家團隊（expert panel）、問卷設計（questionnaire design）、過程和回饋設計（process and feedback design）、個人回應行為（individual response behavior）、共識或異議整體結果（overall results）等影響形成團體共識，而前三類因素影響後兩類因素，且個人回應行為影響共識或異議整體結果；若再納入評估機制（assessment mechanism）所造成的評者倦怠和沮喪，據此能構成圖2的影響架構。在專家團隊部分包括專業知識和經驗、意圖、個人特質、個人評估等項，問卷設計部分包括問題類型、評定量尺、複雜性、問題順序等項，過程和回饋設計部分則包括溝通共識、彙總與同儕回饋（aggregated vs. peer feedback）、個人回饋等項；至於評估機制部分是指Delphi分析採用多輪回評估達共識的設計，造成評者因耗時和多次費心評估，覺得自己看法屬於異議而感到疲憊（即評估倦怠），或者察覺該分析目的偏頗或預期效益不佳，自己苦心反映看法實屬枉然而放棄堅持（即評估沮喪，assessment depression），進而改變前次看法為順從多數人意見。簡言之，此架構指出需要全面關注和有效處理衝擊Delphi分析共識性的諸多因素，方能確保和產出高品質結果。

圖2
Delphi分析共識性之影響因素結構



註：增修自 “How Delphi studies in the health sciences find consensus: A scoping review” , Schifano, J., & Niederberger, M., 2025, *Systematic Reviews*, 14(14), p.4.

其次，聚焦於獲取共識決的Delphi分析，需要假設評估人員不受認知偏執、政治正確、團體思考、團體壓力等因素，不會出現扭曲或違反專業倫理的評估結果。為抑制和遏止違逆專業自覺的達共識評估結果變成現實，需要進行檢測，或者提供其他佐證資訊，能夠說明評估結果達共識是合理和有效的。可是觀察既有研究文獻罕見提出檢測前述評估人員產出不專業和不合理現象的資訊和證據，似有不足，有待思考因應做法。

至於可採取因應策略包括1. 慎選參與評估的人員，如事先探聽該人員關於討論事項的立場傾向、思考問題習慣、決策慣性和決策倫理、對專業認知等，避免邀請可能導致Delphi分析結果違反專業倫理者；2. 施測題項詢問評估人員關於參與Delphi分析認知和態度，如參與多輪回評估的熱誠、參與評估遭遇異議和爭執的態度、對討論事項的認同和立場；3. 詢問在兩輪回評估發生看法改變的評估人員，了解造成其前後看法和立場逆轉的原因和理由，推測是否有受到團體非議或輿論壓力、評估倦怠、評估沮喪等情形；4. 邀請其他專家組成諮詢團隊，針對各輪回實施結果和評者表現提出審查看法和建議意見。

對此，Kobus和Westner（2016, April）就提出確認專家專精領域、確認專家名單、增額提名其他專家、評等專家（rank experts）、邀請專家等五階段做法和運作示例，頗具參考價值。其中“增額提名其他專家”階段是將Delphi分析實施做法摘要資訊，主要包括擬邀請專家背景特性資訊（如專精領域），不包括明確擬邀專家名單以保持後續評估的中立和客觀，提供給列名邀請名單的專家，請其提名其他可邀請的專家，充實諮詢專家資料庫；而“評等專家”階段係根據擬邀名單的專家特性進行評等排序。

二、未考驗評定者專業、理解程度案特性，無法確保Delphi分析專業品質

Delphi分析結果主要依賴參與評估者的認定，若評估者不熟悉和了解待評事項，欠缺專業認知和相關經驗，又無法依據專業倫理進行評判，則最終結果難堪採用；觀察Schifano和Niederberger（2025）提出的五類影響Delphi分析因素，就包括個人回應行為一類。Skulmoski、Hartman和Krahn（2007）就彙整出選擇評估專家的要件，包括關於待討論議題的知識和經驗、參與能力（capacity）和意願、充足時間參與、有效溝通技巧、良好洞察（insight）議題等；另有論者Munier和Rondé（Hanafin, 2004）提出評者說明其評定項目的知識依據，據以分成極佳（very good）、良好（good）、有限（limited）和虛無（null）知識（knowledge），分別意指評者持有知識專精且契合項目主題、持有過去知識且依

賴同行見解、持有來自報章、論文或親身經驗所得知識、持有知識與項目專業和了解無關。因此可知需要了解和控制評估專家專業是有必要，除事先徵詢參與意願時進行了解，事後實際參與諮詢時也需要進行調查和了解。

對此，一些研究論述會報告有關評估者的專業素養調查結果，包括質性描述或呈現計量統計資訊做為判斷依據。檢視多數的臺灣地區研究者鮮少報告關於評估者專業認知、經驗和對事項看法的調查和分析資訊，包括專家權威程度值（Cr）、專家能力係數（coefficient of k-expert competence, k）（陳青山、王聲湧、董曉梅等，2004）、專業指數（P）（葉連祺，2018）等，僅概括性一語說明專家具專業知能，顯然有改善必要。

針對上述討論，綜合文獻彙整可採用以評估評定者專業的調查題項和評定做法見表6臚列。大致上，論者提出關於各調查焦點適用係數/指數的判斷標準有些顯得分歧，如回覆率 $\geq 50\%$ 、 $\geq 60\%$ 、 $\geq 70\%$ ；權威係數(Cr) > 0.5 、 > 0.7 。而少數研究論述還針對評估者的學經歷、職務、性別等基本特性資料，進行表6一些焦點項看法的差異考驗，如t考驗、ANOVA、 χ^2 考驗等，提供評估者條件對後續檢視看法共識性評估結果是否穩定的參考依據。

表6

評估評定者專業可用調查項目和評定做法

調查焦點	調查內容/項目	評定和計算方式	適用係數/指數
1.項目重要性 ^c	重要性,可行性,必要性,影響性,成效性,認同度,敏感性,理解度,成本性,便利性等	以1~5評定重要性程度,計算平均值或等級和	集中係數(M) ^a 協調係數(CV) ^a
2.判斷依據事項	理論依據,實踐依據,文獻成果,主觀直覺,同行共識	勾選各項目或自由陳述項目,合計勾選項數或合計各選項加權值(即選項加權數×勾選項數)	
3.判斷依據影響性		針對判斷依據事項,以1~3或其他評定值(如0.1~0.3)評定各判斷依據影響程度,計算合計值或平均值	判斷係數(Ca) ^b
4.對項目熟悉度		以1~5或其他評定值(如0.2,0.4,0.6,0.8,1)評定熟悉程度,計算平均值	熟悉係數(Cs) ^b
5.有關項目專業性	出版著作/報告,參加研習,擔任教學,擔任評選,獲獎/授證,從事實施,參與規劃/研究,專精領域,工作年限/職務年資,相關科系學歷	勾選各項目或自由陳述項目,合計勾選項數、或勾選項數占比率	專業指數(P)
6.參與積極性	回覆率,提出建議率	計算全體評定者的達成比率	積極係數(Cp),回覆率

註：^a集中係數是全體評定者的重要性評定值平均數(M)，協調係數指全體評定者的重要性評定值變異係數(CV)。^b權威係數(C_i)=(C₁+C₂)/2。^c通常調查對項目重要性的看法，或改採調查對可行性、敏感性等看法。

再者使用8位評估者專業性評估的實際調查資料（表7），進行評估者特性對熟悉係數、判斷係數，權威係數等影響考驗，結果見表8所示。其採用單因子ANOVA（one way ANOVA）協助考驗，顯示僅職務一項特性對評估者熟悉度產生影響，其他則無顯著差異，這表示後續分析共識度時不必考慮評估者特性的衝擊，提供有利的統計考驗佐證證據，可直接進行全體評估者評估項目達共識性檢視，可惜觀察大多數Delphi分析研究未考慮這項考驗，無法確保Delphi分析的專業品質，著實有待改善。

表7
 評估者特性、熟悉度、判斷依據、對項目重要性評估之實證資料示例

評者	熟悉度	判斷依據				特性						
		理論依據	實踐依據	同行共識	主觀直覺	判斷係數	權威係數	重要性	職位	職務	專業	學歷
1	0.6	0.1	0.5	0.05	0.05	0.7	0.65	4	2	1	1	1
2	0.8	0.3	0.5	0.05	0.05	0.9	0.85	4	1	2	5	2
3	0.6	0.2	0.4	0.1	0.1	0.8	0.70	3	1	1	5	1
4	0.8	0.3	0.2	0.05	0.05	0.6	0.70	5	1	1	5	1
5	0.6	0.2	0.4	0.1	0	0.7	0.65	4	2	1	4	2
6	0.8	0.3	0.4	0.1	0.05	0.85	0.83	4	1	3	3	1
7	0.6	0.2	0.4	0.1	0.05	0.75	0.68	3	1	1	2	2
8	0.8	0.2	0.4	0.1	0.05	0.75	0.78	3	2	3	9	1

註：資料取自某實證研究。權威係數=(熟悉度+判斷依據)/2，判斷係數=理論依據+實踐依據+同行共識+主觀直覺。

表8
 評估者特性對熟悉度、判斷依據、對項目重要性評估影響之考驗結果示例

	職位	職務	專業	學歷
熟悉度(C _s)	0.429 (0.067)	6.818* (0.532)	2.586 (0.301)	0.429 (0.067)
判斷係數(C _a)	1.1 (0.733)	0.775 (0.66)	0.172 (0.301)	0.35 (0.467)
1. 理論分析	2.188 (0.467)	0.438 (0.149)	1.119 (0.309)	0.313 (0.111)
2. 實際經驗	0.257 (0.093)	0.271 (0.098)	0.289(0.104)	0.257 (0.093)
3. 國內外同行	0.027 (0.004)	0.448 (0.07)	0.246 (0.039)	0.027 (0.004)
4. 直覺	1.016 (0.289)	0.539 (0.177)	0.041 (0.016)	1.016 (0.289)
權威係數(C _r)	0	0	3.289 (0.892)	1.1 (0.733)
項目重要性	0.313 (0.111)	0.212 (0.078)	0.605 (0.195)	0.313 (0.111)

註：括號外為ANOVA的F值及考驗p值，括號內為固定效果模式的 η^2 效果量。* $p<.05$

三、僅使用描述統計量數簡略評估共識性，未提供更嚴謹共識性評估資訊

檢視 Delphi 分析的成功關鍵在於提供嚴謹、周全和有用的共識性評估資訊，對此僅依賴質性的理性判斷實力有未逮，採用有效的統計分析技術提供較為客觀和嚴謹的共識評估資訊就有必要。一些論述彙總 Delphi 分析文獻採行做法（陳青山等，2004；Beiderbeck, *et al.*, 2021; Diamond, *et al.*, 2014; Grant, Kroger, & Bauman, 2023; Kalaian & Kasim, 2012; Khodyakov, Giannarou & Zervas, 2014; Naisola-Ruiter, 2022; Schifano & Niederberger, 2025; Von Der Gracht, 2012），大致包括描述統計、推論統計和分析指數見表 9 的簡略說明，其顯示每一輪回結果的分析焦點可置諸檢測共識性（consensus）和穩定性（stability），前者包括檢視多評估者對單項和多項看法的達共識程度，即分析焦點 1 和 3；後者檢視評估者在前後兩輪回之間對於評估項目看法變化情形的穩定程度，指分析焦點 2。

檢視臺灣博碩士論文加值系統、期刊文獻資訊網、臺灣人文及社會科學引文索引資料庫、及關於 Delphi 分析研究文獻的系統評論論述、Google Scholar Search 系統等，發現多數 Delphi 法應用研究最多檢視單項共識，如使用平均數、標準差、CV、滿分比、四分差 (Q) 等，以表示項目評分的集中性和離散性；僅少數注重檢視多項共識，如使用 Kappa 一致性係數、Kendall W 係數、Spearman 等級相關 (r_s)、Kruskal-Wallis 單因子等級 ANOVA (H 考驗)、Kolmogorov-Smirnov 單一樣本考驗；而關注單項共識變化者則非常罕見，如採用 McNemar 考驗、改變率。此表明當前應用 Delphi 分析存在資料分析角度狹隘的缺失一只關注單項共識查核，提供的共識性檢核證據顯得薄弱。

另外，也得見 Delphi 分析採用主成分分析 (PCA)、因素分析、項目階層集群分析 (hierarchical cluster analysis) (Kaufmann, 2016; Tapio, 2003)。再者，儘管論者指出針對 Delphi 分析，能使用 Fleiss Kappa 係數、Conger Kappa 係數、Cronbach α 、組內相關係數 (intra-class correlation coefficient, ICC)、Kendall W 係數等統計方法 (葉連祺，2018；Gisev, Bell, & Chen, 2013; Holey, Feeley, Dixon, & Whittaker, 2007)，及 De Moivre 指數 (De Moivre index, DM, DMI)、嚴格一致性指數 (Strict Agreement index, SA, SAI)、校正嚴格一致性指數 (Corrected Strict Agreement index, CSA, CSAI)、Light Kappa 係數、集群成對一致性係數 (Clustered Pairwise Agreement, CPWA)、成對一致性係數 (Pairwise Agreement, PWA)、偏差集群成對一致性係數 (extremities version of the Clustered Pairwise Agreement, XCPWA)、集群狀態係數 (Clustered Mode,

CM) 等 (Birko, Dove, & Özdemir, 2015; Meijering, Kampen, & Tobi, 2013)、主要意見均率決斷率 (Average Percent of Majority Opinions Cut-off Rate, APMO) (Cottam, Roe, & Challacombe, 2004), 但是臺灣地區研究論述卻罕見採用, 此可能是常用統計軟體 IBM SPSS Statistics 不提供 Fleiss Kappa 係數等分析功能、未廣泛知悉和深入了解可用分析係數和統計方法所致。

表9
 Delphi分析應用統計分析焦點比較

分析焦點	統計量數和分析方法		分析指數
1 單項共識	描述統計	平均數*, 中數*, 眾數*, Q1, Q3 等	標準差*, 變異係數(CV)*等
	推論統計	全距, 滿分比*, APMO等 平均數, 滿分比等95%CI和顯著性考驗, Kolmogorov-Smirnov單一樣本考驗 主成分分析, 因素分析	變異數, IQR*, Q*, IPRAS, DI 標準差, CV等95%CI和顯著性考驗, Kruskal-Wallis單因子等級ANOVA(H考驗) 項目集群分析
2 單項共識變化	描述統計	改變率	CVI, CVR
	推論統計	McNemar考驗, Bowker考驗, χ^2 考驗, Wilcoxon符號等級考驗等	Pearson r, Spearman rs等
3 多項共識	推論統計	Kappa一致性係數*, Kendall w*, Mann-Whitney U考驗等	ICC, Cochran Q等 Cronbach α , 主成分分析等 MCV, DM, SA, CSA

註：*標示被多數研究採用的統計量數。

關於表9在單項共識分析部分常採用一些檢核標準, 做為達共識與否的依據, 彙總文獻可見滿分比 (指評定者評分達預定分數的比率) 多以逾70%、75%、80%、90%為判斷規準; 其他如改變率 <15%或25%; 平均數達60%、75%、80%的評定量尺值 (如5點量尺值的3、3.75、4); 標準差 <0.8、1、1.5或2.5; 變異數 <0.5; IQR值 (IQR=Q₃-Q₁) <1、2.5、1.5、2、1.2; Q值 (係Q=(Q₃-Q₁)/2) <0.6、1; CV值 <0.25、<0.2、<0.3、<0.4、<0.5; 中數 >70%或75%的最大評定量尺值; 另有計算眾數和平均數絕對差, 即|Mo-M|≤1, 但罕見應用。分析指數部分採用標準如CVI值 >0.8、>0.78; CDI值 <0.1、0.2、0.05、0.5、0.3; DC值 >0.8。以及Cronbach α 值 >0.8; Kappa係數值 >0.75; Pearson r值 >0.7; Kendall w值 <.05。值得注意的是, 這些被採用的標準值頗多是提出者的直覺經驗談, 未必有堅實的學理或實證證據支持, 所以有些標準設定值顯得分歧, 是故儘管被採用, 仍有可質疑和討論之處。

以一篇臺灣地區已發表研究論述所刊載實證資料為例 (表10), 其僅說明22項的平均數(M)、標準差(SD)和IQR等統計資訊, 根據M>4、滿分比>50%、SD<1和CV<0.25為判斷標準, 可知項目6、8、11和13表現較差, 僅項目11的CV值不佳, 表明共識度差。若進一步分析平均數和CV的95%信賴區間(CI) (表

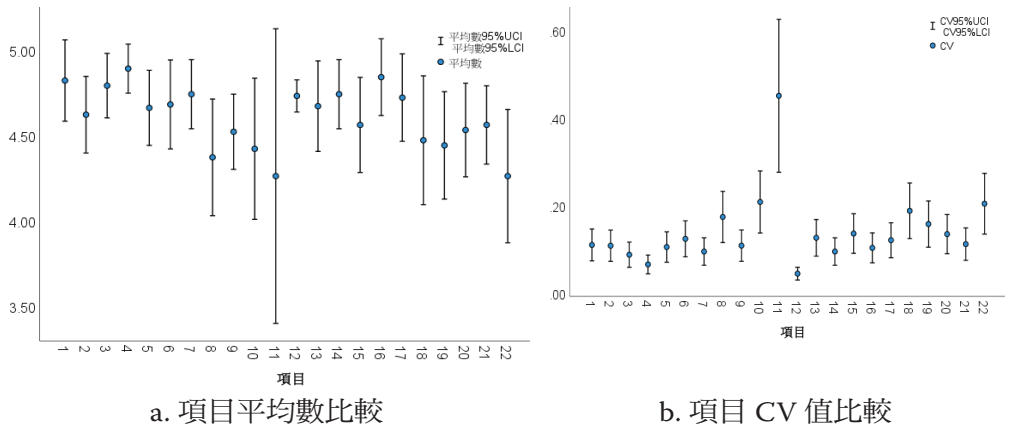
10)，圖3明顯指出專家對項目11的評分歧異性高，其平均數和CV的95%信賴區間甚大，CV比平均數更清楚地表明項目11的共識度明顯低於其他項目，其95%CI不與其他項交疊，即比較CV的95%CI區間是否交疊為檢視項目共識性的頗佳做法，可惜很少Delphi分析揭露此資訊。

表10
Delphi分析實證資料之描述統計結果

項目	平均數	95%LCI	95%UCI	標準差	IQR	CV	95%LCI	95%UCI	滿分比	評估
1	4.820	4.582	5.058	0.530	0	0.110	0.074	0.146	52.632	
2	4.620	4.395	4.845	0.500	0	0.108	0.072	0.144	68.421	
3	4.790	4.601	4.979	0.420	0	0.088	0.059	0.117	68.421	
4	4.890	4.746	5.034	0.320	0	0.065	0.044	0.087	52.632	
5	4.660	4.440	4.880	0.490	0.5	0.105	0.070	0.140	73.684	
6	4.680	4.419	4.941	0.580	0.5	0.124	0.083	0.165	47.368	滿分比差
7	4.740	4.538	4.942	0.450	0.5	0.095	0.064	0.126	78.947	
8	4.370	4.028	4.712	0.760	0.5	0.174	0.115	0.232	26.316	滿分比差
9	4.520	4.300	4.740	0.490	0.5	0.108	0.073	0.144	63.158	
10	4.420	4.006	4.834	0.920	0.5	0.208	0.137	0.279	63.158	
11	4.260	3.397	5.123	1.920	0.5	0.451	0.276	0.625	57.895	CV,SD差
12	4.730	4.636	4.824	0.210	0.5	0.044	0.030	0.059	68.421	
13	4.670	4.405	4.935	0.590	0.5	0.126	0.084	0.168	47.368	滿分比差
14	4.740	4.538	4.942	0.450	5	0.095	0.064	0.126	68.421	
15	4.560	4.281	4.839	0.620	5	0.136	0.091	0.181	63.158	
16	4.840	4.615	5.065	0.500	0	0.103	0.069	0.137	63.158	
17	4.720	4.464	4.976	0.570	0.13	0.121	0.081	0.161	63.158	
18	4.470	4.092	4.848	0.840	0.5	0.188	0.124	0.251	89.474	
19	4.440	4.125	4.755	0.700	0.5	0.158	0.105	0.210	63.158	
20	4.530	4.256	4.804	0.610	0.5	0.135	0.090	0.179	84.211	
21	4.560	4.331	4.789	0.510	0.5	0.112	0.075	0.149	57.895	
22	4.260	3.869	4.651	0.870	0.5	0.204	0.135	0.274	78.947	

註：N=19。平均數、標準差和IQR值取自某實證研究，其餘係本研究計算所得。

圖3
Delphi分析實證資料之項目平均數和CV值比較



四、使用Delphi分析共識性評估結果做為指標信度和效度考驗證據，忽視共識性評估功能

Hasson和Keeney（2011）指出為確保嚴謹的Delphi分析成果，需要注意和思考提出信度（reliability）、效度（validity）和信賴度（trustworthiness）資訊。觀諸一些研究（尤其是臺灣地區學位論文）採用此方法發展評估指標（indicators），如研議應用於課程評估、領導能力評鑑、政策評估等用途的指標，通常就將通過單項共識檢視（如採用平均數、標準差）的結果視為可用的指標。這個做法忽視了Delphi分析使用CV、CDI等協助檢視共識性，若通過只是表示全體評估者對單項看法達共識，若再通過Kendall w 考驗，也是表示對多個項目看法達共識，並不意味著可據此看出關於信度和效度的訊息。因此有極少數Delphi分析研究是進行了Cronbach α 分析，沒有一併做因素分析或主成分分析以確認項目的構念信度；所以對於發展指標的研究而言，仍有提出效度證據不足的問題，有待改進。

使用實證資料進行Fleiss Kappa一致性考驗（ κ ），顯示 $\kappa=0.151$ ， $p=.328$ ，評估者對四項目的看法達共識，可是這不表示全部項目評定看法具高信度和高構念效度。故再進行Cronbach α 和主成分分析，呈現於表11。其說明CV和CDI值頗小，似乎顯示評估者對各項目看法達共識，但此不能夠說明四項目具有頗佳的內部一致性信度和構念效度。觀察整體Cronbach $\alpha=.519$ ，標準化Cronbach $\alpha=.563$ ， $\text{theta}=.598$ ，依據Schober、Mascha和Vetter（2021）所臚列的Cronbach α 值評估標準： $<.6$ 為不佳（poor）， $\geq.6$ 且 $<.7$ 為須質疑（questionable）， $\geq.7$ 且 $<.8$ 是可接受（acceptable）， $\geq.8$ 且 $<.9$ 是良好（good）， $\geq.9$ 為極佳（very good），顯然是內部一致性信度不佳。檢視調整題項和總分相關（ r_{itc} ）值和刪項後Cronbach α 值，表11指出項目4表現不佳，其 $r_{itc}=.158$ 最小，刪項 α 是 $0.646 >$ 整體Cronbach α 的 0.519 ，也是說明項目4反映的評估資訊異於其他項。而主成分分析結果（表11）指出根據K1法則，宜選取2個主成分，解釋變異量%達77.379%甚佳，依據負荷量值判斷項目歸屬，可見項目1~3應歸屬成分1，項目4歸屬成分2，顯然四項目可精簡為2個向度，這與Cronbach α 分析結果透露訊息相同。據此得知，不能僅依賴Kappa係數等整體項目共識性評估結果，做為判斷項目具高信度和高構念效度的證據，應該再確實地進行Cronbach α 、主成分分析（或因素分析）。

表11

Delphi分析實證評估資料Cronbach α 和主成分分析之分析結果

項目	M	SD	CV	CDI	DC	調整rit	刪項	主成分	特徵值	解釋量%	負荷量	
											成分1	成分2
1	4	0.853	0.213	0.171	0.829	.255	0.488	1	1.813	45.316	.7947	-.4435
2	4.083	0.515	0.126	0.103	0.897	.475	0.373	2	1.283	32.064	.8452	-.2422
3	3.583	0.9	0.251	0.180	0.820	.459	0.412	3	0.560	14.004	.6364	.5407
4	3.667	0.888	0.242	0.178	0.822	.158	0.646	4	0.345	8.617	.2485	.8573

註：項目1~項目4資料取自某學位論文實證研究。N=12。整體Cronbach α =0.519。

五、簡略介紹CV造成誤解和誤用，缺乏完整說明和有效應用

CV是適用於檢視評估看法達共識與否的有效分析係數，也被Delphi分析量化研究所採用。常見一般中文統計教科書介紹CV，CV或稱為相對差異係數（coefficient of relative variability, CRV）、正規化RMSD（normalized root-mean-square deviation, NRMSD）、RMS比率（percent of root-mean-square deviation, percent RMS）、關聯標準差（relative standard deviation, RSD），但對照CV的研究發展成果而言，實屬簡略；檢視一般研究應用CV的論述也屬稀少，不如平均數和標準差已成為量化研究必須報告的資訊，至於論者提出的CV衍生指數如團體穩定指數（Group Stability, GS）亦有可議之處，該值有出現負值的可能性，將造成解讀數值意義時困擾。又一般統計教科書登載的是計算樣本CV（sample coefficient of variation），而母群體CV（population coefficient of variation）估計公式為 $\widehat{CV} = \left(1 + \frac{1}{4N}\right) CV$ ，N是樣本數，標準誤為 $\widehat{SE}_{CV} = \frac{CV}{\sqrt{2N}}$ ，樣本CV的標準誤為 $SE_{CV} = \frac{CV}{\sqrt{2N}}$ （Abdi, 2010）。再者研究論述中提及的CV有多個計算公式，包括 $CV = \frac{SD}{M}$ 、 $CV = \frac{\sum|x-\bar{x}|/N}{M}$ 、 $CV = \frac{\sum|x-Md|/N}{Md}$ 、 $RCV = 0.75 \times \frac{IQR}{M}$ 、 $RCV_M = 1.4826 \times \frac{MAD}{M}$ 等（Arachchige, Prendergast, & Staudte, 2022; Martin & Gray, 1971），在教育領域Delphi分析較常應用 $CV = \frac{SD}{M}$ ，其他則少見採用。

以下討論三個有關於CV的問題：

問題 1: 誤解CV是說明標準差值佔平均數值的百分比

若干統計教科書根據CV計算公式 $CV = \frac{SD}{M} \times 100$ ，指出該CV值表示標準差值佔平均數值的百分比，這個說法是有盲點，其或稱百分比變異係數（percentage CV, %CV）（Hart, 2012）見表12說明；另一說法是CV值普遍偏小，乘以100是將CV值變大，便利於辨識，與百分比%無關，例如將CV值以50.34%形式呈現和運用。觀察CV的另一個計算公式 $CV = \frac{SD}{M}$ ，該數值直接以小數形式呈現，如1.23或0.56以表明CV值是標準差值和平均數值的比率，對此可知其計算標準差值是平均數值的倍率，供判讀標準差值大小的意義。而Martin和

Gray (1971)、Abdi (2010) 也都指出%CV有誤導之嫌，因為%CV值可能大於100，應將該值視為比率 (ratio) 而非百分比 (percentage)。

為區別起見，此處將前者稱為 CV_{100} ，後者仍稱為CV。 CV_{100} 的使用價值除了將SD/M值放大之外，通過數值代表的百分比意義，可從百分數制角度便利了解CV值的優劣，越接近100越佳，愈趨近0則顯然愈差。細思此論點似乎是奠基於 CV_{100} 的值域是[0,100]的假設，可是CV值可能大於1，屆時當 CV_{100} 值>100，那麼就容易產生解讀該值優劣的困難，有賴借助 CV_{100} 的上限值做為比較基礎。

根據推導 (見附錄二說明)，CV值的上限為 $\sqrt{N-1}$ ，此已被Martin和Gray談及，N是樣本數，所以 CV_{100} 值的上限就是 $100\sqrt{N-1}$ ，得使用此上限值來輔助判讀 CV_{100} 值的優劣與否，這明顯比直接採用CV值來得麻煩。據此觀之，採用 $CV = \frac{SD}{M} \times 100$ 並未見明顯優勢，反而可能造成誤解，顯見將SD/M值乘以100，就是放大CV值以便於檢視而已，並不具有表示百分比的意義，故不建議使用。

另外 CV_{100} 也被應用衍生出團體穩定係數 (GS)，可見Surveylet軟體 (https://calibrum.net/surveylethelp/consensus_calculation_example___radio_button_combobox.htm)， $GS = 100 - \frac{SD}{M} \times 100$ ，若 $CV_{100}=35.12$ ， $GS=100-35.12=64.88$ ，表示團體共識度達64.88%，而Dodd (1952) 稱此為百分比平均數 (percentage mean, %M)；依據前面討論，GS值的計算和對分析值的解釋著實存在問題，當GS值<0時難以解釋，是故也不建議採用。

前述討論是就學理而言，進一步採用permutation法產生模擬資料，了解實際CV值的值域可能為何？設計程式 (見附錄十一SIMULATECV.sps說明)，協助模擬分析CV值。分析步驟為：1. 設定每批次模擬分析的條件包括量尺值、樣本數、項目數和模擬次數。2. 採用均等分配，依據條件產生模擬資料。3. 針對每批次資料分析平均數、標準差和CV值，並記錄出現的最小CV值、最大CV值、最小平均CV值、及最大平均CV值。4. 視需要儲存模擬資料和繪製CV值分布圖等。

表12和圖4說明在評定量尺值為5和7，樣本數是3、5、10、20和50時的CV值分布資訊摘要。整體觀之，可見幾項發現，並做簡單討論：

1. 可能出現CV值=0和CV值>1的情形。
2. 樣本數與CV值=0有關。可知樣本數小如3或5人時，容易出現CV值最小值為0的情形，樣本數越大則此情形會改善，如樣本數 ≥ 7 人時，這與量尺值無太大關聯。
3. 樣本數與CV值>1有關。可知樣本數偏小如 ≤ 5 時，易出現CV值>1情形，樣本數越大將改善此情形，如樣本數 ≥ 7 人時，這與量尺值關聯性不大。

4. 隨者樣本數增加和採用較大量尺值，影響產出CV值的值域大小。可看出顯示樣本數越大，最小和最大CV值出現的值域越趨縮小，如評定值=5，N=3時是[0,0.9897]，N=5時是[0.0932,0.9938]，而N=50時變為[0.3225, 0.6222]，顯然值域由寬收窄，若評定值=7時也是呈現類似變化情形。相同地，檢視每批次CV值平均的最大值和最小值之間差距變化亦形成隨著樣本數增加，而逐漸縮小。

表12

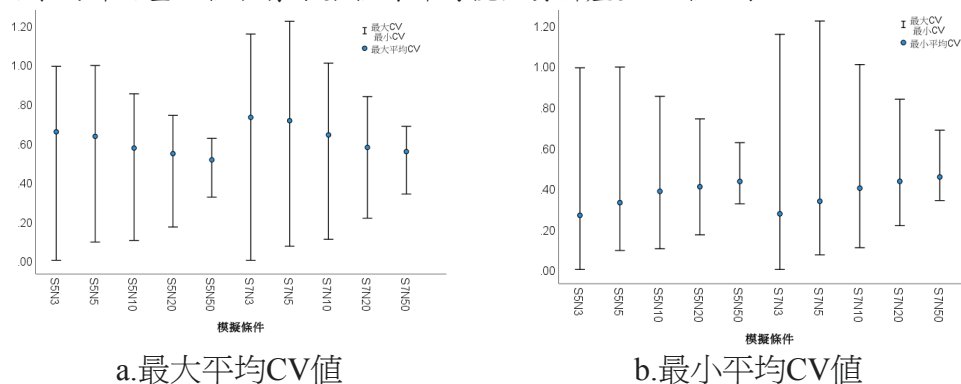
設定不同分析條件產生模擬資料分析CV值之比較

樣本數		量尺值5				量尺值7			
		最小值	最大值	最小M	最大M	最小值	最大值	最小M	最大M
3	最小	0.0000	0.9897	0.2658	0.5857	0.0000	1.1547	0.2734	0.6444
	最大	0.0000	0.9897	0.3261	0.6557	0.0000	1.1547	0.3522	0.7294
5	最小	0.0932	0.8660	0.3281	0.5646	0.0721	1.0030	0.3350	0.6040
	最大	0.0932	0.9938	0.3660	0.6326	0.1014	1.2197	0.3998	0.7125
10	最小	0.1020	0.7658	0.3836	0.5337	0.1071	0.8437	0.3992	0.5297
	最大	0.1942	0.8498	0.4068	0.5730	0.1991	1.0063	0.4400	0.6400
20	最小	0.1704	0.6780	0.4062	0.5140	0.2152	0.7167	0.4327	0.5425
	最大	0.2857	0.7393	0.4312	0.5444	0.2907	0.8357	0.4487	0.5763
50	最小	0.3225	0.5864	0.4321	0.4986	0.3384	0.6283	0.4541	0.5269
	最大	0.3638	0.6222	0.4493	0.5127	0.3835	0.6837	0.4702	0.5548

註：採取均勻分配設定和Permutation法進行10批每批次100次模擬產生10個項目資料，再計算CV值進行彙總。最小值指最小CV值，最大值指最大CV值，最小M值指最小的平均CV值，最大M值指最大的平均CV值。

圖4

設定不同評定量尺值和樣本數值條件時模擬資料產生CV值比較



註：針對Permutation法模擬資料所得CV值彙總分析結果。模擬條件的S5指評定量尺值為5，N3指樣本數為3，餘依此類推。

縱觀上述針對模擬資料的實證分析，可知CV值受樣本數和採用評定量尺值有關，過去Martin和Gray (1971)、Smithson (1982)、Abdi (2010) 研究就談到類似現象，或進一步構思不受樣本數影響的改進CV指數如CV''，但仍存在

計算問題有待克服，值得後續致力深究。其次，會出現CV值>1情形，故GS係數、CV100係數皆不宜使用。而樣本數變大，CV值域隨之變小，若CV值表示樣本看法差異程度，那麼進行Delphi分析宜邀請較多的評定者參與，考量實施成本，宜至少採用5點評定量尺，邀請至少5位評定專家參與，能避免出現CV值>1的情形。

問題2：誤解CV是描述統計方法

若干統計教科書指出CV是描述統計方法，不能做推論統計的考驗，這也是誤會。檢視文獻就可察知統計學者已發展CV值的顯著性考驗、分析95%信賴區間、考驗多個CV值均等、考驗母群體CV值等方法（Abdi, 2010; Arachchige, Prendergast, & Staudte, 2022; Banik, Kibria, & Sharma, 2012; Forkman, 2009; Fung & Tsang, 1998; Kalkur & Rao, 2015; Krishnamoorthy & Lee, 2014; Shoukri, 2011; Smithson, 1982; Thangjai, Niwitpong, & Niwitpong, 2021），表13彙總一些考驗方法供參考，其中正規化變異係數（normalized coefficient of variation, scaled coefficient of variation, CVs）納入考慮了CV的最大值，Bedeian和Mossholder（2000）稱之為CV(n)，係將CV值規範於[0,1]，便利於判讀CV值的優劣，很值得採用。Kvålseth（2017）構思出二階變異係數（second-order coefficient of variation, CV₂），該係數值≥0且<1，數值越小越好，表示資料值間差異越小，頗便利於直觀理解和闡釋，該係數的最大值上限和最小值下限討論見附錄四。又Smithson（1982）提出跨樣本變異係數（CV''），宣稱可使CV值不受樣本數的影響，但是如何設定適當的Q值頗為困難，故應用該係數者不多。而Quan和Shih（1996）提出樣本內變異係數（within-subject coefficient of variation, WCV, WSCV），以評估測量的重製性（reproducibility），指在相同條件下，以相同樣本使用相同測量工具或不同場合進行重複測量的近似程度（degree of closeness），爾後Shoukri、Colak、Kaya和Donner（2008）補充提出顯著性考驗、信賴區間估計、多個WSCV值差異比較考驗等做法，但是教育領域罕用此係數。又Kalaian和Kasim（2012）提出變異係數差異（CV_{diff}），係估算兩輪回評估的兩個CV值差異，以絕對值表示前後輪回的CV值變化程度，CV_{diff}值越接近0表示評估者看法變化越小，達共識性情形越好。Sharma提出CVS（Sharma's coefficient of variation）（Tovohery, André, & Rajaonasy, 2022）其以最大值和最小值與平均數差距為比較基礎，異於CV採用平均數。另見反變異係數（inverse coefficient of variation, inverse CV, ICV）， $ICV = \frac{M}{SD}$ ，係計算CV的倒數（Sharma & Krishna, 1994），並有論述（Albatineh, Kibria, & Zogheib, 2014; Kalkur & Rao, 2015; Potas & Gamgam, 2019）探討其信賴區間估計、顯著性考驗等做法，及分析

值的分配型態，而Pélabon、Armbruster和Hansen（2011）也提出變異係數平方係數（square of CV, CV^2 ），其值等於基於平均數的正規化變異數（mean-scaled variances）， $CV^2 = \frac{S^2}{M^2}$ ， S^2 是變異數，可用於比較不同樣本數時的變異數，惟教育領域似乎未見應用這些係數。

再者，表14彙總考驗CV值95%信賴區間的方法，但未包括採取Bayesian法理念計算的信賴區間方法，其可見Thangjai、Niwitpong和Niwitpong（2021）論著的說明，表15說明考驗母群體CV值方法，表16臚列考驗二或多個CV值差異的做法供參考。

問題3：對於評估CV值標準存在分歧看法

如何辨識CV值的適切程度，已見Pearson、English和Keran、Costa等人提出判斷分類經驗值，如表17所示。至於應用實務，則見採用CV值 <0.2 、 <0.25 、 <0.3 、 <0.4 、 <0.5 等多種評估標準。不難瞧出儘管看法分歧，但是CV值應 <0.5 似乎是一個可接受的觀點。

又前述觀點是採取固定標準的觀點，即選取一個固定值為標準，以適用於全部情形，Lorenzo等（2015）提出一個變動標準的做法，是以該次分析中最大和最小CV值的差為基礎，平均分成三類，計算公式為 $CV_{width} = (CV_{max} - CV_{min}) / 3$ ， $CV_{LL} = CV_{min} + CV_{width}$ ， $CV_{RL} = CV_{max} - CV_{width}$ 低度（low）是 $CV \leq CV_{LL}$ ，中度（medium）是 $CV_{LL} < CV \leq CV_{RL}$ ，高度（high）是 $CV_{RL} < CV$ 。另外Costa等人也提出結合使用中數、四分差（Q）和四分位距（IQR）構成的四分類觀點（Costa, Seraphin, & Zimmermann, 2002; Couto, Peternelli, & Barbosa, 2013），Garcia則納入CV值的平均數和標準差，亦形成四分類（Ferreira *et al.*, 2016），Pimentel和Gomes採取0.1、0.2和0.3為分類依據，構成四分類看法（Lopes *et al.*, 2021），或者可採取 $M \pm C \times SD$ ，設定C值，以取得明確的分類標準，以上所提觀點的計算公式見表17臚列。若以 $C=1$ ，則分類標準為 $M \pm SD$ ，形成三分類，即表17的Yeh三分類做法，因為三分類做法的中間分類值域極大，可納入平均數做為切割點，形成Yeh四分類做法。大抵上，採用固定評估標準觀點可謂簡易，但是學理依據偏弱，而變動標準則計算較繁瑣，多有資料分配型態的統計學理依據，這兩類點各具優劣，有待進一步檢驗其分類效果。

又需要注意，Pearson（1932）、English和Keran（1976）這兩者的CV值分類命名是根據CV值顯示意義，即CV值越大越差，表示項目值彼此差異越大；反觀，Pimentel和Gomes（2009）等其他的CV值分類觀點是基於CV值大小，命名反映CV值的相對大小，所以其CV值極大就將分類命名為“極高”，但是其實該分

表13

CV及有關係統計量數

統計量數	分析公式	顯著考驗	95%信賴區間
變異係數 CV, V	樣本 CV $CV = \frac{SD}{M} = \frac{\sqrt{\frac{1}{N} \sum (x_i - \bar{x})^2}}{\frac{1}{N} \sum x_i}$	H ₀ : CV=0 $t_{CV} = \frac{\sqrt{N}}{CV}$ $t_{CV} \sim t_{\alpha, N-1}$	$CV_{LCI} = CV - z_{1-\alpha/2} \sqrt{\frac{CV^2(1+2CV^2)}{2(N-1)}}$ $CV_{UCI} = CV + z_{1-\alpha/2} \sqrt{\frac{CV^2(1+2CV^2)}{2(N-1)}}$
有偏差母群體 CV	$CV_{bias} = \frac{\sqrt{\frac{1}{N-1} \sum (x_i - \bar{x})^2}}{\frac{1}{N} \sum x_i}$		
母群體變異係數, CV _{pop}	無偏差母群體 CV $CV_{unbias} = \left(1 + \frac{1}{4N}\right) CV$	H ₀ : CV=CV _{pop} =0 $SE_{CV} = \frac{CV}{\sqrt{2N}}$ $t_{CV} = \frac{CV}{SE_{CV}}$ $t_{CV} \sim t_{\alpha, N-1}$	$CV_{LCI} = CV - t_{\alpha, N-1} \times \frac{CV}{\sqrt{2N}}$ $CV_{UCI} = CV + t_{\alpha, N-1} \times \frac{CV}{\sqrt{2N}}$
變異係數, CV	$CV = \frac{SD}{M} \times 100$		
百分比變異係數, %CV, CV(%), OCV	$\%CV = \frac{SD}{M} \times 100\%$		
變異係數, CV _{QME}	$CV_{QME} = 100 \times \frac{\sqrt{QME}}{M}$		
正規化變異係數, CV _s , CV(n)	$CV_s = \frac{CV}{\sqrt{N-1}}$		
跨樣本變異係數, CV _{CG} , CV''	$CV_{CG} = \sqrt{\frac{\sum x_i^2 - N\bar{x}^2}{\sum Q_i^2 - N\bar{x}^2}}$		
百分比平均數 (%M), 團體穩定係數 (GS)	$\%M = 100 - 100 \frac{SD}{M}$ $GS = 100 - \frac{SD}{M} \times 100$		
二階變異係數, CV ₂ , V ₂	$CV_2 = \sqrt{\frac{CV^2}{1+CV^2}} = \sqrt{\frac{SD^2}{SD^2+M^2}}$		
變異係數差異, CV _{diff}	$CV_{diff} = CV_{R1} - CV_{R2} $		
反變異係數, ICV, InvCV	$ICV = \frac{1}{CV} = \frac{M}{SD} = \frac{\frac{1}{N} \sum x_i}{\sqrt{\frac{1}{N-1} \sum (x_i - \bar{x})^2}}$		$ICV_{LCI} = ICV - z_{\alpha/2} \frac{1}{\sqrt{N}}$ $ICV_{UCI} = ICV + z_{\alpha/2} \frac{1}{\sqrt{N}}$
Sharma's CVS	$CVS = \frac{SD}{\sqrt{(\max(x) - \bar{x})(\bar{x} - \min(x))}}$		
變異係數平方, CV ²	$CV^2 = \frac{S^2}{M^2}$		

註：N是樣本數，M是平均數，S²是變異數，SD是標準差，x是原始評定資料。z_{Miller}、t_{KOO}、χ²_{MC}指其分析值近似z、t、χ²分配，並標示提出者做為區別。CV_{LCI}指CV的95%信賴區間下界，CV_{UCI}指CV的95%信賴區間上界，Q是設定會產出最大變異的數值。CV_{R1}和CV_{R2}表示兩個輪回的CV值。通常設定α=.05、.01或.10。

表14
CV信賴區間考驗方

提出者	信賴區間下界(LCI)	信賴區間上界(UCI)
Miller(1991), Miller & Feltz (1997)	$CV = \frac{SD}{M} = \frac{\sqrt{\sum(x_i - \bar{x})^2}}{\bar{x}}$ $LCI = CV - z_{\alpha/2} \sqrt{\frac{CV^2}{N-1}}(0.5 + CV^2)$	$UCI = CV + z_{\alpha/2} \sqrt{\frac{CV^2}{M}}(0.5 + CV^2)$
Gulhar, Kibria, Albatineh, & Ahmed (2012)	$CV = \frac{SD}{M} = \frac{\sqrt{\sum(x_i - \bar{x})^2}}{\bar{x}}$ $LCI = CV - z_{\alpha/2} \sqrt{\frac{CV^2}{N-1}}(0.5 + CV^2)$	$UCI = CV + z_{\alpha/2} \sqrt{\frac{CV^2}{M}}(0.5 + CV^2)$
Miller & Feltz (1997)	$LCI = \frac{CV}{1 + z_{1-\alpha/2} \sqrt{\frac{1+2CV^2}{2(N-1)}}}$	$UCI = \frac{CV}{1 - z_{1-\alpha/2} \sqrt{\frac{1+2CV^2}{2(N-1)}}}$
McKay (1932) Shao (2005)	$u_1 = \chi_{N-1, 1-\alpha/2}^2$ $LCI = \frac{CV}{\sqrt{\frac{u_1}{N-1} + CV^2} \left(\frac{u_1}{N} - 1\right)}$ $LCI = CV \sqrt{\frac{u_1}{N-1} + CV^2} \left(\frac{u_1}{N} - 1\right)$	$u_2 = \chi_{N-1, \alpha/2}^2$ $UCI = \frac{CV}{\sqrt{\frac{u_2}{N-1} + CV^2} \left(\frac{u_2}{N} - 1\right)}$ $UCI = CV \sqrt{\frac{u_2}{N-1} + CV^2} \left(\frac{u_2}{N} - 1\right)$
Vangl(1996)	$u_1 = \chi_{N-1, 1-\alpha/2}^2$ $LCI = \frac{CV}{\sqrt{\frac{u_1}{N-1} + CV^2} \left(\frac{u_1 + 2}{N} - 1\right)}$	$u_2 = \chi_{N-1, \alpha/2}^2$ $UCI = \frac{CV}{\sqrt{\frac{u_2}{N-1} + CV^2} \left(\frac{u_2 + 2}{N} - 1\right)}$
Gulhar, Kibria, Albatineh, & Ahmed (2012)	$u_1 = \chi_{N-1, 1-\alpha/2}^2$ $LCI = CV \sqrt{\frac{u_1}{N-1} + CV^2} \left(\frac{u_1 + 2}{N} - 1\right)$	$u_2 = \chi_{N-1, \alpha/2}^2$ $UCI = CV \sqrt{\frac{u_2}{N-1} + CV^2} \left(\frac{u_2 + 2}{N} - 1\right)$
Forkman(2005)	$LCI = \frac{\sqrt{(N-1)SD_{\log X}^2}}{\chi_{N-1, 1-\alpha/2}^2}$	$UCI = \frac{\sqrt{(N-1)SD_{\log X}^2}}{\chi_{N-1, \alpha/2}^2}$
Gulhar, Kibria, Albatineh, & Ahmed (2012)	$LCI = \frac{\sqrt{(N-1)CV^2}}{\chi_{N-1, 1-\alpha/2}^2}$	$UCI = \frac{\sqrt{(N-1)CV^2}}{\chi_{N-1, \alpha/2}^2}$
Forkman(2005) Panichkitkosolkul (2007)	$\tilde{k} = \frac{\sqrt{\sum(x_i - \bar{x})^2}}{\sqrt{N} \times \bar{x}}$ $LCI = \tilde{k} \sqrt{\left(\frac{\chi_{N-1, 1-\alpha/2}^2 + 2}{N} - 1\right) \tilde{k}^2 + \frac{\chi_{N-1, 1-\alpha/2}^2}{N-1}}$	$UCI = \tilde{k} \sqrt{\left(\frac{\chi_{N-1, \alpha/2}^2 + 2}{N} - 1\right) \tilde{k}^2 + \frac{\chi_{N-1, \alpha/2}^2}{N-1}}$
Sharma & Krishna(1994)	$CV^{-1} = \frac{1}{CV} = \frac{\bar{x}}{SD}$ $LCI = \frac{1}{CV^{-1} - \frac{1}{\sqrt{N} \times Z_{\alpha/2}}}$	$UCI = \frac{1}{CV^{-1} + \frac{1}{\sqrt{N} \times Z_{\alpha/2}}}$
Arachchige, Prendergast, & Staudte (2022)	$LCI = \frac{1}{\frac{1}{CV} + Z_{1-\alpha/2} \frac{1}{\sqrt{N}}}$	$UCI = \frac{1}{\frac{1}{CV} - Z_{1-\alpha/2} \frac{1}{\sqrt{N}}}$

註：N是樣本數，B是Bootstrap模擬次數，CV*是Bootstrap模擬所得CV值。通常設定 $\alpha=0.05$ ，或設定為0.01、0.1。

提出者	信賴區間下界(LCI)	信賴區間上界(UCI)
Curto & Pinto(2009)	$LCI = CV - Z_{\alpha/2} \sqrt{\frac{1}{N} \left(CV^4 + \frac{CV^2}{2} \right)}$	$UCI = CV + Z_{\alpha/2} \sqrt{\frac{1}{N} \left(CV^4 + \frac{CV^2}{2} \right)}$
非參數 bootstrap 法(NP bootstrap) Gulhar, Kibria, Albatineh, & Ahmed (2012)	<p>$LCI = \widehat{CV}_{(\alpha/2)B}$,B 是模擬次數 若 $\alpha=.05$, $CV_{\alpha/2}$ 是 P25 的模擬值</p>	<p>$UCI = \widehat{CV}_{(1-\alpha/2)B}$ 若 $\alpha=.05$, $CV_{1-\alpha/2}$ 是 P97.5 的模擬值</p>
Bootstrap t 法 Gulhar, Kibria, Albatineh, & Ahmed (2012)	<p>$LCI = CV + T_{\alpha/2}^* \times SD_{CV}$</p> $T_i^* = \frac{CV_i^* - \overline{CV}}{\widehat{\sigma}_{CV}}$ $\widehat{\sigma}_{CV} = \sqrt{\frac{1}{B} \sum_{i=1}^B (CV_i^* - \overline{CV})^2}$ $\overline{CV} = \frac{1}{B} \sum_{i=1}^B CV_i^*$ <p>若 $\alpha=.05$, $T_{\alpha/2}^*$ 是 P25 的模擬值</p>	<p>$UCI = CV + T_{1-\alpha/2}^* \times SD_{CV}$ 若 $\alpha=.05$, $T_{1-\alpha/2}^*$ 是 P97.5 的模擬值</p>
Mahmoudvand & Hassani (2009)	$C_N = \sqrt{\frac{2}{N-1}} \times \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}$ $LCI = \frac{CV}{2 - C_N + Z_{1-\alpha/2} \sqrt{1 - C_N^2}}$	$UCI = \frac{CV}{2 - C_N - Z_{1-\alpha/2} \sqrt{1 - C_N^2}}$
Mahmoudvand & Hassani (2009)	$\widetilde{CV} = \frac{CV}{2 - C_N}$ $C_N = \sqrt{\frac{2}{N-1}} \times \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}$ $LCI = \widetilde{CV} - \frac{\widetilde{CV}}{2 - C_N} \times Z_{1-\alpha/2} \sqrt{1 - C_N^2 + \frac{\widetilde{CV}^2}{N}}$	$UCI = \widetilde{CV} + \frac{\widetilde{CV}}{2 - C_N} \times Z_{1-\alpha/2} \sqrt{1 - C_N^2 + \frac{\widetilde{CV}^2}{N}}$
MOVER 取向 Donner & Zou (2012)	<p>$u_1 = \chi_{N-1,1-\alpha/2}^2$</p> $LCI = \frac{SD}{d} \left(\bar{x} \sqrt{\max(0, \bar{x}^2 + ad(a-2))} \right)$ $a = \sqrt{\frac{N-1}{u_1}}$ $d = \bar{x}^2 - Z_{\alpha/2} \times \frac{SD^2}{N}$	<p>$u_2 = \chi_{N-1,\alpha/2}^2$</p> $UCI = \frac{SD}{d} \left(\bar{x} \sqrt{\max(0, \bar{x}^2 + bd(b-2))} \right)$ $b = \sqrt{\frac{N-1}{u_2}}$
GCI 取向 Donner & Zou (2012)	<p>$LCI = R_{\theta(\alpha/2)}$ 若 $\alpha=.05$, $R_{\theta(\alpha/2)}$ 是 P25 的 R_{θ} 資料值</p> $R_{\theta} = \frac{\sqrt{R_{\sigma^2}}}{R_{\mu}}$	<p>$UCI = R_{\theta(1-\alpha/2)}$ 若 $\alpha=.05$, $R_{\theta(1-\alpha/2)}$ 是 P97.5 的 R_{θ} 資料值</p>

提出者	信賴區間下界(LCI)	信賴區間上界(UCI)
	$R_{\sigma^2} = \frac{(N-1)SD^2}{\chi_{N-1, 1-\alpha/2}^2}$	
	$R_{\mu} = \bar{x} - Z \sqrt{\frac{R_{\sigma^2}}{N}}$	
調整 GCI 取向 Donner & Zou (2012)	$LCI = R_{\bar{\theta}(\alpha/2)}$ <p>若 $\alpha=0.05$, $R_{\bar{\theta}(\alpha/2)}$ 是 P25 的 $R_{\bar{\theta}}$ 資料值</p> $R_{\bar{\theta}} = \frac{R_{\theta}}{2 - C_N}$ $= \frac{1}{2 - C_N} \left(\frac{\sqrt{\frac{(N-1)SD^2}{\chi_{N-1}^2}}}{\bar{x} - Z \sqrt{\frac{(N-1)SD^2}{N \times \chi_{N-1}^2}}} \right)$ $C_N = \sqrt{\frac{2}{N-1}} \times \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}$	$UCI = R_{\bar{\theta}(1-\alpha/2)}$ <p>若 $\alpha=0.05$, $R_{\bar{\theta}(1-\alpha/2)}$ 是 P97.5 的 $R_{\bar{\theta}}$ 資料值</p>

註：N 是樣本數，B 是 Bootstrap 模擬次數，CV* 是 Bootstrap 模擬所得 CV 值。通常設定 $\alpha=0.05$ ，或設定為 0.01、0.1。

表 15
考驗母群體 CV 值方法

提出者	計算公式	考驗方法
McKay (1932)	$DCV_{Mc} = \left(1 + \frac{1}{CV_{pop}^2}\right) \left(\frac{N \times CV^2}{1 + CV^2}\right)$	$DCV \sim \chi_{df=N-1}^2$
Hendricks & Robey (1936)	$DCV_{HR} = \frac{CV - CV_{pop}}{\sqrt{2N}}$	$DCV \sim t_{\alpha, N-1}$
Miller (1991)	採用調查資料 $DCV_{Mi} = \frac{CV - CV_{pop}}{\sqrt{\frac{CV^4 + \frac{CV^2}{2}}{N}}}$	$DCV \sim Z$
Sharma & Krishna (1994)	$DCV_{SK} = N \times (CV - CV_{pop})$	$DCV \sim Z$
Bootstrap t 法 Banik, Kibria, & Sharma (2012)	採用調查資料 $T_i^* = \frac{CV_i^* - \overline{CV}}{\frac{\hat{\sigma}_{CV}}{B}}$ $\hat{\sigma}_{CV} = \frac{1}{B} \sum_{i=1}^B (CV_i^* - \overline{CV})^2$ $\overline{CV} = \frac{1}{B} \sum_{i=1}^B CV_i^*$	P(1- α) 的 T* 值為 決斷值上限
Miller Bootstrap 法 取向 1 Banik, Kibria, & Sharma (2012)	採用模擬資料	$DCV \sim Z$

提出者	計算公式	考驗方法
	$DCV_{Mi} = \frac{CV - CV_{pop}}{\sqrt{\frac{CV^4 + \frac{CV^2}{2}}{N}}}$	
Miller Bootstrap 法 取向 2 Banik, Kibria, & Sharma (2012)	採用模擬資料 $T_i^* = \frac{CV_i - \overline{CV}}{\widehat{\sigma}_{CV}}$ $\widehat{\sigma}_{CV} = \frac{1}{B} \sum_{i=1}^B (CV_i^* - \overline{CV})^2$ $\overline{CV} = \frac{1}{B} \sum_{i=1}^B CV_i^*$	

註： $H_0: CV = CV_{pop}$ ，CV 是樣本變異係數， CV_{pop} 指母群體變異係數，N 是樣本數，k 是變項數。

表 16
多個 CV 值差異比較方法

提出者	計算公式	考驗方法
獨立樣本 $H_0: CV_1 = CV_2$ Hendricks & Robey (1936)	$DCV_{HR} = \frac{CV - 0}{\frac{CV}{\sqrt{2N}}}$	$DCV \sim t_{\alpha, N-1}$
Likelihood ratio test Gerig & Sen (1980)	$\widehat{\mu}_1 = \frac{\frac{CV}{\sqrt{2N}}}{(N_1 + N_2)\widehat{\mu}_2 - N_2 M_2}$ $\widehat{\mu}_2 = -\frac{q}{2p} + \sqrt{\frac{q^2}{4p^2} - \frac{r}{p}}$ $\widehat{p} = \sqrt{\frac{\frac{N_2 - 1}{N_2} CV_2^2 M_2^2 + M_2^2 - M_2 \widehat{\mu}_2}{\widehat{\mu}_2}}$ $p = (N_1 + N_2) CV_1^2 + N_2$ $q = -(2N_2 C_1^2 + (2N_2 - N_1) M_2)$ $r = \frac{(N_2^2 (CV_1^2 + 1) - N_1^2 (CV_2^2 + 1)) M_2^2}{N_1 + N_2}$	$DCV \sim \chi_{df=1}^2$
Bennett (1976)	$DCV_R = N_1 \log \left(\frac{(\widehat{p}\widehat{\mu}_1)^2}{\left(\frac{N_1 - 1}{N_1}\right) CV_1^2 M_1^2} \right) + N_2 \log \left(\frac{(\widehat{p}\widehat{\mu}_2)^2}{\left(\frac{N_2 - 1}{N_2}\right) CV_2^2 M_2^2} \right)$ $\theta_1 = \frac{N_1 - 1}{N_1}$ $\theta_2 = \frac{N_2 - 1}{N_2}$ $DCV_B = (N_1 + N_2 - 2) \log \left(\frac{\frac{N_1 \theta_1 CV_1^2}{1 + \theta_1 CV_1^2} + \frac{N_2 \theta_2 CV_2^2}{1 + \theta_2 CV_2^2}}{N_1 + N_2 - 2} \right) - (N_1 - 1) \log \left(\frac{N_1 \theta_1 CV_1^2}{1 + \theta_1 CV_1^2} \right) - (N_2 - 1) \log \left(\frac{N_2 \theta_2 CV_2^2}{1 + \theta_2 CV_2^2} \right)$	$DCV \sim \chi_{df=1}^2$

提出者	計算公式	考驗方法
Miller (1991)	$\gamma_w = \frac{(N_1 - 1)CV_1 + (N_2 - 1)CV_2}{N_1 + N_2 - 2}$ $DCV_M = \frac{CV_1 - CV_2}{\sqrt{\frac{\gamma_w^2}{2(N_1 - 1)} + \frac{\gamma_w^4}{N_1 - 1} + \frac{\gamma_w^2}{2(N_2 - 1)} + \frac{\gamma_w^4}{N_2 - 1}}}$	$DCV \sim Z$
Wald test Gupta & Ma (1996)	$DCV_W = \frac{(CV_1 - CV_2)^2}{\frac{CV_1^2}{2N_1} + \frac{CV_1^4}{N_1} + \frac{CV_2^2}{2N_2} + \frac{CV_2^4}{N_2}}$	$DCV \sim \chi_{df=1}^2$
Score test Gupta & Ma (1996)	$DCV_S = \left(\frac{1}{2}\hat{p}^2 + \hat{p}^4\right) \left(\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2}\right)$ $a_i = \frac{1}{\sqrt{\hat{\mu}_i}} \times \frac{1}{\sqrt[3]{\hat{p}}} \times \sum_{j=1}^{N_i} (x_{ij} - \hat{\mu}_i)^2 - N_i \frac{1}{\hat{p}}, i=1,2$	$DCV \sim \chi_{df=1}^2$
Doorbos & Dijkstra (1983)	$b_i = \frac{1}{CV_i}$ $b_w = \frac{N_1 b_1 + N_2 b_2}{N_1 + N_2}$ $T = N_1(b_1 - b_w)^2 + N_2(b_2 - b_w)^2$ $\overline{E(T)} = \frac{N_2(N_1 - 1)}{(N_1 + N_2)(N_1 - 3)} + \frac{N_1(N_2 - 1)}{(N_1 + N_2)(N_2 - 3)}$ $+ \frac{1}{C_p^2} \left[\frac{N_1 N_2 (N_1 - 1)}{(N_1 + N_2)(N_1 - 3)} + \frac{N_1 N_2 (N_2 - 1)}{(N_1 + N_2)(N_2 - 3)} + \frac{1}{N_1 + N_2} (N_1^2 e_1^2 + N_2^2 e_2^2 - (N_1 e_1 + N_2 e_2)^2) \right]$ $C_p^2 = \frac{\frac{N_1(N_1 - 1)}{N_1 - 3} + \frac{N_2(N_2 - 1)}{N_2 - 3}}{N_1 b_1^2 + N_2 b_2^2 - \left(\frac{N_1 - 1}{N_1 - 3} + \frac{N_2 - 1}{N_2 - 3}\right)}$ $e_i = \sqrt{\frac{N_i - 1}{2}} \times \frac{\Gamma\left[\frac{1}{2}(N_i - 2)\right]}{\Gamma\left[\frac{1}{2}(N_i - 1)\right]}$ $DCV_{DD} = \frac{T}{\overline{E(T)}}$	$DCV \sim \chi_{df=1}^2$
Log test	$DCV_{log} = \frac{SD_1^2 \times \log(y)}{SD_2^2 \times \log(y)}$	$DCV \sim F_{N_1-1, N_2-1}$
Naive test	$DCV_{NA} = \frac{CV_1^2}{CV_2^2}, CV_1 \geq CV_2$	$DCV \sim F_{N_1-1, N_2-1}$
關聯樣本 $H_0: CV_1 = CV_2$ Wald test Kalkur & Rao (2015)	$DCV_{KR} = \frac{CV_1 - CV_2}{\sqrt{\frac{CV_1^4}{N} + \frac{CV_1^4}{N} + \frac{CV_1^2}{2N} + \frac{CV_2^2}{2N} - \frac{2rCV_1^2 CV_2^2}{N} - \frac{r^2 CV_1 CV_2}{N}}}$	$DCV \sim \chi_{df=1}^2$ $CR_{DCV} \sim \chi_{1-(1-\alpha)^k, 1}^2$ $CR_{DCV} \sim \chi_{\alpha, k-1}^2$ $CR_{DCV} \sim F_{\alpha, (1, k-1)}$ $CR_{DCV} \sim F_{1-(1-\alpha)^k, (1, k-1)}$ $CR_{DCV} \sim F_{\alpha, (1, n-k)}$
$H_0: CV_1 = CV_2 = \dots = CV_k$ Shafer & Sullivan (1986)	$DCV_{SS} = (N - k) \ln \sum_{i=1}^k \frac{d_i}{N - k} - \sum_{i=1}^k (n_i - 1) \ln \left(\frac{d_i}{n_i - k} \right)$	$DCV \sim \chi_{df=k-1}^2$

提出者	計算公式	考驗方法
Miller (1991)	$N = \sum_{i=1}^k n_i$ $d_i = \frac{n_i CV_i^2}{CV_i^2 + 1}$ $DCV_{Mi} = \frac{1}{R^2(0.5 + R^2)} \left[\sum_{i=1}^k (n_i - 1) CV_i^2 - \frac{1}{N} \left(\sum_{i=1}^k (n_i - 1) CV_i \right)^2 \right]$ $R = \frac{1}{n - k} \sum_{i=1}^k (n_i - 1) CV_i$	
Miller (1991) 平方等級考驗 (squared ranks test, SRT)	$DCV_{Mi} = \frac{1}{D^2} \left[\sum \frac{R_i^2}{n_i} - N\bar{R}^2 \right]$ $R_i = \sum_{j=1}^{n_i} R_{ij}^2$ $R_{ij} = \text{rank of } x_{ij} - 1 $ $\bar{R} = \sum_{i=1}^k \frac{R_i}{N} = (N + 1)(2N + 1)/6$ $D^2 = N(N + 1)(2N + 1)(8N + 11)/180$	$DCV \sim \chi_{df=k-1}^2$

註：各方法考驗 $H_0: CV_1 = CV_2$, N, N_1, N_2 是樣本數, μ_1, μ_2 是平均數, r 是積差相關值。

表17

CV值評估標準之比較

Pearson (1932)	English 和 Keran (1976)	Pimentel 和 Gomes (2009) ^c	Yeh 三分類(2025) ^b
極精準 $0 < CV \leq 0.33$	良好 $0 < CV \leq 0.5$	極高 $0.3 < CV$	高 $(M + SD) < CV$
合理精準 $0.33 < CV \leq 0.67$	不滿意 $0.5 < CV \leq 0.8$	高 $0.2 < CV \leq 0.3$	中 $(M - SD) < CV \leq (M + SD)$
不佳 $0.67 < CV$	差 $0.8 < CV$	中 $0.1 < CV \leq 0.2$	低 $CV \leq (M - SD)$
		低 $CV \leq 0.1$	
Garcia (1989) ^b	Costa 等(2002) ^b	Lorenzo 等(2015) ^b	Yeh 四分類(2025) ^{ab}
極高 $(M + 2SD) < CV$	極高 $(M_d + 2PS) < CV$	高 $CV_{RL} < CV$	極高 $(M + C \times SD) < CV$
高 $(M + 1SD) < CV \leq (M + 2SD)$	高 $(M_d + 1PS) < CV \leq (M_d + 2PS)$	中 $CV_{LL} < CV \leq CV_{RL}$	高 $M < CV \leq (M + C \times SD)$
中 $(M - 1SD) < CV \leq (M + 1SD)$	中 $(M_d - 1PS) < CV \leq (M_d + 1PS)$	低 $CV \leq CV_{LL}$	中 $(M - C \times SD) < CV \leq M$
低 $CV \leq (M - 1SD)$	低 $CV \leq (M_d - 1PS)$		低 $CV \leq (M - C \times SD)$

註： $M_d = (Q_1 + Q_3)/2$, $PS = IQR/1.35$, $IQR = Q_3 - Q_1$, $CV_w = (CV_{max} - CV_{min})/3$, $CV_{LL} = CV_{min} + CV_w$,

$CV_{RL} = CV_{max} - CV_w$ 。Costa 等人的 CV 是 $\frac{\sqrt{OME}}{M}$ 近似 $\frac{SD}{M}$ 。^a 可設定 $C=1.96$ 或其他值如 0.5、1。^b M、SD、 M_d 和 PS 值應以全部項的 CV 值計算取得。^c 其基於 %CV，提出 10%、20% 和 30% 為區分値。

類卻是表示 CV 值極差，這是與 Pearson (1932) 等對 CV 值分類命名觀點相左，所以這是根據分類命名判斷 CV 值優劣時必須特別小心之處。

此外，觀察 Garcia (1989) 和 Costa 等 (2002) 所提評估 CV 值分類標準有幾分神似，可簡化為一個通則公式，Ferreira 等人 (2018) 就提出如表 18 的通則計算公式及可採行的六種參數設定方式，以形成多樣化的評估標準系統，其實證研究比較發現使用 MDQ 和 PS 的做法與採用 M 和 SD 的做法產生相似的結果。簡言之，這個通則公式是基於某種分配型態的假設，Garcia (1989)、Costa 等 (2002) 等評估標準及分類系統成立與否有賴於這個假設是真實成立，通常係假設 CV 值呈現常態分配型態，但是不少論者 (Forkman, 2005; McKay, 1932) 也提

表18
CV值評估標準分類之設定方法比較

分類及通則計算公式	設定方式	A 設定 ^a	B 設定 ^a	提倡者
4.極高 (A+2B)<CV	方法 1	M _d	PS	Costa 等(2002)
3.高 (A+1B)<CV≤(A+2B)	方法 2	M _d	SD	
2.中 (A-1B)<CV≤(A+1B)	方法 3	M	PS	Garcia (1989)
1.低 CV≤(A-1B)	方法 4	M	SD	
	方法 5	M _{DQ} =(Q ₁ +Q ₃)/2	SD	
	方法 6	M _{DQ} =(Q ₁ +Q ₃)/2	PS	

註：Costa 等人將 M_{DQ} 視同為 M_d，但是兩者數值未必相等，M_{DQ}=(Q₁+Q₃)/2。^aA 和 B 設定值應以全部項的 CV 值計算取得。

出其他可能的資料分配型態如 t 分配、F 分配、 χ^2 分配等，則評估標準劃分可能隨之變動，亦有建議使用 bootstrap 法和 Bayesian 取向分析 CV 值 (Banik, Kibria, & Sharma, 2012; Thangjai, Niwitpong, & Niwitpong, 2021)，則表 18 所列 A 和 B 值將以模擬資料進行分析，或許更能接近母群體真實值。

收集多個實證研究文獻所載平均數和標準差資料，分析出 CV 值見表 19。就考驗固定標準而言，幾乎全部研究都符合兩類標準的最佳分類，CV 值 < 0.5。表 20 是使用一個實際研究的 11 個變項平均數、標準差等資料，分析六個評估 CV 值標準等級方法的效果，顯示 English & Keran、Garcia 等人、Lorenzo 等人所提分類方法出現多個分類結果，另外三種分類觀點僅出現一種分類結果，經投入平均數、標準差和 CV 值進行區別分析，得出 English & Keran 法區別正確率皆 100%，Garcia 等人分類觀點略優於 Lorenzo 等人分類看法，區別正確率都逾 90%，這表示 Garcia 等人和 Lorenzo 等人的分類方法可以採用，其均具有統計資料分配學理依據，且分類正確率亦高。

再使用實際資料進行 bootstrap 法 (Stine, 1989) 模擬 1000 次分析 CV 值，表 21 呈現分析結果。不難看出採用 Pearson 標準，達 96% 的 CV 值達 0.33 標準值，若選取 0.67 則全部都達共識，而依據 English 和 Keran 的標準，所有項都通過，表示達到共識。這表明 Pearson 標準頗適用，尚具有區分效果，反觀 English 和 Keran 所提標準可能偏於寬鬆，故後續或許可以 0.3 為嚴格 CV 值評估標準。

表19
Delphi分析實證研究CV值及採用判斷標準值評估結果比較

研究	項數	量尺值	CV ^b	Pearson ^a			English & Keran ^a		
				≤0.33	0.33~0.67	>0.67	≤0.5	0.5~0.8	>0.8
1	7	5	[0.121,0.259]	100	0	0	100	0	0
2	4	5	[0.126,0.251]	100	0	0	100	0	0
3	22	5	[0.044,0.451]	95.45	4.55	0	100	0	0
4	14	5	[0.081,0.249]	100	0	0	100	0	0
5	18	5	[0.016,0.274]	100	0	0	100	0	0
6	15	5	[0.088,0.227]	100	0	0	100	0	0

研究 項數	量尺值	CV ^b	Pearson ^a			English & Keran ^a			
			≤0.33	0.33~0.67	>0.67	≤0.5	0.5~0.8	>0.8	
7	22	5	[0.046,0.34]	95.45	4.55	0	100	0	0
8	19	5	[0.046,0.268]	100	0	0	100	0	0
9	13	5	[0.136,0.246]	100	0	0	100	0	0
10	14	5	[0.1,0.266]	100	0	0	100	0	0

註：針對取自學位論文和期刊論文的平均數和標準差資料分析所得CV值。^a顯示數值為歸類數占全部項數%。
^b顯示CV值的最小值和最大值域。

表20
 Delphi分析實證研究CV值及採用判斷標準值評估結果比較

項目	平均數	中數	PS	標準差	CV	CV值評定等級 ^a					
						Pearson	English	Pimentel	Garcia	Costa	Lorenzo
1	12.06	12.43	4.91	4.3	0.357	2	1	4	1	1	1
2	14.35	14.25	5.27	4.93	0.344	2	1	4	1	1	1
3	10.59	9.12	3.61	4.86	0.459	2	1	4	2	1	2
4	22.16	17.54	7.14	14.12	0.637	2	2	4	3	1	3
5	20.76	19.35	9.25	12.37	0.596	2	2	4	3	1	3
6	11.62	12.59	4.67	5.15	0.443	2	1	4	2	1	2
7	10.27	9.9	3	4	0.389	2	1	4	2	1	1
8	9.97	9.52	7.05	4.08	0.409	2	1	4	2	1	1
9	18.29	16.62	12.31	11.5	0.629	2	2	4	3	1	3
10	9.49	9.47	7.01	4.16	0.438	2	1	4	2	1	1
11	15.67	16.53	12.24	6.4	0.408	2	1	4	2	1	1
					考驗1 ^b	100	100	100	100	100	90.9
					考驗2 ^b	100	100	100	90.9	100	90.9

註：平均數、標準差、PS和CV值取自外文期刊論文。^a顯示數值為評定等級。^b考驗1是投入CV值，考驗2是投入平均數和標準差值進行區別分析，數值為區別正確率(%)。

表21
 Delphi分析模擬研究CV值及採用判斷標準值評估結果比較

變項	最小	最大	平均數	中數	P5 ^b	Pearson ^a			English & Keran ^a		
						cv≤0.33	0.33~0.67	>0.67	cv≤0.50	0.50~0.80	>0.80
1	0.071	0.384	0.203	0.202	0.107	98.2	1.8	0	100	0	0
2	0.071	0.250	0.124	0.126	0.071	100	0	0	100	0	0
3	0.071	0.378	0.240	0.242	0.132	95.9	4.1	0	100	0	0
4	0.071	0.374	0.230	0.242	0.107	96.2	3.8	0	100	0	0

註：針對4個變項實際資料(量尺值為5)採用bootstrap法模擬1000次分析所得CV值。^a顯示數值為歸類數占全部項數%。^b為百分位數5。

六、採用CDI指數和偏於主觀的評估標準，卻乏嚴謹統計考驗

關於評估共識性的指數，近年來由鄧振源(2012)修改CV，增加調整常數形成CDI指數和DC指數，計算公式見表22，根據公式可知CDI指數測量未達共識的程度，DC指數表示達共識程度，兩者擇一即可。觀察其逐漸見諸中、英文

期刊論文和學位論文，然而檢視 Google Scholar Search，採用該指數的英文期刊論文卻大多數是臺灣地區研究者撰寫，少部分是國外和大陸地區研究者。前面提到將 CV 納入考量 CV 的最大值，可形成 CVs，依此構造理念，能將 CDI 指數的最大值納入，以形成 CDIs 指數（normalized Consensus Deviation Index, scaled Consensus Deviation Index），使得分析值規範於 [0,1]，並建構出 DCs，另外亦能提出調整共識性差異指數（adjusted Consensus Deviation Index, CDIA），為將 CVs 根據 CDI 理念所建構試圖去除量尺值影響的新指數。彙總 CDI 及衍生指數的計算公式如表 22。

表 22
CDI 及有關衍生統計量數

統計量數	分析公式
共識性差異指數, CDI	$CDI_1 = CV_i \frac{M_i}{\max(M_i)} = \frac{SD_i}{\max(M_i)}$ $CDI_2 = CV_i \frac{M_i}{VS} = \frac{SD_i}{VS}$
共識程度, DC	$DC = 1 - CDI$
正規化共識性差異指數, CDIs	$CDI_s = CDI \frac{VS}{M\sqrt{N} - 1}$
正規化共識程度, DCs	$DC_s = 1 - CDI_s = 1 - CDI \frac{VS}{M\sqrt{N} - 1}$
調整共識性差異指數, CDIA	$CDI_A = CV_s \frac{M_i}{VS} = \frac{M_i \times CV_i}{VS\sqrt{N} - 1} = \frac{SD_i}{VS\sqrt{N} - 1}$

註：N 是樣本數，M 是平均數，SD 是標準差。VS 是最大的評定量尺值， $\max(\bar{x}_i)$ 是向度內項目或全部項目的平均數最大者。

關於 CDI 指數存在五個問題待思考，以下分別討論：

問題 1：將 CDI 指數是同於 CV 係數？

檢視一些應用文獻使用 CDI 指數的情形，明顯或隱約地透露出研究者認為 CDI 指數 = CV 係數，然而細視 CDI 指數的計算公式 $\frac{SD_i}{\max(\bar{x}_i)}$ 明顯和 CV 係數的 $\frac{SD}{M}$ 不同，因此將兩者混為一談，實在不妥。CDI 指數和 CV 係數的計算公式說明兩者數值反映的資訊不同，前者聚焦於試圖控制評定量尺值對 CV 值的影響，即 CDI 指數是調整的 CV 係數。綜言之，需要明辨兩者差異，不能混淆。

問題 2：對 CDI 指數計算方法存在不同理解？

其次，觀察多數研究中呈現至少兩種計算 CDI 指數的想法，所以計算 CDI 值需要確定 $\max(\bar{x}_i)$ 的定義。關於分析該值大致有三種做法，一是選取某向度內實際最大的項目平均數，二是不分向度，選取全部項目中實際最大的項目平均數，三是選取該次分析採用最大的評定量尺值（scale maxima），如採用 Likert 五點量尺評分，則 5 是最大的評定量尺值。假設兩個項目的平均數和標準差值如表 23，在狀況 1 時，最大評定值為 5，觀察兩種計算 $\max(\bar{x}_i)$ 值做法所得 CDI 指標值，

CV值顯示項目2共識性表現優於項目1，可是 CDI_1 值卻顯示兩者相同，顯然與事實矛盾，而 CDI_2 值與CV值一樣都指出項目2表現優於項目1。在狀況2時，最大評定值為7，兩項目的 CDI_1 值完全相同， CDI_2 值則隨之改變，亦顯示項目2共識性優於項目1。這表明不應採取實際最大的項目平均數去計算CDI值， CDI_1 無法調整因評定量尺值改變的影響，可是不少臺灣地區學位論文使用這個計算方法，明顯有待思考改善，換言之宜採用 CDI_2 計算做法，能反映出評定量尺值改變的影響，利於進行跨評定量尺比較。

表23

虛擬兩項目資料分析CDI結果之比較

			評定量尺	平均數	標準差	CV	CDI_1	CDI_2
狀況1	向度1	項目1	5	4	2	0.5	0.4444	0.4
		項目2	5	4.5	2	0.4444	0.4444	0.3996
狀況2	向度1	項目1	7	4	2	0.5	0.4444	0.28571
		項目2	7	4.5	2	0.4444	0.4444	0.28569
狀況3	向度1	項目1	5	4	2	0.5	0.4167	0.4
		項目2	5	4.5	2	0.4444	0.4167	0.3996
	向度2	項目1	5	4.3	2.15	0.5	0.4479	0.43
		項目2	5	4.8	2.1331	0.4444	0.4444	0.4266

註：最大評定值=評定量尺。狀況1和2時最大平均數= $\max(4,4.5)=4.5$ ， $CDI_1=CV \times$ 平均數/4.5， $CDI_2=CV \times$ 平均數/最大評定值。狀況3時最大平均數= $\max(4,4.5,4.3,4.8)=4.8$ ， $CDI_1=CV \times$ 平均數/4.8。

接著討論狀況三， CDI_1 指數是要選取向度內或全體項目的最大平均數進行計算。表23指出向度1的兩項目 CDI_1 值相同，CV值則不同，反觀 CDI_2 值卻展示在四個項目的指數值完全不同。這表示不同的CDI指數計算公式會造成不同的分析結果值，透露出 CDI_2 較佳於 CDI_1 的表現效果，利於做跨向度和跨項目的比較。

問題3: CDI指數可取代CV係數?

可以CDI指數取代CV係數嗎? 主要得考察這些指數或係數所體現的達高共識性資訊是否合理? 能否進行跨項目、跨向度和跨研究比較? 觀察CDI指數在於調整CV值可能受到評量項目時採用最大尺度值不同的影響，故 CDI 值 $<CV$ 值，但是CDI值仍未規範於0~1之間，可能會 >1 。相對地，正規化變異係數(CVs)是以最大值 $\sqrt{N-1}$ 為調整依據，其值落於0~1，利於解釋和理解，可說是比CV值和CDI值更利於判斷和說明共識性程度；反觀尚未見論述指出CDI值的上限，提出類似CVs的正規化指數，可將CDI值規範於[0,1]，實有待研議。

對此，經過一番推導，見附錄三說明，得知CDI值(係指基於 CDI_2 的計算公式和理念)的理論最大值为 $\frac{M\sqrt{N-1}}{VS}$ ，故可提出正規化共識性差異指數(normalized Consensus Deviation Index, scaled Consensus Deviation Index, CDI_s)， $CDI_s = CDI \frac{VS}{M\sqrt{N-1}}$ ，其值規範於[0,1]， $CDI_s=0$ 最佳， $CDI_s=1$ 最差，該值

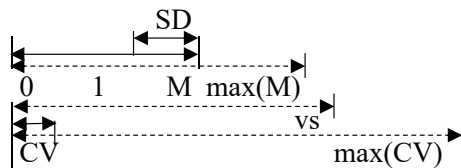
越小越好，此便利解讀和判斷。據此，也能進一步推導（見附錄三說明）正規化共識程度（DCs）， $DC_s = 1 - CDI_s = 1 - CDI \frac{vs}{M\sqrt{N-1}}$ ，其值亦規範在 $[0,1]$ ， $DC_s=0$ 最差，是毫無共識， $DC_s=1$ 最佳，是完全共識，該值越大越好，頗便利解讀和判斷。

再審思表 23 表明 CV 值未反映出評定量尺值的改變影響，但是否表示 CDI_2 就比較好？也就是說得思考評定量尺值是否會對共識性評估產生不可忽視的影響？CV 值作用是顯示標準差值佔平均數值的比例，SD 值和 M 值決定 CV 值，SD 能視為各值距離平均數的離均差（見圖 5），SD 值=0 即各值相同，SD 值越大即各值和平均數差距越大，而 M 值則做為衡量的基準依據， $M \leq$ 最大評定量尺值， $CV=SD/M$ 就同時納入考量各項目評分的集中程度值和歧異程度值資訊，當 M 值越大，SD 值越小，CV 值隨之越小，SD 值=0，CV 值=0 表示各項目評分無差異，卻無關於 M 值大小，只反映出評分歧異而無法反映評分重要度資訊，這是 CV 的限制。

又圖 5 指出 SD 值缺少比較基準，M 值可以最大評定量尺值 (vs) 確定優劣，CV 值亦能以最大 CV 值（即 $\sqrt{N-1}$ ）做比較基準，而 CDI 衍生自 CV，是調整的 CV，卻缺少比較基準，目前若干研究採用的 0.2、0.05 等判斷基準值實際上都缺少學理方面的支持證據，這是應用 CDI 的困境，有待改善。

圖 5

平均數、標準差、最大評定量尺值、CV 和最大 CV 值之意義比較



註：vs 是最大評定量尺值。

進一步以六個變項 300 個樣本的實證資料，採取 Bootstrap 法抽樣 1000 次進行 CV、CVs、CDI 和 CDI_s 分析值的比較，結果見表 24，另分析一個變項的 CV 等四個指數分析值分配型態見圖 6。表 24 和圖 6 指出四個指數值分配型態近乎常態分配型態，CV 值的常態分配特性更明顯於 CDI 值，進行 Kolmogorov-Smirnov 常態考驗 (KS) 及 Shapiro-Wilk 常態考驗 (SW) 考驗結果（見表 24 的常態考驗部分）顯示多數 CV 值呈常態分配，而 CDI 值僅部分呈常態分配；至於 CVs 值和 SDI_s 值的常態分配考驗結果與前述關於 CV 值和 CDI 值的考驗結果相仿。採用積差相關係數考驗六個變項 CV 值和 CDI 值的關聯性，為 $r=.9816\sim.9931$ 皆 $p<.001$ ，顯然兩者達高關聯。而衍生的 CVs 和 CDI_s （表 24）兩者分析值關聯性

亦達 $r=.9816\sim.9931$ ，皆 $p<.001$ ，表示兩者係高關聯關係。因為CV值和CDI值為極高度關聯，則CV和CDI兩指數似乎擇一即可，CVs值介於 $[0,1]$ ，CV值介於 $[0, \sqrt{N-1}]$ ；所以就解釋便利性和易理解性來看，CVs顯然較優於CV，當然CDIs也明顯較優於CDI。大抵觀之，圖6顯示結果與論者（Arachchige, Prendergast, & Staudte, 2022）指出CV值為漸近對稱常態分配的說法相符。

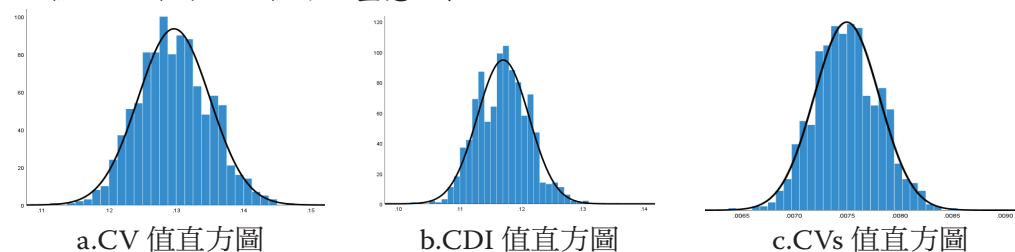
表24
 CV、CVs、CDI和CDIs分析值比較

	CV						CVs					
	x1	x2	x3	x4	x5	x6	nx1	nx2	nx3	nx4	nx5	nx6
M	0.1297	0.1145	0.1154	0.1481	0.1224	0.1222	0.0075	0.0066	0.0067	0.0086	0.0071	0.0071
SD	0.0053	0.0040	0.0049	0.0070	0.0041	0.0043	0.0003	0.0002	0.0003	0.0004	0.0002	0.00025
最小	0.1115	0.1018	0.0988	0.1272	0.1105	0.1107	0.0064	0.0059	0.0057	0.0074	0.0064	0.01
最大	0.1474	0.1281	0.1311	0.1730	0.1353	0.1374	0.0085	0.0074	0.0076	0.0100	0.0078	0.01
偏態	0.1092	0.0688	0.1319	0.0777	0.1804	0.0564	0.1092	0.0688	0.1319	0.0777	0.1804	0.056
峰態	-0.1558	0.0102	0.0467	0.0719	-0.0490	0.0363	-0.1558	0.0102	0.0467	0.0719	-0.0490	0.036
常態	0.025	0.023	0.025	0.037**	0.041***	0.021	0.025	0.023	0.025	0.037**	0.041***	0.021
考驗	0.998	0.999	0.998	0.999	0.997*	0.998	0.998	0.999	0.998	0.999	0.997*	0.998

	CDI						CDIs					
	x1	x2	x3	x4	x5	x6	nx1	nx2	nx3	nx4	nx5	nx6
M	0.1170	0.1050	0.1065	0.1289	0.1100	0.1099	0.0068	0.0061	0.0062	0.0075	0.0064	0.0064
SD	0.0042	0.0032	0.0040	0.0055	0.0033	0.0034	0.0002	0.0002	0.0002	0.0003	0.0002	0.0002
最小	0.1029	0.0955	0.0932	0.1103	0.1008	0.1009	0.0060	0.0055	0.0054	0.0064	0.0058	0.0058
最大	0.1307	0.1161	0.1204	0.1491	0.1205	0.1228	0.0076	0.0067	0.0070	0.0086	0.0070	0.0071
偏態	0.1384	0.0930	0.1550	0.0555	0.2432	0.0954	0.1384	0.0930	0.1550	0.0555	0.2432	0.0954
峰態	-0.1825	-0.0203	0.0636	0.1436	-0.1068	-0.0363	-0.1825	-0.0203	0.0636	0.1436	-0.1068	-0.0363
常態	0.054***	0.024	0.03*	0.023	0.088***	0.063***	0.054***	0.024	0.03*	0.023	0.088***	0.063***
考驗	0.995**	0.999	0.997	0.999	0.991***	0.994***	0.995**	0.999	0.997	0.999	0.991***	0.994***
r	.9890	.9879	.9931	.9853	.9816	.9823	.9890	.9879	.9931	.9853	.9816	.9823

註：n=300。以Bootstrap法抽樣1000次所得資料進行分析。r是CV值和CDI值、CVs值和CDIs值的積差相關值，皆 $p<.001$ 。畫底線的偏態係數值表示顯著考驗 $p<.05$ 為正偏態，無底線表示常態。常態考驗部分上方為Kolmogorov-Smirnov常態考驗統計量值(KS)及顯著性考驗結果，下方為Shapiro-Wilk常態考驗統計量值(SW)及顯著性考驗結果。* $p<.05$ ，** $p<.01$ ，*** $p<.001$

圖6
 CV值、CDI值和CVs值分配型態比較



註：a~c是採用Bootstrap法抽樣1000次產生300個樣本資料的分析結果。

問題 4：如何選擇 CDI 指數評估標準值？

應用 CDI 指數必須選定一個評估標準值 (ϵ)，以判斷共識度是否達到可接受的程度 (鄧振源, 2012)，可是採行標準值卻為分歧，已見 0.1、0.2、0.05、0.5 或 0.3，其中 0.5 最為寬鬆。為何選定這些標準值，多數研究未提學理依據，少數指出是徵詢多位專家的共識，更多是沿用他人的慣例，如此看來主觀直覺和慣例係選取的依據。而評估標準值應用於實際研究的效果為何，並未見討論，採用多個實證研究文獻刊載的平均數和標準差資料，探討 CDI 標準值的影響見表 25。不難看出採取 $\epsilon > 0.2$ 時逾 83% 的分析結果可視為達共識，隨著 ϵ 值收窄，至 0.1 達共識率低於 43%，若採取 0.05 則達共識率更低至不逾 11%，也就是說幾乎近 89% 的 Delphi 分析結果會被判定為未達共識。

表 25

Delphi 分析實證研究 CDI 值及採用判斷標準值評估結果比較

研究	項數	量尺值	CDI ^b	CDI 判斷標準值 ^a				
				0.05	0.1	0.2	0.3	0.5
1	7	5	[0.112,0.21]	0	0	85.71	100	100
2	4	5	[0.103,0.18]	0	0	100	100	100
3	22	5	[0.042,0.384]	4.55	31.82	95.45	95.45	100
4	14	5	[0.078,0.2]	0	28.57	92.86	100	100
5	18	5	[0.014,0.216]	5.56	5.56	94.44	100	100
6	15	5	[0.084,0.174]	0	13.33	100	100	100
7	22	5	[0.046,0.236]	4.55	40.91	95.45	100	100
8	19	5	[0.046,0.218]	10.53	42.11	94.74	100	100
9	13	5	[0.126,0.2]	0	0	92.31	100	100
10	14	5	[0.094,0.22]	0	7.14	92.86	100	100

註：針對取自學位論文和期刊論文的平均數和標準差資料分析所得 CDI 值。^a 顯示數值為%。^b 顯示 CDI 值的最小值和最大值域。

另以實際資料進行 bootstrap 法模擬 1000 次分析 CDI 值，結果見表 26。不難發現設定 $\epsilon = 0.2$ 時，皆全部達標準值，若選取 0.1 則至多僅約 6% 會達共識，此較諸前述表 25，更顯得嚴格。該表也說明 PR5 時對應的 CDI 值，在 0.1~0.12 之間，而設定 $\epsilon = 0.1$ ，該值的 PR 是 4.3 或 4.5，因為 CDI 值一如 CV 值都是越小越好，其似乎表明 $\epsilon = 0.1$ 約能視為可能 P5 的下限，此為 ϵ 值設定提供了其統計學理的依據。據此可見，不論針對實際資料或模擬資料，都顯示 0.1 能視為嚴格評估標準值 ϵ 的上限， $\epsilon < 0.1$ 時說明評估共識性的標準頗高，0.2 則是寬鬆評估標準值的下限；這可供研究者參考：考量嚴格性，應選 0.1，若是想寬鬆些，那麼 0.2 是適當的評估標準值。

表26

Delphi分析模擬研究CDI值及採用判斷標準值評估結果比較

變項	CDI					CDI判斷標準值 ^a				
	最小	最大	平均數	中數	P5	0.05	0.1	0.2	0.3	0.5
1	0.105	0.129	0.117	0.117	0.11	0	0	100	100	100
2	0.095	0.118	0.105	0.105	0.10	0	5.5 (PR4.3)	100	100	100
3	0.094	0.120	0.106	0.106	0.10	0	5.9 (PR4.5)	100	100	100
4	0.115	0.146	0.129	0.128	0.12	0	0	100	100	100
5	0.101	0.121	0.110	0.110	0.105	0	0	100	100	100
6	0.100	0.122	0.110	0.110	0.109	0	0	100	100	100

註：針對6個變項實際資料（量尺值為5）採用bootstrap法模擬1000次分析所得CDI值。^a顯示數值為%。

進一步思考，表25和表26皆顯示約30%的分析若採取 $\epsilon=0.1$ 為標準，將導致該輪回評估出現無一項目看法是達共識的窘境，若改為 $\epsilon=0.2$ 就可以解決此問題。這指出採取固定 ϵ 值的做法（稱為固定標準，fixed criterion-value）無法完全適用於所有情況，可能有將近30%機率會碰到得調高 ϵ 值，才能有效處理CDI值評估共識適當性的難題；表示或許得思考另一個選項：使用變動 ϵ 值設定的思維，即變動標準（flexible criterion-value），依靠設定CDI值分配型態，運用bootstrap法配合百分位數（Pp）確認來協助，此構想尚待深入討論。

問題5：CDIs和CDIA兩指數如何選擇？

CDIs和CDIA兩指數都是將CDI值進行調整，CDIs將分析值規範於[0,1]，CDIA則試圖去除量尺值的影響，兩者構建公式和依據理念並不相同，可能導致最終分析值也有差異。針對三筆實證資料比較CDIs、CDIA等指數分析結果（表27），不難看出CVs值和CDIs值相同，分析值呈現CVs值，CDIs值<CDIA值<CDI值<CV值的規律（見表27的係數值部分），觀察分析值的排序則能看出：CDI值和CDIA值兩者值排序有時相同，有時不同，而CV值、CVs值和CDIs值三者排序卻完全相同。追根究柢，這實在是肇因於CDI和CDIA採用調整數 $M_i/\max(M_i)$ 產出值受到 M_i 影響極大，其造成的調整效果不像CVs和CDIs值採用調整數 $\max(CV_i)$ 和 $\max(CD_i)$ 是固定的，乃造成調整效果恆定。

據此觀之，因為CVs值和CDIs值相同，故兩者擇一即可。CDIA調整效果略佳於CDI，CDIA值<CDI值，表27顯示比較這兩個指數和CV等係數的係數值排序相似度，可見CDIA值的排序變動高於CDI，表示其與CV等係數展示的變異差異訊息略異。綜言之，應優先選擇CDIs，CDIA為次選項。

表27

Delphi分析實證研究CDIs和CDIA值分析結果比較

項目	M	SD	係數值					係數值排序				
			CV	CVs	CDI	CDIs	CDIA	CV	CVs	CDI	CDIs	CDIA
資料1												
1	4.57	0.57	0.125	0.024	0.114	0.024	0.104	4	4	4	4	4
2	4.07	0.77	0.189	0.036	0.154	0.036	0.125	1	1	1	1	2
3	4.39	0.74	0.169	0.032	0.148	0.032	0.130	3	3	2	3	1
4	4.18	0.72	0.172	0.033	0.144	0.033	0.120	2	2	3	2	3
5	4.61	0.5	0.108	0.021	0.100	0.021	0.092	5	5	5	5	5
資料2												
1	4.21	0.79	0.188	0.036	0.158	0.036	0.133	1	1	1	1	1
2	4.54	0.69	0.152	0.029	0.138	0.029	0.125	4	4	4	4	2
3	4.18	0.72	0.172	0.033	0.144	0.033	0.120	3	3	3	3	3
4	3.96	0.74	0.187	0.036	0.148	0.036	0.117	2	2	2	2	4
資料3												
1	4.39	0.57	0.130	0.025	0.114	0.025	0.100	8	8	8.5	8	9
2	4.21	0.83	0.197	0.038	0.166	0.038	0.140	5	5	4	5	4
3	3.93	0.94	0.239	0.046	0.188	0.046	0.148	2	2	2	2	2
4	4.57	0.57	0.125	0.024	0.114	0.024	0.104	9	9	8.5	9	8
5	4.25	1.08	0.254	0.049	0.216	0.049	0.184	1	1	1	1	1
6	4.43	0.69	0.156	0.030	0.138	0.030	0.122	6	6	6	6	5
7	4.64	0.62	0.134	0.026	0.124	0.026	0.115	7	7	7	7	7
8	3.77	0.81	0.215	0.041	0.162	0.041	0.122	4	4	5	4	6
9	3.89	0.92	0.237	0.046	0.184	0.046	0.143	3	3	3	3	3

註：取自實際調查資料，N=28，量尺值=5。

七、未重視和應用RAND/UCLA適當法（RAM法）、ROMPER法

RAM法和ROMPER法係強化經典Delphi分析的新方法，國外已見不少應用。前者創新使用對稱調整百分位距（Inter-Percentile Range Adjusted for Symmetry, IPRAS）指數和DI指數（disagreement index）值，以評估是否達成共識（Fitch, *et al.*, 2000）。後者採用GRADE評定法（Grading of Recommendations Assessment, Development and Evaluation），增進分析嚴謹性，採擷後設分析（meta analysis）使用的森林圖（forest plot），以分析各項評值的分布情形，適用於政策Delphi分析（Grant & Smart, 2024），可惜罕見臺灣地區論者討論這兩者。

至於計算IPRAS指數及其他指數的公式如下，其中IPR是百分位距（Inter-Percentile Range），IPRL是IPR的下限（Lower percentile for the IPR），通常取P30，IPRU是IPR的上限（upper percentile for the IPR），通常取P70，IPRCP（IPR central point, IPRc）是IPR的中間值，AI（Asymmetry Index）是評定量尺值的中間值與IPRCP值差的絕對值，IPRr（Inter-percentile Range required

for disagreement when perfect symmetry exists) 是當完美對稱時達不一致的百分位距，通常設定 $IPR_r = 2.35$ ，CFA 是不對稱校正因子 (Correction Factor for Asymmetry)，通常設定 $CFA = 1.5$ ，DI (disagreement index) 是不一致指數，若 $DI \geq 1$ 或 IPR 值 $> IPRAS$ 值都表示不一致 (disagreement)，即評估者看法未達共識。需注意的是，此處的 $IPRAS$ 指數和 DI 指數應用是設定使用 9 點量尺，即評定量尺值為 1~9 時，一般 Delphi 分析多使用 Likert 5 點量尺，則計算 $IPRAS$ 指數和 DI 指數值可能需要以調整值進行分析值調整，後續將討論之。

$$IPRAS = IPR_r + (AI \times CFA) \quad IPR_r = 2.35$$

$$CFA = 1.5$$

$$IPRCP = \frac{IPR_L + IPR_U}{2} \quad IPR_L = P_{30} \quad IPR_U = P_{70}$$

$$IPR = IPR_U - IPR_L$$

$$AI = |SV_{middle} - IPRCP| \quad SV_{middle} = \text{評定量尺值的中間值, 設定為 } 5$$

$$DI = \frac{IPR}{IPRAS}$$

使用實證評估資料分析 CDI 、 DC 、 $IPRAS$ 和 DI 指數，分析結果如表 28。其顯示 4 個項目的 DI 值都 < 1 ， $IPRAS$ 值都 $> IPR$ 值，表示皆達共識。反觀 CDI_1 值為 0.126~0.22， CDI_2 值為 0.103~0.18，以 $CDI < 0.2$ 為衡量標準則僅有項目 2 達共識，以 < 0.05 為判斷標準則無一項目顯示評估意見達共識；這表明 $IPRAS$ 和 DI 指數比 CDI 和 DC 指數更有應用的參考價值，其能明確決斷達共識與否，而 CDI 和 DC 指數有待提出明確和合理的決斷標準，後續將深入探討。

表 28

Delphi 分析實證評估資料 CDI 、 DC 、 $IPRAS$ 和 DI 指數之分析結果

項目	CDI_1	DC_1	CDI_2	DC_2	IPR	$IPRAS$	DI
1	0.209	0.791	0.171	0.829	0	50.620	0
2	0.126	0.874	0.103	0.897	0	50.620	0
3	0.220	0.780	0.180	0.820	1.800	40.270	0.422
4	0.217	0.783	0.178	0.822	0.900	40.945	0.182

註：項目 1~項目 4 資料取自某學位論文實證研究。N=12。

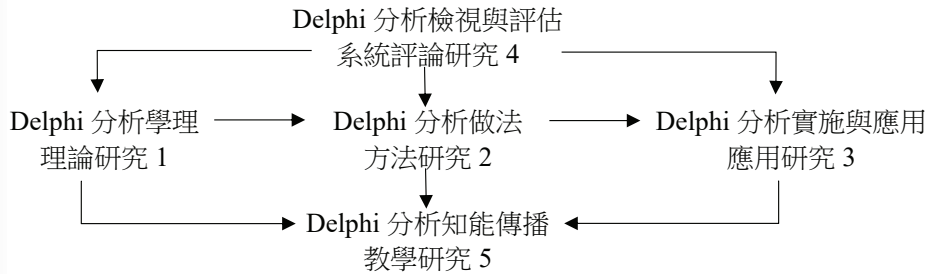
八、未建立 Delphi 分析研究與應用生態體系，高度聚焦低層次的應用

欲強化應用 Delphi 分析，其實背後有賴一個研究與應用生態體系支持和協助，圖 7 說明一個模型。此模型揭示出對於有關 Delphi 分析學理、方法、應用和傳播的評論研究部分 (圖 7 的部分 4) 佔著極關鍵影響地位，目前這方面的討論主要是系統性彙整和討論已知某領域內期刊論文、學位論文、研討會議論文、研究報告等採用 Delphi 分析做法的做法，甚至誤用和誤解，以描繪出一個具普遍

性的應用形態，供研究者了解和採用；觀察這類外文論述不少，但是中文論述則稀缺，影響所及就是中文實務應用和前沿學術研究有不小的落差，前面討論已表明不少研究者不知使用較佳、較新的分析理念和方法，這衝擊應用成果的品質確保，有待突破。

圖7

Delphi分析研究與應用生態體系



簡言之，臺灣地區應用 Delphi 分析出現幾項問題：包括 1. 高度依賴既有論述，但是其概介內容與當前 Delphi 分析及共識性評估等理論和方法層面的研究有嚴重落差，缺少對新近十年內理論發展的討論和介紹，即圖 7 部分 1 和 2 發展存在代差問題；2. 多數應用研究尤其是量化 Delphi 分析側重針對單項使用簡單描述統計資訊如平均數、標準差等，缺乏採用 Kappa 一致性係數等進行向度項目共識性檢視，即圖 7 部分 2 和 3 發展存在應用效能 (utility) 可慮問題；3. 多數教學參考用教科書仍刊載較舊的 Delphi 分析方法說明，未及時更新較新發展成果，即圖 7 部分 5 發展存有新知感知落差問題。上述幾項問題環環相扣，產生複雜的負交互作用效果，又缺乏圖 7 部分 4 可發揮提醒和引導的作用，整體而言就是沒有形成完整的生態體系，長久以往，可能衝擊 Delphi 分析的應用效益，著實有待思考改進。

九、綜合討論

上述討論的八項問題其實可歸併為評者專業、評估工具有效、應用嚴謹、發展生態等四類，前三者屬於微觀層次，可以逐一進行細項解決和因應，而發展生態項係屬宏觀層次，有待研究者集體自發自覺地持續按部就班建設，才有可能構建出健全生態系。簡言之，就是健全由理論研究、方法研究、系統評論研究三者構成的上層系統，去影響由應用研究、教學研究組成的下層系統，當理論、方法發展健全，加上系統評論研究協作指出理論和方法的缺失、及改善和發展方向和策略，便能給應用和教學研究者奠定良好基礎和資源支援，當然應用和教學品質及效益就提升。因此面對前述 Delphi 分析問題得採取兩手思維，一者專注於厚植

基礎策略，從有益於Delphi分析有關理論、方法與評論三個角度攜手並進，此有待長時間持續累積，二者關注於巧用改進策略，從增益教學、應用兩角度協作互助，此必須即時推動。

肆、Delphi分析應用改善探討

以下揀選前述討論的若干問題，提出一些改善做法的思考。

一、發展更契合情境需求的Delphi分析學理和方法，完善Delphi分析理論和應用做法

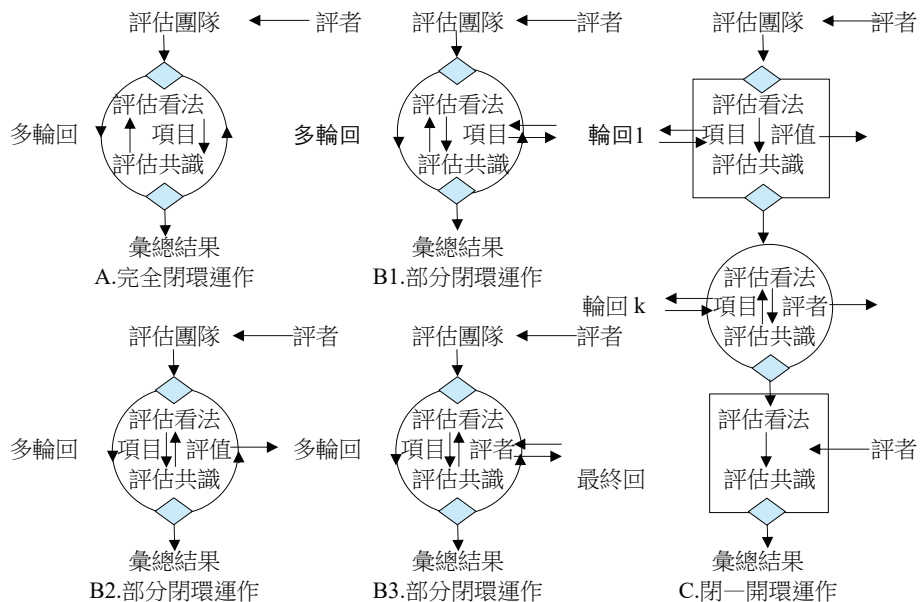
1. 深思Delphi分析基本原則調整的可行性和衝擊

如前所述Delphi分析是以匿名、循環、控制回饋、統計總合等做法為原則，實現民主參與決策、匯聚眾人智慧的理念。因此觀察衍生的眾多Delphi分析方法基本上都保持匿名等精神，調整修改做法圍繞在改變取得資料方法如1. 借助網絡收集評估看法和提供回饋資訊，以加速執行效率；2. 採取更合理有效評估共識方法，如使用模糊數集（fuzzy set）評定，以縮短運作多輪回諮詢的次數；3. 採用更多元統計方法，除了常用的標準差、IQR等簡單統計外，納入使用Kendall w、Kappa係數、甚至Cronbach α ，以增進共識性評估的合理性。基本上前述做法主要針對循環、控制回饋和統計總合，至於匿名則做法維持不變，乃避免因為開放討論難免造成評者身分透漏，因而產生團體壓力，發生評者從眾和違逆專業性考量行為，進而影響最終結果的合理性和專業性，又在最終階段採取開放討論方式，似有打破匿名原則的跡象。檢視文獻，亦可見論者（Firth, *et al.*, 2019）提出類似的討論聲音，包括眾多修正式Delphi分析將Delphi定義擴展和模糊化、Delphi分析評估共識和異識的平衡考量等，因此值得關心和思考改善Delphi分析是否可能一併衝擊了其堅持的理念和原則，反而造成傷害。

大抵上，Delphi分析係以閉環（closed circle）運作為主，圖8的A模式（完全閉環運作）就是基本型，係選取評者組評估團隊，然後提出項目供評估看法，經由統計分析評估看法的共識性，若達評估標準，則進入最終階段，否則提供全部評者評估看法彙總資訊，再據以進行新一輪評估，直至確認評估已達共識為止，此運作模式不刪增項，至多潤飾項目陳述。一般研究多採取B1模式，允許評者提出增減項要求，固有時導致各輪回的評估項數和內容可能不盡相同。而B2模式考量極端評估值對於評估共識性達成的衝擊，允許刪去被判為偏離值的評者看法，以促使這些扮演噪音角色的評值無法影響整體項目評估的共識性，與B1

模式不同，此為分析者發動刪項呼籲，具有合理性是有採用價值。B3 模式則允許除去被認為固執的異議評者所有評估資料，通常若無令人信服的證據，如檢視其在多輪回評估看法都明顯異於其他人，且無合理理由可支持其評估作為，故勿輕易啟動本模式。至於 C 模式屬於採取多輪回調整做法，可見諸使用混合方法取向的 Delphi 分析如 Firth 等人（2019），不僅允許增刪項、調項和潤項敘述，也認可刪去異常評值、調整參與評估的評者，其夾雜循環圈運作和單向運作的方式，最後一輪回係評者針對項目進行開放性討論，不關注獲取共識，可增進討論深度及廣度和最終決策品質，但是本模式並非適合所有取向的 Delphi 分析。簡言之，B1 模式為多數研究採用，為增進最終決策品質和應用效益，可思考採用 B2 或 B3 模式，而 C 模式較適合政策推演或預測規劃，以最終階段開放討論來增益最後提案的品質，其是否較佳於傳統上採取的模式 A 或 B 是值得進行效益和可能問題方面的探討。

圖8
Delphi分析閉開環運作模式比較



註：菱形框指進入下階段的決策關卡。圓圈指多輪回閉環循環運作，矩形框指開環單輪回單向運作。

2. 發展考量噪音影響的 Delphi 分析共識評估方法，以快速和合理偵測極端異識者

如前所述 CV 係數較諸標準差是較佳檢視共識性的係數，一般進行 Delphi 分析多未偵測和檢視評者看法屬於偏離值所產生的影響，實有可慮之處。其一，這些具備離值性質的評估值係屬於噪音（noise），會造成共識無法達到合理要求的滿意標準值，致使得進行較多次輪回評估，若有評者一直堅持己見，則很難達成滿意的共識，此時無疑得思考是否研議刪除經辨識為異識評估者（dissensusor）

的評估看法，或是針對偏離值進行合理處理以降低其對共識達成的衝擊，似乎少見討論此問題的論述，其涉及Delphi分析結果的生態效度（ecological validity），指分析結果能夠通則化（generalization）適用於其他類似情境，有待從統計學理方面尋求突破，構思適切的分析係數。其二，乃偵別偏離值是重要課題，研究成果甚多（Rousseuw & Hubert, 2011），但所提區辨做法多半繁瑣，涉及繁雜統計分析，如能提出較簡易操作且辨識效果佳的做法當有必要。

就第二個問題而言，一般常用z分數、IQR來協助。對此，可思考以CV係數為基礎，發展新係數來協助辨識，稱為 CV_{ln} ， $M_{ln} = \frac{\sum \ln(x_i)}{N}$ ， $SD_{ln} = \sqrt{\frac{\sum (\ln(x_i) - M_{ln})^2}{N-1}}$ ， $CV_{ln} = \frac{SD_{ln}}{M_{ln}}$ ，其基本上是先將資料值進行ln轉換，再計算平均數、標準差和CV，其不同於Nakagawa等人（2015）提出用於比較的lnCV。該係數特性是若無較大偏離值資料，通常 CV_{ln} 值 $<$ CV值，當出現較大偏離值時， CV_{ln} 會有放大或縮小偏離值影響的效果，因此造成 CV_{ln} 值 $>$ CV值。觀察表29應用實例可以發現X6~X8有一個極端偏離值，其全距達到4， CV_{ln} 值都 $>$ CV值，反之其他變項X1~X5的 CV_{ln} 值都 $<$ CV值，對此提出 CV_{ln} 比率（ratio of CV_{ln} to CV, RCV_{ln} ）， $RCV_{ln} = \frac{CV_{ln}}{CV}$ ，當 $RCV_{ln} > 1$ 表示資料值存在極端偏離值（指資料值的z值 > 2 或 < -2 ），而 RCV_{ln} 值 < 1 就顯示可能不存在極端偏離值，在Delphi分析評估共識時能毋庸考慮極端異議評者（extreme dissensusor）的影響。因此，能使用 CV_{ln} 於Delphi分析，惟其對於敏感於極端的偏離值，若評值的偏離效果較小如其z值在 $[-2, 2]$ 之間則 CV_{ln} 較無法敏銳地反映。

表29
 CV和 CV_{ln} 應用實例比較

資料	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1	3	4	4	4	1	1	3	4
2	4	5	3	3	1	1	4	3
3	4	4	5	4	1	1	4	4
4	4	4	4	4	1	1	4	4
5	2	3	2	4	2	2	2	4
6	5	5	4	5	2	2	5	5
7	5	4	2	2	2	2	5	2
8	4	4	3	4	2	2	4	4
9	4	4	4	4	2	2	4	4
10	4	4	4	2	2	5*	1*	1*
全距	3	2	3	3	1	4	4	4
M	3.900	4.100	3.500	3.600	1.600	1.900	3.600	3.500
SD	0.876	0.568	0.972	0.966	0.516	1.197	1.265	1.179
CV	0.225	0.138	0.278	0.268	0.323	0.630	0.351	0.337
CV_{ln}	0.199	0.101	0.257	0.252	0.861	1.021	0.415	0.410
RCV_{ln}	0.887	0.728	0.926	0.940	2.667	1.620	1.182	1.217
影響% ^a	-0.113	-0.272	-0.074	-0.060	1.667	0.620	0.182	0.217

註：*是極端偏離值，其z值 > 2 或 < -2 。^a影響%=(CV_{ln} -CV)/CV，數值越大表示 CV_{ln} 影響效果越大。

3. 結合魚骨圖 擴展質性Delphi 分析共識評估效益

Ishikawa (1974) 提出特性要因圖 (cause-and-effect diagram, CE diagram), 或稱魚骨圖 (fishbone diagram), 使用於品質管控 (quality control), 就 Delphi 分析而言亦可適用。論述提出甚多魚骨圖構形, 此處歸納提出溯因型、解析型、對立型、影響型等四類有益於 Delphi 分析運用者, 見圖 9 說明。溯因型係指陳造成問題 (或災難、危機、衝突等) 發生的子要素項, 包括事件、政策、法規、原因等, 其設定左項為因 (cause), 右項為果 (effect), 說明因與果的影響關係, 為最早提出和常見採用的形式。解析型則闡述構成理念、想法、事務、器物、制度、規定等內涵項, 其設定左項為主要的或基本的概念 (main or bases concept), 右項為構成內容 (content), 說明概念和內容的隸屬關係。對立型則說明兩組或多組具對立關係的概念或問題, 如危機和解決方案、阻力和助力、支持和反對等, 利於瞭解造成對立的癥結點。影響型是從多時間點、多地域、多個體等角度, 組成多個溯因型、多個解析型或是兩者所形成, 可以更宏觀、全面地或完整的知悉概念、想法、事件、爭議等演變或衍生的複雜因果關係網絡。大抵來說, 圖 9 所展示的只是雛形, 使用者能擴展或變形, 產出更多複合形式。

圖 9
魚骨圖類型及表現形式

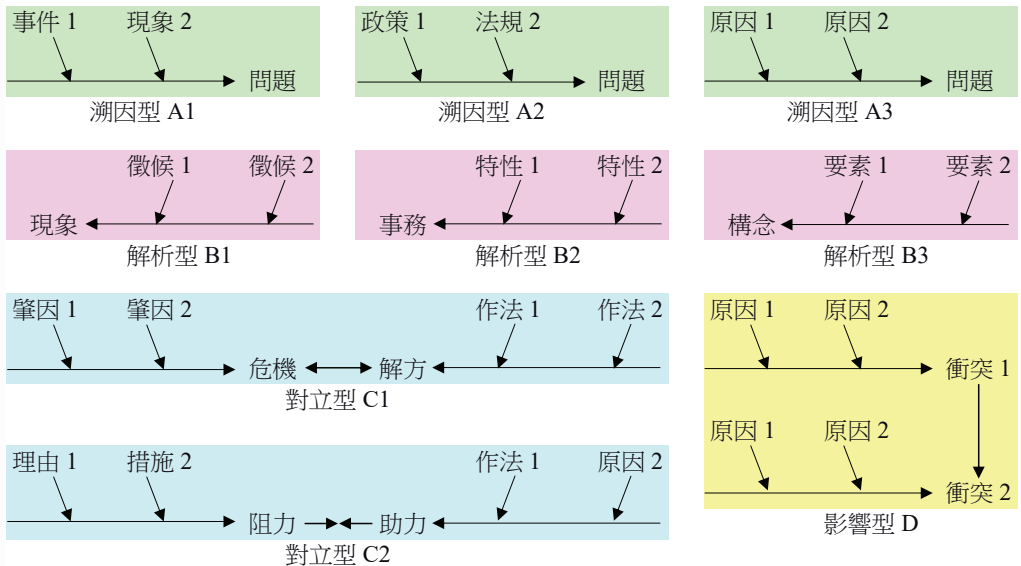
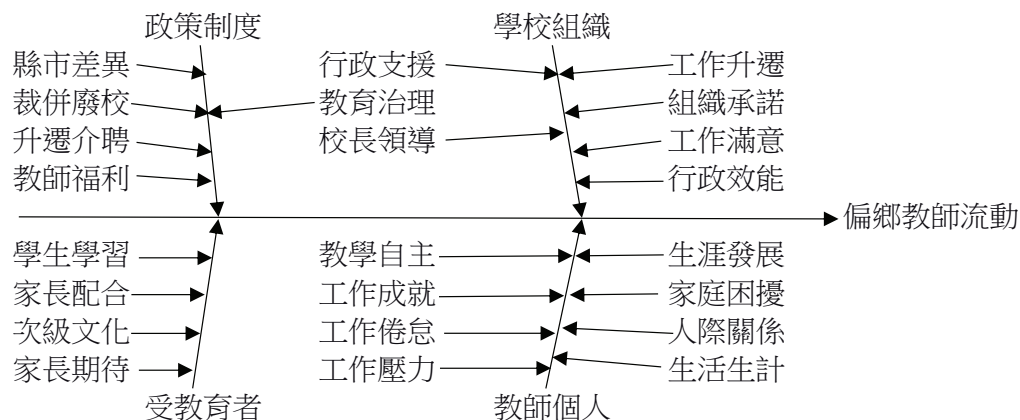


圖 10 顯示一個說明關於偏鄉教師流動影響因素的魚骨圖示例, 指出歸納自研究文獻屬於四個向度的影響因素, 包括政策與制度、學校組織、教師個人、受教育者 (指學生) 等向度。Delphi 分析應用該圖時, 在輪回 1 可要求評者對該圖所列向度和項目的影響因素, 提出增項、刪項、移項、修項 (指修改或潤飾項目

敘述文字)等回饋意見，輪回2再針對提供的新修改魚骨圖提出諮詢意見，如此持續提出圖例、蒐集回饋、修改圖例等工作循環，直至無修改必要為止。後續亦能要求評者對各項目給予評分值，然後彙總以平均值做為各向度和各項目的代表性綜合分數，據以評選最優項目，此為先質性後量化的混合方法Delphi分析做法。

圖10
 偏鄉教師流動影響因素之魚骨圖示例



就Delphi分析而言，進行爭論Delphi、政策Delphi時，可將爭論議題、政策課題加以分解為多個子部分及內容項，讓評估者表達其對各子議題、政策相關事項彼此隸屬或關聯關係的看法，能夠補充量化Delphi分析只能針對既定架構內的小項目提供看法，可能無益於取得最優解或滿意解的缺失，加速共識性達成和提升評估結果滿意度。若就應運用Delphi分析建構指標或解析理念構成要素時，配合使用魚骨圖，可讓評估者針對指標項、向度項和構念項三者之間隸屬關係提出定位看法，能獲取類內容效度的效益，並提供考驗構念效度的參考架構，可謂甚具採用價值。

簡言之，應用魚骨圖可有兩個時機：一是開始收集對項目的評估意見前，做為前置作業或試探性研究，可使用半開放性諮詢做法，即提供評者一張自繪的魚骨圖，讓評者參考和提出改進意見，包括刪項、增項、移項、改項（指修改項目名稱），再彙總評者意見，提出初步架構，並說明異議處的處理做法，或者再做一次徵詢，以提出妥協版本架構，做為後續實施量化評估的基礎依據；也就是說增加一個輪回來處理待評估事項架構合理性或最佳化的問題。二是在實施量化評估時一併徵詢評估者看法，惟此作法亦遭到魚骨圖諮詢意見指出應增、刪或移項，造成後續實施量化評估考驗共識性時的不確定性大增，影響量化分析共識性的困擾。



4. 選用適當評定量尺值 降低對評估共識性結果的衝擊

採用量化Delphi分析主要使用Likert量尺值協助取得評者的評估看法，已知使用3、5、6、7、9、10或11點量尺值，然而使用量尺值是否影響評估結果的合理性或穩定性，有賴實證考驗。de Meyer等（2019）比較採用3和9點量尺值的評估結果影響，指出9點值適用於收集評估看法過程，3點值適用於決定最終共識時；Lange等（2020）比較採用3、5和9點量尺值對共識性評估和重測信度影響，發現3點值是最佳選擇；Schifano和Niederberger（2025）綜析幾項探討量尺值和共識性評估的研究成果不一，有建議3和9點值最佳、5點值表現不佳。據此觀之，可知使用Likert量尺值的確和Delphi分析評估結果穩定性有關聯，但是何者較佳，前述討論顯示9點和3點值似乎較好，可是5點值卻是最常見採用，因此後續宜參考測量理論發展，優先探討評估量尺值影響的有效方法。

