

心理幸福感量表簡式中文版信效度及測量 不變性：以大學生為樣本並兼論測量 不變性議題*

李仁豪

余民寧**

中山醫學大學心理學系暨
附設醫院臨床心理室

國立政治大學教育學系

本研究目的有二，一是獲得更多有關 Li (2014) 18 題簡式中文版心理幸福感量表的信、效度訊息，二是進行跨樣本交叉效度及跨時間測量不變性的檢定。研究樣本有三套，以大學生為對象，分別命名為樣本一、樣本二及共同樣本，人數分別為 442、349、394 位參與者。以結構方程式模型技術來處理資料。研究發現，符合理論架構的六因素斜交模型獲得良好的適配；在信度方面，分量表內部一致性 α 係數在 .73 ~ .87，總量表高達 .91；分量表間隔 2 ~ 3 週的再測信度介於 .59 ~ .65，總量表則達到 .70。在效度方面，因素負荷量在 .65 ~ .92，具有聚斂效度，而因素間的相關性在 .30 ~ .72，通過區辨效度的檢驗。另外，跨樣本的交叉效度亦獲得一定程度的支持；而測量不變性檢驗亦獲得因素負荷量及截距在跨時間上符合量尺不變性及純量不變性的要求，有利於構念在跨時間上的因素間關係比較及因素平均數比較。

關鍵詞：心理幸福感、交叉效度、純量不變性、測量不變性、量尺不變性

壹、緒論

人類自古以來便努力與周遭的人事物環境進行接觸交流，希望獲得生活上的舒適快樂，並進一步達成個人理想的實現，因此發展出兩種哲學思考方向，一種是對快樂原則的重視，另一種是強調潛能發展的追求。這兩類取向形成兩種幸福感研究主流，在概念上相關但實徵上則不同 (Keyes, Shmotkin, & Ryff, 2002)。無論是何種取向的幸福研究，欲將複雜的幸福構念進行科學化測量是一種挑戰，尤其是良好的心理計量實務，通常要求

* 感謝國科會經費補助 (編號：NSC 102-2410-H-040-009-SSS 和 NSC 102-2410-H-004-191-SS2)。

** 通訊作者：余民寧，11699 台北市木柵郵政 1-433 號信箱，e-mail: mnyu@nccu.edu.tw；(02)2938-7445。

每個構念須有多個測量試題或觀察指標來評估，而主要構念通常又包含數個相關但有區別性的次構念，因此使得部份量表的長度頗為可觀。從內部一致性信度的觀點而言，試題數目多的量表之信度通常不低；從內容效度的觀點而言，量表試題的編寫過程通常須有特定主題之心理學理論為依據，以保證所測量的構念不會偏離預定主題。如此一來，在上述質性效度及量化信度的配合下，曾經出現包含六向度（構念／因素）長達 120 題的心理幸福感量表（Ryff, 1989b），乃至後來的 84 題版（Ryff, Lee, Essex, & Schmutte, 1994）。然而，早在 19 世紀末 20 世紀初，利用因素分析將抽象構念加以捕捉的做法就已經出現（Mulaik, 2010），逐漸成為評估量表建構效度的重要技術，一般稱為探索性因素分析。但是，這種分析方式對於長題數多構念的量表而言，通常無法獲得試題與構念間歸屬一致的驗證結果（Kafka & Kozma, 2002），這對於未能提供較佳因素分析結果之量表而言頗為不利，尤其是心理測驗學強調健全的量表應該提供合理的量化效度證據，以利客觀科學研究的進行。

隨著時代的演進，出現結構方程式模型方法學，可利用預設理論模型將量表所收集到的樣本資料模式化，藉由模型與資料的適配情形及相關參數的解讀，來提供量表建構效度乃至其它信度證據，一般稱為驗證性因素分析。這改善了早期使用探索性因素分析的缺點，例如容易因樣本不同而產生聚焦問題、萃取或轉軸方式選用不一致、因素結構係數與因素組型係數的誤用等等（Mulaik, 2010）。因此，驗證性因素分析改進這些缺點，並提供了許多有利量表信效度檢證的方式（Brown, 2006; Kline, 2011; Richardson, Ratner, & Zumbo, 2009）；例如：測量／因素不變性（measurement/factorial invariance）的檢定，就必須在驗證性因素分析架構下才能同時完整進行。然而，即使如此，在實務上，過多試題仍然使驗證性因素分析模型出現適配指標不佳的可能性大增，理由是，即使是參酌相關理論所編製出的試題，或多或少都可能包含一些非研究預期的構念，這些非研究預期的構念會導致試題測量殘差間出現可能的相關，迫使正確模型與樣本資料的適配度隨著試題增多而越糟。事實上，從驗證性因素分析技術來看，在多構念量表中，每個構念只需兩個試題即可讓模型獲得辨識，但考量可能出現其它實務上的辨識問題，Brown（2006）建議每個構念最少有三個測量（觀察）指標，而 Little（2013）也認為具三個觀察指標的構念就好比是有三條腿的凳子，十分穩當，甚至比四條腿的凳子安穩。雖然 Ryff 與 Keyes（1995）曾將心理幸福感量表精簡為 18 題版本，但卻有信度過低的疑慮。因此，Li（2014）將 Ryff 等人（1994）的 84 題心理幸福感量表翻譯後，在原量表六向度架構下，以中老年人為樣本，重新選出 18 題的簡式中文版本，使用驗證性因素分析模型確認後，獲得較佳的信效度訊息。本研究目的除了以大學生樣本再次驗證此量表的模型適配度以及擴展其它信效度訊息外，將進一步利用驗證性因素分析，確認本量表具有跨樣本的交叉效度（cross valida-

tion)，以及確認本量表在跨時間上具有測量不變性，以利跨時間的量表分數比較具有意義。

遠在古希臘時代就曾經出現兩種與幸福感有關的哲學思想，一個是被稱為伊比鳩魯（Epicurus）的派別，主張生存原則是避免痛苦和尋求快樂，被稱為快樂主義（hedonism）。另一學派以亞里斯多德（Aristotle）為代表，認為理性積極（eudaimonia）在本質上就是愉快的來源，也是善的至高境界；個體會積極追求自身潛能的實現，以達到出色及完美，此派學說被稱為完善論或理性積極的幸福主義（eudaimonism）。這兩派學說的論點持續到十八、十九世紀時，快樂主義的思想與洛克（Locke）等人的經驗主義哲學產生連結，以感官的覺受為經驗來源，形成主觀幸福感 bottom-up 模式的理論依據；另一方面，理性積極的幸福主義則與康德（Kant）等人的理性主義哲學一致，鼓勵個體努力追求理想實現或以道德完美為目標，並主張人類先天就有一種解釋生活經驗的存在，形成主觀幸福感 Top-down 模式的理論依據。值得注意的是，理性積極幸福的追求並不強調逃避受苦，受苦對個體而言可能是達到完美或真理的必經之路（更多這兩種派典的討論請參考 Joshanloo, 2014; Keyes et al., 2002; Ryan & Deci, 2001; Shmotkin, 2011; Waterman, 1993）。

近代 Diener（1984）提出主觀幸福感（subjective well-being）這個術語，而「主觀」兩字強調幸福應該是來自個體自身主觀感受，而非來自他人客觀衡量，可說是在某種程度上承襲了伊比鳩魯學派的快樂主義傳統。而 Diener、Emmons、Larsen 和 Griffin（1985）接著編製了 5 題的生活滿意度量表（Satisfaction With Life Scale, SWLS），將認知和情感的成分納入考量，被視為是簡單且常用的主觀幸福感測量工具，此後主觀幸福感逐漸形成一個重要的學術主題，例如 Wu 與 Yao（2006）曾使用臺灣大學生樣本對此量表進行性別的因素不變性檢定。然而，由於主觀幸福感的測量過於模糊，包括快樂、生活滿意度、正負向情感、生活品質、士氣、自尊等等，都常用來單獨地作為對主觀幸福感的量測或定義（如 Stanton, Sinar, Balzer, & Smith, 2002; Steel, Schmidt, & Shultz, 2008 所做的分類），Salmela-Aro 與 Schoon（2005）亦建議相關幸福感術語的使用須有一致性，避免紛雜。

而早在上世紀 80 年代末期，Ryff（1989a, 1989b）也認為主觀幸福感的定義不夠嚴謹，沒有理論基礎來作為主觀幸福感的內涵，因此，她將過去傳統主觀幸福感過於模糊的一面重新予以嚴謹的定義確認，偏重將主觀幸福感重新定義為較理性積極的（eudaimonic）層面，因而提出「心理幸福感」（psychological well-being）術語，使得未來從事主觀幸福感的研究可以有較為精確的解釋。值得注意的是，從 Ryff 定義所發展出來的量表依然是需要受試者以自陳方式進行主觀的評價，因此，仍然可視為是主觀幸福感的測量，只是關注焦點在受試者理性積極的心理功能層面上，承襲了亞里斯多德、康德學派的論點。

Ryff 定義的心理幸福感整合了更早期學者所提出的關於正向心理健康功能的概念，

包括發展心理學中 Erikson 的心理社會發展階段 (psychosocial stages) (Erikson, 1959)、Bühler 的基本生活傾向 (basic life tendencies) (Bühler, 1935; Bühler & Massarik, 1968)、Neugarten 的成人及老年人格變化 (personality change) (Neugarten, 1973) 等觀點來闡述幸福是橫跨生命全程 (life span) 的持續成長軌跡；另外，從臨床心理學中，藉由 Maslow 的自我實現 (self-actualization) 概念 (Maslow, 1968)、Allport 的成熟 (maturity) (Allport, 1961)、Roger 的功能發揮的人 (fully functioning person) (Roger, 1961)、Jung 的個體化 (individuation) (Jung, 1933) 以及 Jahoda 的正向心理健康 (Jahoda, 1958) 等觀點提供對幸福更一步的描述。Ryff 整合上述觀點為心理幸福感的理論結構，其結構包含以下六個向度 (Ryff, 1989a, 1989b)：(1) 自我接納 (Self-Acceptance, SA)：對自我擁有正向的態度；承認並接受自我的多個面向，包括好的及壞的特質；對於過去的生活有正向感；(2) 正向關係 (Positive Relations with Others, PR)：與他人有溫暖的、令人滿意、信賴的關係；關心他人的福利；能夠有強烈的同理心、情感與親密性；(3) 自主 (Autonomy, AU)：自我決定及獨立；能夠抵抗社會壓力造成的想法及行動；從內在調整行為；用個人的標準來評價自我；(4) 環境精熟 (Environmental Mastery, EM)：對於環境的掌握具有精熟及勝任感；控制大量複雜的外在活動；有效利用周遭的機會；能夠選擇或創造適合個人需求及價值的背景脈絡；(5) 生活目的 (Purpose in Life, PL)：生活有目標及方向感；感覺目前及過去的生活有意義；堅持具有生活目的之信念；生活有目的及目標；(6) 個人成長 (Personal Growth, PG)：有持續發展的感覺；視自我在成長及擴展中；對新經驗開放；有實現潛能的感覺；自我及行為隨著時間進步；改變的方式能反映出更多的自我知識及效能。更詳細地說，自我接納是心理健康的中心特徵，也是自我實現、功能發揮及成熟者的特點；生命全程理論也強調接受自己及過去生活的重要性。正向關係中的愛人能力是心理健康的中心成分，自我實現者對所有人類都擁有強烈的同理及情感，而成熟者的標準之一則是能夠與他人建立溫暖關係，且成人發展階段理論也強調親密關係的發展任務。自主被認為與自我實現者表現出自治功能及抵抗文化適應有關，功能發揮的人會顯示出內控評價的特徵，而個體化則涉及從傳統中釋放，不再抱持集體的恐懼、信念及大眾主義，這種轉向內在的過程在人生晚年也被生命全程發展論視為是一種免於常模掌控日常生活的自由感。環境精熟涉及成熟者應具有能參與外在於自我的明顯活動之能力，生命全程發展論也要求個體需具有操控複雜環境的能力，而擁有能力去選擇或創造適合自己精神條件的環境也被定義為是心理健康者的特徵。生活目的是心理健康者具有的一種對生活懷有目的及意義感的信念，成熟的定義強調對生活目標的清楚理解、具方向感及企圖性，而生命全程發展論認為生命過程中會有不同的目標追求。個人成長需要自我實現及瞭解自身潛能，而非停滯不前，因

此對經驗的開放也是功能發揮者的特徵，且生命全程發展論也強調持續成長並面對不同階段的挑戰或任務。

就實務操作層面來說，Ryff 所編製的心理幸福感量表最原始版高達 120 題 (Ryff, 1989b)，隨後也出現稍短的 84 題以利用使用 (Ryff et al., 1994)，此外，為了提供驗證性因素分析的效度訊息，Ryff 與 Keyes (1995) 將之精簡為 18 題的版本。然而，此版本雖然在一階六因素結構或包含二階一因素結構的模型適配度上勉可接受，但各向度內部一致性信度 α 係數太低，僅在 .33 ~ .52 之間，其他研究亦發現此版本相似的問題 (余民寧、謝進昌、林士郁、陳柏霖、曾筱婕, 2011; Cheng & Chan, 2005; Clarke, Marshall, Ryff, & Wheaton, 2001; Van Dierendonck, 2004)，以致於隨後各國研究者相繼進行各種精簡版的發展工作，例如 Fernandes、Vasconcelos-Raposo 與 Teixeira (2010) 選出 30 題的葡萄牙文版，Van Dierendonck (2004) 選出 39 題的荷蘭文版；Cheng 與 Chan (2005) 選出 24 題的中文版。然而，這些版本仍然都有一些信效度的問題存在，例如 Cheng 與 Chan (2005) 根據在儘量提高各向度信度的原則下進行選題，但卻仍然導致因素負荷量太低的問題，其中有三題在 .30 ~ .36，五題在 .41 ~ .49，僅一題達到 .70，他們認為部分試題仍有重新修飾的必要。因此，Li (2014) 重新翻譯 84 題版本，他認為 84 題版在質性編題的過程中已有一定的專家效度 (Ryff, 1989b; Ryff et al., 1994)，所以，透過幾位專家進行倒譯程序後，直接利用驗證性因素分析進行選題，在考量內部一致性信度及因素負荷量後，選出 18 題的中文新版本，各題因素負荷量在 .55 ~ .80，各分量表內部一致性信度 α 係數在 .60 ~ .74。以下分析將針對這個版本在大學生樣本上的信效度、跨樣本測量不變性 (交叉效度)、以及跨時間的測量不變性進行檢定。

貳、研究方法

一、研究樣本及設計

樣本來源為中部一所大學修習心理學相關課程的大學部學生，主要以大學一、二年級學生為主，課程包括普通心理學、視覺心理學、醫學心理學、心理疾病學、心理統計學、心理測驗學等，請心理系任課老師幫忙施測及宣導，說明學生有權選擇不填問卷且與個人成績無關，但願意參與的學生可獲得高級原子筆的酬賞。共施測兩次，間隔 2 ~ 3 週的時間，問卷回收後，由於部分學生同時修習好幾門這些課程，因此，先將同一次施測期間在不同課程重覆填寫的少數參與者問卷排除。

統計後發現，第一次施測共獲得 442 位有效樣本，本研究將之稱為樣本一資料，其中男性 146 位，女性 296 位；第二次施測共獲得 743 位有效樣本，在扣除了與樣本一共同的樣本後，獲得 349 位未重複的樣本，本研究將之稱為樣本二資料，其中男性 158 位，女性 191 位。兩次施測都有填寫的學生共 394 位，本研究將之稱為共同樣本，其中男性 128 位，女性 266 位。樣本一及樣本二的參與者未重複，而共同樣本則是跨時間的同一群大學生樣本，這三個樣本之間的參與者都未重複，故總共有 1185 人。樣本一施測的量表包括心理幸福感量表 18 題版、生活滿意度量表、正負向情感量表。樣本二施測的量表包括心理幸福感量表 18 題版、自尊量表。對共同樣本而言，上述量表都會施測到，且心理幸福感量表更是跨時間被重複施測。

研究樣本的規劃是根據本研究的兩個目的所進行的設計考量，一是針對 Li (2014) 的 18 題心理幸福感版本在大學生樣本上的信效度檢驗，乃利用樣本一及樣本二分別進行驗證性因素分析，以提供適配度、因素負荷量、聚斂及區辨效度、組成信度等訊息，並利用共同樣本提供間隔 2 ~ 3 週的再測信度。此外，也針對不同樣本所施測的其它量表，提供效標關聯效度及區辨效度等證據。二是針對 Li 的心理幸福感版本檢定其交叉效度及測量不變性，利用多群組結構方程式模型檢驗該量表各參數矩陣或向量在組間的恆等性，以呈現樣本一及樣本二間具有交叉效度的程度，以及在共同樣本上的跨時間測量不變性結果。

二、研究工具

(一) 心理幸福感量表 18 題版

Ryff 等人 (1994) 的心理幸福感量表共 84 題，即每向度 14 題，該量表涵蓋六個向度，即六個分量表，以李克特氏六點量尺的形式來計分，1 代表完全不符合、2 代表相當不符合、3 代表有點不符合、4 代表有點符合、5 代表相當符合、6 代表非常符合。本研究透過三位心理學家及一位英語專家的協助，並參考過去的翻譯版本 (董智慧, 1998)，將 Ryff 提供的 84 題心理幸福感量表重新翻譯完成，同時也進行倒譯的程序，針對差異較大的部分進行檢視及討論。

在翻譯過程中，儘量按照字面句子的用語來翻譯成對應的中文，若按照原句的意義來翻譯可能會有許多句子的語意太過接近而無法區辨，即便如此，在 84 題翻譯完成後，再請幾位熟識的成年人試答，仍有許多人不耐煩的表示：「怎麼都在問同樣的問題？」，這也是本研究在翻譯此量表時感受到的現象，即同一向度 14 題中，嚴格來講，各題內容雖然保有必要的差異性，但這些語句差異性對於一般成年人而言，仍然感到十分雷同，這與一般認知心理學教科書提到的人類訊息處理原則頗為一致，亦即以意義性來收錄訊息而非純

粹按字面敘述來記憶。因此，對於意義十分接近的試題一直出現，受試者會容易煩躁，故有縮減試題數的空間。另外，有部分成人反應：「問卷試題問得很奇怪」，推測其原因，可能是部分試題因東西文化不同而產生的用語差異。因此，將其中幾個較爭議的句子改採華人較常用的表達方式來翻譯。例如：原量表第 1 題「Most people see me as loving and affectionate」，若照字面來翻譯「loving」、「affectionate」這兩個形容詞，可能很不符合華人的常用語，華人很少會用「可愛或令人喜愛、熱情」來形容成年人，更何況是對男性。因此，我們討論後，在不違反該向度「正向關係」(PR)的原則下，決定將之翻譯為更常見的用語，即「大多數的人認為我是平易近人的」。又如原量表第 82 題「There is truth to the saying that you can't teach an old dog new tricks」，若按照字面翻譯，可能會引起中老年人強烈的反感，導致量表信效度降低。因此，在「個人成長」(PG)這個向度下，將之翻譯為「<人要是老了，就學不會新的東西>，我覺得這句話是對的」，盡量減少歧視成份的出現，應較為適當可行。

最後，在不違反 Ryff 心理幸福感六向度的原則下，利用中老年人為樣本獲得新的 18 題心理幸福感量表 (Li, 2014)，每個向度三題，依序為 PR、AU、EM、PG、PL、SA，信效度訊息如文獻探討所述。目前，本研究將再以大學生為樣本進行信、效度及測量不變性的檢測，並於附錄一摘要 1185 位大學生樣本合併後的描述統計。

(二) 生活滿意度量表

本研究採用 Diener 等人 (1985) 發展的 5 題生活滿意度量表 (Satisfaction With Life Scale, SWLS)。生活滿意度量表是傳統主觀幸福感定義下常使用的測量工具。SWLS 為單一因素量表，以李克特式 7 點量尺計分，從非常不同意 1 分到非常同意 7 分。因素負荷量在 .61 ~ .84，它與測量主觀幸福感的其它量表之間有中等程度相關，具有良好的效標關聯效度；而該量表試題與總分的相關在 .61 ~ .81，具有良好的內部一致性。SWLS 在樣本一的內部一致性 α 信度係數為 .87，因素負荷量在 .56 ~ .88，在本研究中主要是用來提供 18 題心理幸福感量表的效標關聯效度及區辨效度證據。

(三) 正負向情感量表

本研究採用 Bradburn (1969) 編製的平衡式情感量表 (Affect Balance Scale, ABS) 共有 10 題，包含 5 題正向情感，5 題負向情感。在內部一致性信度上，正向情感的題間相關在 .19 ~ .75，負向情感的題間相關在 .38 ~ .72，間隔三天的再測信度分別為 .83 及 .81；且正負情感兩個分量表間的相關低於 .10，顯示區辨效度良好。該量表中文版取自楊中芳譯 (1997/1991)，原量表以是／否二元方式計分，本研究將之改為非常不符合到非常符合的

4 點量尺來計分，較適合連續構念的測量。ABS 在本研究樣本一的內部一致性信度 α 係數分別為 .66 及 .63，因素負荷量除了負向情感的最後一題為 .25 外，其餘都在 .43 ~ .71。ABS 也將用來提供 18 題心理幸福感量表的效標關聯效度及區辨效度證據。

(四) 自尊量表

本研究採用 Rosenberg (1965) 編製的自尊量表 (Self-Esteem Scale, SES)，係奠基在自我接納及自我價值感上所發展出來的簡短量表，為單一因素共 10 題，以李克特式 4 點計分，分別為非常不同意、不同意、同意、非常同意，其中第 3、5、8、9、10 題為反向題。該量表中文版亦取自楊中芳譯 (1997/1991)，其內部一致性 α 係數在 .77 ~ .88，間隔 1 ~ 2 週的再測信度在 .82 ~ .85。而 SES 與自信的相關為 .65，與學業自我概念等方面的相關在 .18 ~ .40，具有良好的效標關聯效度。SES 在樣本二的內部一致性 α 係數為 .88，因素負荷量在 .43 ~ .79，在本研究中也用來提供 18 題心理幸福感量表的效標關聯效度及區辨效度證據。

三、統計分析

本研究利用結構方程式模型 (Structural Equation Modeling, SEM) 中的測量模型，採用 LISREL 8.8 版軟體，以預設的最大似估計法來進行參數估計。分別驗證樣本一及樣本二的模型適配度及信效度，並使用多群組 SEM 進行測量不變性或稱因素不變性/相當 (factorial invariance/equivalence) 檢定，來評估 18 題心理幸福感量表在樣本一及樣本二的測量不變性 (交叉效度)、以及在共同樣本跨時間的測量不變性結果。

以跨時間測量不變性 (longitudinal measurement invariance) 檢定為例，它關心的重點是構念在時間上觀察到的變化是真正的改變，抑或因測量本身隨時間發生改變所致 (Brown, 2006)；亦即，若構念的平均數、變異數及共變數隨時間出現變化，是否僅僅是因為測量工具的因素結構、因素負荷量及測量截距等等隨時間發生變化而導致的。為了評估前者，須確認後者的測量不變性程度。Byrne、Shavelson 與 Muthén (1989) 曾將涉及觀察指標與潛在變項間關係的參數等化限制視為真正的測量不變性，而將潛在變項本身的不變性檢定稱為結構不變性 (structure invariance)，前者包括因素結構不變性 (factor structure invariance)、因素負荷量不變性 (factor loading invariance)、截距不變性 (intercept invariance)、測量殘差不變性 (measurement error/residual invariance)，後者包括潛在變項的平均數、變異數及共變數不變性。本研究認為可將前者視為狹義的測量不變性，而將上述兩類不變性檢定合併視為廣義的測量不變性，這樣有利於本文後續的討論。

在測量不變性的檢定中，因素結構不變性又常稱為構型不變性 (configural invari-

ance)；而因素負荷量不變性又常稱為量尺不變性 (metric invariance)；且截距不變性又常稱為純量不變性 (scalar invariance)，這三種不變性檢定需依序進行且成立後，後續的結構不變性檢定才有意義；亦即量表的使用者必須先瞭解自己正在測量什麼東西，之後才能檢定所測量到的東西之間的關聯 (Anderson & Gerbing, 1988; Vandenberg & Lance, 2000)。構型不變性是指在各時間點或組別間的因素個數相等以及因素組型一致，通常被視為是基準模型，以利後續等化限制依序建立模型巢套關係之參照依據。量尺不變性是指因素負荷量的值在組間是相等的，被視為是「弱的測量不變性」(weak measurement invariance) (Meredith, 1993)。由於因素負荷量是潛在變項與觀察指標間的迴歸斜率，因此，若潛在變項每改變一個單位，對應的觀察變項改變量之期望值在組間是相等的，此時量尺單位 (scaling units) 在兩組可視為等值 (Schmitt, 1982)，故而因素間的關係在跨時間或跨組別的比較是有意義的。而純量不變性是指上述迴歸關係中的截距項在跨時間或跨組別間具不變性，被視為「強的測量不變性」(strong measurement invariance) (Meredith, 1993)，由於截距是指當潛在變項為 0 時的觀察指標期望值，因此，當截距在各組不變，可視為量尺原點在各組是相同的 (Hair, Black, Babin, & Anderson, 2010)，此時因素平均數 (潛在變項平均數) 在各時間點或組別間的差異比較是有意義的。

至於測量殘差不變性，則被視為是「嚴格的測量不變性」(strict measurement invariance) (Meredith, 1993)。測量不變性是否須包含測量殘差不變性，相對而言，不若因素負荷量及截距不變性來得重要 (Brown, 2006)。量尺不變性及純量不變性足以保證測量不變性，此時，具有相同水準的構念將預期會有相同的測量 (觀察) 變項分數期望值，因為測量殘差項的變異雖然不相等，但其平均數即期望值為 0 (Little, Card, Slegers, & Ledford, 2007)。因此，當量表達到量尺及純量不變性，若觀察變項平均數在組別間有差異，可推論出主要是因為潛在變項平均數在組間的差異所導致，而非由無關的測量殘差所造成的 (Brown, 2006; Little, 2013)。另外，在分析操作上，由於本研究使用多群組而非單一組 SEM 來進行分析，因此，兩個時間點或組別間相同試題的測量殘差之相關不需因方法學效應 (method effect) 而自由估計。

在測量不變性的檢定中，值得一提的是部分測量不變性 (partial measurement invariance) 的議題。所謂的部分測量不變性，是指在限制等化特定某類參數矩陣或向量時，由於無法達到合理的適配標準，因此在仍可滿足後續等化限制模型進行的前提下，僅需限制部分參數具不變性即可。特別是針對量尺不變性及純量不變性這兩種檢定而言，每個潛在變項至少需兩個因素負荷量及兩個截距項具有不變性，這樣才能滿足接續結構不變性的檢定條件，讓潛在變項平均數、變異數及共變數在跨組或跨時間的比較仍有意義

(Baumgartner & Steenkamp, 1998; Hair et al., 2010)。有些學者甚至認為在量尺不變性限制中，除了參照指標外，每一個因素僅需一個觀察指標受到限制即可，其餘皆可放寬(Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998)。然而，Vandenberg 與 Lance (2000) 建議僅限制兩個參數的情形最好是在每個因素僅有少數觀察指標的條件下使用。而 Little (2013) 的相似建議則是，若大部份的觀察指標 (例如三分之二) 達到不變性，則有較佳的基礎去探討結構不變性。

此外，由於結構不變性的檢定涉及潛在變項變異數不變性的檢定，因此，在模型辨識時，潛在變項量尺設定不宜使用固定因子法 (fixed factor method)，這樣會讓潛在變項的變異數自動設定為 1，失去跨組潛在變異數比較的意義；宜使用參照指標法 (marked variable method) 來設定潛在變項的量尺單位，但如此一來，會使得每個潛在變項量尺的參照變項之因素負荷量不估計，而無法讓所有因素負荷量都接受不變性檢定，導致檢定結果僅反映出部分量尺不變性；另外，參照指標法會讓其它非參照指標的參數 (如其它因素負荷量或截距) 之不變性不容易達成，因為被強迫設定因素負荷量為 1 的參照指標在各組間可能與潛在變項的關係本來就不相等，因而產生一種人造的潛在變項量尺，導致後續的等化限制無法彰顯組間其它參數真正的異同 (Brown, 2006)，而且使用此法會讓潛在變項的變異及平均數非常依賴所選擇的參照指標，在某些情況下，這種任意選擇參照指標的方式在跨組比較的解釋中會產生實質的差異 (Gonzalez & Griffin, 2001; Little et al., 2007)。同樣地，為了辨識模型平均數結構，讓潛在變項平均數不變性檢定可以進行，不應讓潛在平均數為 0，故參照指標的截距項不估計而設為 0，讓潛在變項平均數可以估計，但這同樣無法讓所有的截距項都接受純量不變性的檢定。近來有學者提倡使用一種稱為效果編碼法 (effect coding method) 的量尺設定方式來解決這個問題 (Little et al., 2007; Little, Slegers, & Card, 2006)，讓所有的因素負荷量及截距都可以列入計算考量，其作法是將每個潛在變項所對應的觀察指標之因素負荷量平均及截距項平均分別設定為 1 及 0。這種方式的主要優點是讓所有因素負荷量及截距的估計是來自所有觀察指標的不偏且最佳的平衡訊息，使得潛在變項的變異數可來自所有觀察指標可信變異的平均貢獻，讓潛在變項的原始量尺更符合實際狀況，比起任意決定參照指標的量尺設定法來得真實且有實際意義 (Little, 2013)，不過這種設定方式在進行因素負荷量及截距不變性考驗時，所增減的自由度數目仍與參照指標法一樣。

在模型好壞的判斷上，傳統單組模型的適配度良窳較不常將 χ^2 作為主要論斷依據，因為卡方統計量容易受樣本大小及變項多寡的影響，因此其它適配指標如 CFI、GFI、RMSEA、SRMR 等等較常被使用。而在測量不變性的檢定中，模擬研究發現 $\Delta\chi^2$ 與其它模型

適配指標差異量進行比較下， $\Delta\chi^2$ 仍然是過於敏感的 (Chen, 2007; Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008)。Cheung 與 Rensvold (2002) 利用模擬研究，比較包含 $\Delta\chi^2$ 在內的 20 個適配度指標差異量的變化情形，建議使用 ΔCFI 、 $\Delta \text{Gamma hat}$ 、 $\Delta \text{McDonald's NCI}$ 這三個較為強韌的統計量來判斷測量不變性。然而 Chen (2007) 比較 $\Delta\chi^2$ 、 ΔRMSEA 、 ΔSRMR 與前述三種指標差異量的穩定性後發現，雖然 $\Delta \text{Gamma hat}$ 的表現與 ΔCFI 一樣好，但兩者相關太高，基於 CFI 的使用較 Gamma hat 廣泛，因此他建議使用 ΔCFI 即可；而 $\Delta \text{McDonald's NCI}$ 的表現並沒有比 ΔRMSEA 及 ΔSRMR 好，因此他並不建議使用。Chen (2007) 分別為 ΔCFI 、 ΔRMSEA 及 ΔSRMR 提出判斷測量不變性的標準，不過他也認為應該以 ΔCFI 為優先考量，因為後兩者在小樣本 (總樣本 300 人以下) 時仍會有過度拒絕等其它問題發生。Cheung 與 Rensvold (2002)、Chen (2007) 針對 ΔCFI 的建議值是 .01，超過這個標準，不變性的假設可能會被拒絕，而 Meade、Johnson 與 Braddy (2008) 則是建議不要超過 .002，但這個標準也許過於嚴格，因此，Little (2013) 根據實務經驗建議狹義的測量不變性可用 $\Delta CFI = .01$ 標準，之後的結構不變性可用 $\Delta CFI = .002$ 或是 $\Delta\chi^2$ 的顯著性 p 值 = .001 (500 人以上大樣本) 或 .005 (200 ~ 500 人小樣本) 來檢視。

參、研究結果

一、模型適配及信效度

(一) 模型適配度

各試題在樣本一的偏態係數介於 -0.62 ~ 0.22，峰度係數介於 -0.20 ~ 0.88；在樣本二的偏態係數則是介於 -0.50 ~ 0.12，峰度係數介於 -0.58 ~ 0.91，絕對值均未超過 1.00，表示各試題分數的分佈非常接近常態分配，有利於驗證性因素分析的進行。在一階六因素斜交結構以及含二階一因素結構之模型適配度的評估上，使用 Kline (2011) 所建議的四種最廣泛使用的適配指標，如表一所示，CFI 及 GFI 符合一般 .90 以上的建議，而 SRMR 及 RMSEA 也都符合 .08 以下的合理適配標準。由此可知，本量表在大學生樣本上具有良好的六向度結構，同時也顯示一個更高階的心理幸福感構念合理存在。

表一 兩樣本的模型適配結果

模型	χ^2	df	p	CFI	GFI	SRMR	RMSEA
一階六因素斜交	374.71	120	< .001	0.97	0.91	0.049	0.069
一階六因素二階一因素	458.32	129	< .001	0.97	0.90	0.061	0.076
一階六因素斜交	288.93	120	< .001	0.98	0.92	0.049	0.064
一階六因素二階一因素	317.43	129	< .001	0.98	0.91	0.057	0.065

註：樣本一442人，樣本二349人。

(二) 量表信效度

兩次樣本的描述統計、各分量表及總量表的內部一致性信度 α 係數，以及共同樣本再測信度整理如表二，而一階六因素斜交模型的其他信效度訊息如表三所示。

表二 大學生樣本在心理幸福感18題六向度上的描述統計及信度

量表或變項	樣本一		樣本二		再測信度 r
	平均數 (標準差)	信度係數 α	平均數 (標準差)	信度係數 α	
正向關係PR	13.29 (2.25)	.73	13.49 (2.28)	.78	.62
自主AU	12.19 (2.53)	.78	12.18 (2.35)	.75	.64
環境精熟EM	12.85 (2.33)	.85	12.82 (2.20)	.80	.59
個人成長PG	14.67 (2.42)	.87	14.59 (2.30)	.86	.60
生活目的PL	13.17 (2.64)	.83	12.90 (2.68)	.84	.65
自我接納SA	12.10 (2.44)	.81	12.29 (2.46)	.82	.63
總量表	78.33 (10.81)	.91	78.37 (10.79)	.92	.70

表三 一階六因素斜交模型的因素負荷量及信度

樣本	試題	正向關係 (PR)	自主 (AU)	環境精熟 (EM)	個人成長 (PG)	生活目的 (PL)	自我接納 (SA)
一	甲	.65	.72	.81	.75	.88	.73
	乙	.74	.82	.86	.89	.79	.80
	丙	.68	.68	.75	.86	.72	.78
	組成信度	.73	.79	.85	.87	.84	.81
	平均潛在變異萃取	.48	.55	.65	.70	.64	.59
二	甲	.69	.73	.80	.77	.86	.70
	乙	.74	.73	.84	.86	.81	.92
	丙	.78	.67	.62	.84	.75	.71
	組成信度	.78	.75	.80	.87	.85	.82
	平均潛在變異萃取	.54	.51	.58	.68	.65	.61

註：甲、乙、丙在PR向度上分別代表對應84題版的第19、49、67題；在AU向度上分別代表第50、68、80題；在EM向度上分別代表第3、21、39題；在PG向度上分別代表第28、64、70題；在PL向度上分別代表第47、53、59題；在SA向度上分別代表第6、12、78題。

各試題因素負荷量的效度訊息、以及衍生計算出的組合信度 (composite reliability) 都呈現不錯的結果，僅「正向關係」向度在樣本一的平均潛在變異萃取量稍低於 .50 的標準。與過去研究對照下，Clarke 等人 (2001) 獲得的因素負荷量有一半不到 .50，而 Cheng 與 Chan (2005) 有三分之一不到 .50，而本研究的因素負荷量大都在 .65 以上，甚至較 Li (2014) 使用中老年人為樣本的結果還要好，因此，本量表具有一定的聚斂效度 (convergent validity)。就信度而言，各分量表信度 α 係數至少在 .73 以上，總量表的 α 係數亦高達 .91 以上，優於過去相關的研究 (如 Cheng & Chan, 2005; Clarke et al., 2001; Fernandes et al., 2010; Ryff & Keyes, 1995)，顯示不管是分量表或總量表都具有不錯的內部一致性。此外，過去研究未曾提出短題版的再測信度資料供參照，本量表提供間隔 2 ~ 3 週的再測信度，其值介於 .59 ~ .65，由於再測信度法又會比內部一致性信度法將更多的

影響因素列入測量誤差中 (Murphy & Davidshofer, 2005)，因此這對於每向度只有 3 題的分量表而言，仍頗具一定的穩定性。

表四 心理幸福感量表的六向度相關情形

樣本	變項	正向關係 (PR)	自主 (AU)	環境精熟 (EM)	個人成長 (PG)	生活目的 (PL)	自我接納 (SA)
一	PR	1.00					
	AU	.21 (.30)	1.00				
	EM	.43 (.56)	.43 (.50)	1.00			
	PG	.53 (.65)	.35 (.40)	.49 (.56)	1.00		
	PL	.43 (.55)	.39 (.44)	.53 (.62)	.60 (.68)	1.00	
	SA	.48 (.63)	.43 (.54)	.59 (.70)	.41 (.47)	.49 (.56)	1.00
二	PR	1.00					
	AU	.39 (.51)	1.00				
	EM	.44 (.58)	.49 (.62)	1.00			
	PG	.59 (.72)	.50 (.59)	.46 (.55)	1.00		
	PL	.54 (.64)	.46 (.57)	.55 (.66)	.59 (.67)	1.00	
	SA	.50 (.62)	.44 (.56)	.47 (.59)	.41 (.48)	.50 (.57)	1.00

註：表內的皮爾森積差相關係數皆達至少 .001 顯著水準，括弧內為潛在變項間相關。

表四呈現兩樣本分別在六向度分量表的相關性及區辨性，樣本一分量表間的相關係數在 .21 ~ .60，潛在變項（因素）間的相關則在 .30 ~ .70；樣本二分量表間的相關係數在 .39 ~ .59，因素間的相關則在 .51 ~ .72，由於因素間的相關是排除測量誤差後獲得的校正值，因此因素相關會比觀察分數相關高。儘管這些相關不算低，但經計算後發現，任兩個構念相關的平方皆小於其對應構念的平均潛在變異萃取 (Fornell & Larcker, 1981)，顯示各向度間仍具有一定程度的區辨效度 (discriminant validity)。與過去研究相比，Ryff (1989b) 各分量表間的相關係數在 .32 ~ .76、Ryff 與 Keyes (1995) 各分量表間相關係數在 .13 ~ .46，潛在變項間相關則在 .24 ~ .85；而 Springer 與 Hauser (2006) 在國家級資料庫 (如 WLS、NSFH II、MIDUS) 中，發現該量表潛在變項間的相關在 .74 ~ .98、.65 ~ .98 及 .48 ~ .87。雖然使用的題數及版本不同，但相較下，本研究兩樣本都顯示出本量表各向度間的區辨性更大，這可能是該版本確實較佳，或者因為本研究使用較為同質的大學生樣本，因此相關會降低，區辨性變大。

在與其它常用的主觀幸福感量表之效標關聯效度分析上，如表五所示，心理幸福感各分量表與生活滿意度、正負向情感及自尊的相關係數之正負情形如預期，可說是具有法則效度 (nomological validity)，且各有關變項與六向度間的相關組型也各有差異，可顯示出心理幸福感六個向度間的區辨性。整體而言，本研究結果與過去研究一致，自我接納分量表與生活滿意度、正負向情感、自尊的相關最高，其它分量表的結果則較為混雜或有弱至中等的相關 (Ryff, 1989b; Ryff & Keyes, 1995; Ryff et al., 1994)。

表五 分量表及總量表與其他效標變項間的相關係數

效標	正向關係 (PR)	自主 (AU)	環境精熟 (EM)	個人成長 (PG)	生活目的 (PL)	自我接納 (SA)	總量表
生活滿意	.28	.10	.30	.19	.21	.52	.36
正向情感	.39	.08 ^{ns}	.30	.34	.35	.44	.43
負向情感	-.21	-.04 ^{ns}	-.23	-.22	-.21	-.28	-.27
自尊	.47	.48	.43	.47	.45	.69	.66

註：^{ns}表示未達 .05顯著水準，其餘未標示則至少達 .05顯著水準。自尊與心理幸福感的相關使用樣本二資料，其餘使用樣本一資料。

表六 心理幸福感各向度與主觀幸福感變項之因素分析 (N = 394)

量表或變項	因素負荷量	
	心理幸福感	主觀幸福感
生活目的PL	.74	-.38
環境精熟EM	.74	-.44
個人成長PG	.73	-.35
正向關係PR	.60	-.44
自主AU	.52	-.16
生活滿意SWLS	.24	-.72
自我接納SA	.67	-.70
自尊SES	.42	-.66
正向情感	.37	-.64
負向情感	-.25	.51
特徵值	4.2 (3.1)	1.5 (2.8)
累積解釋變異比例%	42.1	56.6

註：心理幸福感各分量表分數是使用共同樣本第一次施測的資料。探索性因素分析是使用主軸因子萃取因素並以Oblimin法進行斜交轉軸。在特徵值中，括弧外代表未轉軸前的結果，括弧內代表轉軸後的結果。

進一步利用探索性因素分析檢查心理幸福感與主觀幸福感之間的區辨性，使用共同樣本獲得如表六顯示的 2 個因素。心理幸福感分量表中的自我接納及正向關係有跨因素的傾向，且自我接納落入主觀幸福感因素下，這與 Keyes 等人（2002）發現的結果一致，即心理幸福感與主觀幸福感雖在概念上相關但仍有所區別，同時也顯示出心理幸福感各向度仍有些許的不同，具有一定的區辨性。

二、測量不變性與交叉效度

（一）共同樣本跨時間的測量不變性

由表七檢定結果可知，兩個時間點下的一階六因素斜交模型都有合理的適配度。接著，將兩個時間點的樣本放在一起，利用多群組 SEM 進行構型不變性的檢定，此時的適配結果亦可接受，即因素結構不變性成立，表示兩組因素負荷量組型是一致的（Horn & McArdle, 1992; Vandenberg & Lance, 2000）。然後以此作為因素負荷量不變性檢定的比較模型，潛在量尺設定是採用效果編碼法，讓所有試題的因素負荷量可在兩組間設定相等，並利用巢套模型（nested model），將因素負荷量不變性與構型不變性兩個模型進行比較，結果卡方值差異量 12.22 在自由度差異為 12 時，並未達到 .05 的顯著水準，表示兩個時間點下的模型因素負荷量具有不變性。因此，本量表在兩個時間點上符合量尺不變性的條件，有利於因素間關係在兩個時間點之間的比較。接著，以量尺不變性模型為比較標準，設定各試題截距在兩時間點上相等，以進行純量不變性檢定，結果發現卡方差異量亦未達到 .05 顯著水準，表示模型在兩個時間點上符合純量不變性要求，有利於潛在平均數在兩個時間點上的比較。由於量尺不變性及純量不變性獲得支持，因此，這 18 題心理幸福度量表的六因素斜交模型在兩個時間點上符合強的測量不變性。

在測量不變性的檢定中，最嚴格的就是測量殘差變異不變性的成立，本研究檢定結果發現，在兩個時間點的測量殘差是不相等的，但這結果並無損於測量不變性的成立。事實上，多數學者認為測量殘差變異相當的假設太過嚴格，很少可以在實務中獲得驗證，且其對測量不變性的評估不是那麼重要（Brown, 2006; Byrne & Stewart, 2006; Chan, 1998）；測量殘差可能包含兩個部分，一是試題特定變異（item-specific factor），一是隨機誤差（random error），前者會因抽樣變異及測量條件變化而有差異，後者的隨機性使得相等性的期待變得沒有理論基礎；因此，強迫測量殘差相等並非可行的命題，且當試題特定變異僅是近似但卻強迫設定測量殘差相等，可能會將殘差散佈到其它模型參數上，導致參數估計的偏誤（Little et al., 2007）。故而在接下來的結構不變性檢定，不以測量殘差不變性模型為比較基礎，而是將截距不變性模型視為比較對象，以利變異數、共變數及潛在平均數依序

在組間進行比較。

在強的測量不變性成立後，所進行的結構不變性之檢定結果發現，卡方差異量 5.09 及 19.03 都未達到 .05 顯著水準。因此，跨時間的六因素變異及共變數可視為相等，不受時間改變的影響。接著，同時比較兩個時間點對應構念的潛在平均數差異，卡方差異量 14.36 達到 .05 顯著水準，表示至少有一個因素的平均數隨時間發生改變，由於本研究是檢定不同時間點整體不變性，因此，不列出個別潛在平均數在兩個時間點的比較結果，除非是想要了解潛在平均數變化的軌跡 (Brown, 2006)。

表七 共同樣本在兩個時間點的測量不變性檢定

模型	χ^2	df	p	RMSEA	CFI	Δ CFI	$\Delta\chi^2$	Δ df	p
單組解									
時間點1	325.56	120	< .001	.066	.9738	---	---	---	---
時間點2	349.86	120	< .001	.070	.9759	---	---	---	---
測量不變性									
構型不變性	675.42	240	< .001	.068	.9749	---	---	---	---
因素負荷量不變性	687.64	252	< .001	.066	.9749	.0000	12.22	12	> .05
截距不變性	698.19	264	< .001	.065	.9749	.0000	10.55	12	> .05
殘差變異不變性	788.11	282	< .001	.068	.9712	-.0037	89.92	18	< .001
結構不變性									
變異數不變性 ^a	703.28	270	< .001	.064	.9748	-.0001	5.09	6	> .05
共變數不變性	722.31	285	< .001	.062	.9745	-.0003	19.03	15	> .05
潛在平均數不變性	736.67	291	< .001	.062	.9742	-.0003	14.36	6	< .05

註：使用效果編碼法。在 .05 顯著水準下的臨界值， $\chi^2(6) = 12.59$ 、 $\chi^2(12) = 21.03$ 、 $\chi^2(15) = 25.00$ 。在 .001 顯著水準下的臨界值， $\chi^2(18) = 42.31$ 。^a變異數不變性模型是以截距不變性模型為比較對象。

(二) 樣本一及樣本二間的交叉效度

由表八可知，樣本一及樣本二間符合因素負荷量相當及截距相當，即量尺及純量不變性成立，但殘差變異不相等。在此條件下，繼續進行結構不變性的考驗，即以截距不變性模型為基礎，分別就因素變異數、共變數及潛在平均數不變性進行檢定。結果發現變異數不變性成立，但共變數在兩組間的差異達到 .05 顯著水準。由於本研究的兩個樣本並非如性別或種族間的比較來得有實質上意義，僅作為量表交叉效度之證據，因此，不擬列出每一個因素共變數在兩組間的詳細差異情形。最後，繼續進行潛在平均數不變性檢定，結果

兩組間並無顯著差異存在。

表八 樣本一及樣本二的測量不變性檢定（交叉效度）

模型	χ^2	df	p	RMSEA	CFI	ΔCFI	$\Delta \chi^2$	Δdf	p
單組解									
樣本一	374.71	120	< .001	.069	.9729	---	---	---	---
樣本二	288.93	120	< .001	.064	.9805	---	---	---	---
測量不變性									
構型不變性	663.64	240	< .001	.067	.9764	---	---	---	---
因素負荷量不變性	677.20	252	< .001	.065	.9763	-.0001	13.56	12	> .05
截踞不變性	689.36	264	< .001	.064	.9762	-.0001	12.16	12	> .05
殘差變異不變性	729.28	282	< .001	.063	.9748	-.0014	39.92	18	< .01
結構不變性									
變異數不變性 ^a	695.68	270	< .001	.063	.9762	.0000	6.32	6	> .05
共變數不變性	722.32	285	< .001	.062	.9757	-.0005	26.64	15	< .05
潛在平均數不變性	729.17	291	< .001	.062	.9756	-.0001	6.85	6	> .05

註：在 .05 顯著水準下的臨界值， $\chi^2(6) = 12.59$ 、 $\chi^2(12) = 21.03$ 、 $\chi^2(15) = 25.00$ 。在 .01 顯著水準下的臨界值， $\chi^2(18) = 34.81$ 。^a變異數不變性模型是以截距不變性模型為比較對象。

肆、討論與建議

一、討論

本研究針對 Li (2014) 的 18 題心理幸福感版本進行的一連串信效度考驗，並與過去研究結果進行比較後發現，該版本不僅保有過去版本的優點，同時也提供過去版本較少進行的驗證性因素分析及再測信度訊息。此外，本研究使用的樣本是大學生，而 Li 使用的樣本是中老年人，由於不論是在模型適配度檢驗或一般信效度上都有不錯的結果。因此，呼應 Ryff (1989b) 所宣稱的，該量表可以適用在所有成人身上的想法。此外，在獲得該量表信效度正面的回饋後，測量不變性執行的目的，主要是希望讓量表具有跨樣本交叉效度，並能夠進行跨時間的測量比較，讓量尺具有相同的測量單位及原點，基於此點，量尺不變性及純量不變性，亦即強的測量不變性成立是本研究的主要考量。

就本研究目的而言，可從兩種不同的觀點來看測量不變性的內容。若研究目的是為了

針對量表的交叉效度，也就是從母群中抽取二套有代表性的樣本來驗證量表的心理計量屬性，那麼廣義測量不變性的各項檢定在這兩個樣本間應該都要能通過，不宜有部分測量不變性，即使有放鬆限制的地方也須說明特別的理論考量；此時的測量不變性並不需要針對特定的性別或其他背景變項來處理，純粹就是交叉效度的檢定。然而，若研究目的是在信任原量表信效度下，進行不同背景或時間下的潛在變項比較研究，那麼狹義的測量不變性應該要能滿足，部分測量不變性可以被允許，亦即至少要達到分量尺及純量不變性，然後才能針對結構不變性進行評估，這時候研究者是希望潛在變項間的關係或潛在平均數的差異是顯著的，也就是共變數不變性及潛在平均數不變性不要成立，然後在適當的學說理論指引下，探討這些差異的意義，若有因果關係的可能性，甚至可進一步設定適當的因素間之路徑關係，來探討組間差異所蘊含的意義。事實上，前一種目的下的交叉效度要求在實務上很少能達成，主要是具代表性樣本的收集不太可行，以及廣義測量不變性的要求不易達成，即使可獲得具有母群代表性的樣本，測量殘差也很難在每一個觀察指標上都獲得組間不變的結果，而 Mulaik (2010) 也認為要從母群中抽到具因素不變性的樣本以及在比較性的因素分析中符合殘差變異不變性的假設，在實務上都有困難。

後一種研究目的，僅要求達到量尺及純量不變性，則較有可能在實務中驗證，例如比較三種學歷程度者在基本人類價值量表上的測量不變性 (Steinmetz, Schmidt, Tina-Booh, Wiczorek, & Schwartz, 2009)、性別在大學生生活滿意度量表上的因素不變性 (Wu & Yao, 2006)、中年與老年人在強的測量不變性建立後所進行的潛在變項路徑結構係數比較 (李仁豪、余民寧, 2014)、或是 Little (2013) 書中介紹的許多跨時間的潛在變項路徑關係的實際資料分析。值得一提的是，Little 對於測量不變性的看法即屬於狹義測量不變性，也就是確認達到強的測量不變性後，即可去探討潛在變項在跨時間上的變化情形。就本研究目的而言，兩個方便取樣的大學生樣本設計，試圖去進行心理幸福感量表交叉效度的確認，但仍可發現測量殘差變異不變的假設確實在實務上很難達到；而共同樣本跨時間的設計可歸屬於後者的應用，亦即在狹義的測量不變性中獲得量尺及純量不變性的結果後，發現大學生心理幸福感的潛在變項平均數會隨時間而改變，這符合心理幸福感理論的期望。

截至目前為止，本研究遵循傳統作法，對於模型差異的考驗方式仍然以卡方差異量的 .05 顯著水準為主要考量，但若改以 $\Delta CFI = .002$ 來重新檢視不變性的判斷，可發現本研究跨樣本的所有不變性檢定都成立，沒有任何特定矩陣或向量在兩樣本間出現顯著不同，達到全然的交叉效度；而在跨時間的共同樣本上，僅測量殘差變異無法獲得全然的不變性，這也符合過去學者認為的實務困難性。此外，值得注意的是，跨時間共同樣本的潛在平均數在 ΔCFI 標準下有達到不變性，但在 $\Delta\chi^2$ 的標準下卻顯示有差異，這意謂著間隔 2 ~ 3

週的大學生心理幸福感雖有變化，但並沒有重大的波動，這符合心理幸福感的理論，因為心理幸福感是較為功能取向的，不同於 Diener (1984) 所提出具有較多情緒成份的主觀幸福感，因此心理幸福感雖會有波動但不太可能在短期間內有重大的改變。

二、結論與建議

(一) 大學生樣本的信效度驗證增加量表的可用性

在量表信效度驗證上，利用臺灣中老年人為對象，選自 Ryff 等人 (1994) 84 題心理幸福感量表而形成的 18 題版本 (Li, 2014)，在本研究大學生樣本上的各項檢驗結果提供了更多信效度的訊息，不僅在樣本一及樣本二上獲得六因素結構的適配結果，且因素負荷量、內部一致性信度、因素間的區辨效度，都比 Li 使用中老年人為樣本時更高。此外，與其它研究常使用的不同主觀幸福感量表進行相關分析後，也提供了更多區辨效度與效標關聯效度的證據。另外，利用共同樣本獲得間隔 2 ~ 3 週的分量表相關介於 .59 ~ .65，總量表相關則達到 .70，提供了過去研究未曾提供的心理幸福感量表之再測信度證據。最後，將心理幸福感六個分量表與其他主觀幸福感變項一起進行探索性因素分析後，結果出現與過去研究相同的兩因素解及相似的變項歸屬情形，說明了心理幸福感與傳統主觀幸福感不僅在理論上有相同及相異之處，在實徵資料上確實也反映出這種異同的支持性證據。

(二) 測量不變性檢定結果提供交叉效度並讓量表的跨時間比較有意義

針對量表測量不變性檢定而言，樣本一及樣本二間的跨組比較以及兩個時間點間共同樣本的跨組比較，都具有強的測量不變性，亦即因素負荷量及截距都具有不變性，因此具有組別間相同的量尺單位及原點，達到量尺不變性及純量不變性的要求，提供了跨樣本的交叉效度證據，並有利於心理幸福感在長期追蹤研究中的因素間關係或因素平均數的跨時間比較。至於其它背景變項的測量不變性亦值得深入檢定，例如該量表在跨性別的測量不變性已獲得支持 (Li, Kao, & Wu, 2015)。但要注意的是，若不針對潛在變項，而是使用觀察變項分數進行比較的話，由於測量殘差不變性並未獲得十分肯定的結果，因此跨組間的比較更適用於群體平均數比較，不宜用在跨組間的個人分數比較。事實上，使用潛在變項來進行跨組比較會比使用觀察變項來得更接近母群真值的比較結果，特別是使用效果編碼法時，量尺的決定是由所有觀察指標的加權平均而產生，這樣產生的量尺與原來觀察指標的量尺更一致，也更接近母群真值，當觀察指標本身量尺是有實質意義時，潛在平均數差異效果量的計算就好像是在觀察指標原始分數上執行一樣地有意義 (Little et al., 2007)。

(三) 未來研究及使用建議

本研究樣本僅取自特定學校的方便樣本，可能無法全然地推論到臺灣所有大學生，未來研究或許可以繼續進行更具代表性樣本的交叉效度檢定，或針對不同年齡層及社經背景樣本的信效度及測量不變性進行檢定，以獲得更多有利的使用訊息，特別是不同年齡層間的測量不變性檢定可以確認本量表適用於所有成人的宣稱。此外，可搭配其它量表進行完整結構方程式模型的驗證，並在模型的測量不變性成立後，依特定理論的指引，進行潛在變項間路徑關係的設定，以驗證該理論。從實用性的角度而言，對於需要長期追蹤的調查研究，較短的量表具有減少受試者調查疲勞以及避免樣本流失等優點 (Stanton et al., 2002)，已逐漸成為一種趨勢 (Rammstedt & John, 2007)，尤其是以年紀較大的中老年人、低教育程度者，甚至是時間匆忙的上班族為研究對象時，簡式量表更能增加其配合的意願，減少遺漏值，使原始資料所需進行的人工干預 (例如平均數取代、多元插補等等) 降至最低，確保研究經費及心力獲得一定的回饋。因此，推薦使用本量表進行心理幸福感的相關研究，以提升研究結果的可信度。

參考文獻

- 余民寧、謝進昌、林士郁、陳柏霖、曾筱婕 (2011)：教師主觀幸福感模式建構與驗證之研究。測驗學刊，58 (1)，55-85。[Yu, M. N., Hsieh, J. C., Lin, S. Y., Chen, P. L., & Tseng, H. C. (2011). Confirmatory study of model of teachers' subjective well being. *Psychological Testing*, 58(1), 55-85.]
- 李仁豪、余民寧 (2014)：臺灣中老年人外向性及神經質人格、社會支持與主觀幸福感關係之模型建構及其結構係數比較。教育心理學報，45 (4)，455-474。DOI: 10.6251/BEP.20130725 [Li, R.-H., & Yu, M.-N. (2014). Modeling extraversion and neurotic personality, social support, and subjective well-being of the middle-aged and older people in Taiwan: Comparison of structural models. *Bulletin of Educational Psychology*, 45(4), 455-474. doi: 10.6251/BEP.20130725]
- 董智慧 (1998)：單身類型、社會支持與心理健康之分析研究。國立彰化師範大學輔導學系碩士論文，未出版，彰化。[Tung, C. H. (1998). *The relations among single types, social support and psychological well-being of never-married*. Unpublished master's thesis. National Changhua University of Education, Changhua, Taiwan.]
- 楊中芳譯 (1997)：性格與社會心理測量總覽 (上)。臺北：遠流。[Robinson, J. P., Shaver, P. R.,

- & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitude*. San Diego, CA: Academic Press.]
- Anderson, J. C., & Gerbing, D. W. (1988). Structure equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411-423. doi: 10.1037/0033-2909.103.3.411
- Allport, G. W. (1961). *Pattern and growth in personality*. New York, NY: Holt, Rinehart & Winston.
- Baumgartner, H., & Steenkamp, J. B. E. M. (1998). Multi-group latent variable models for varying numbers of items and factors with cross-national and longitudinal applications. *Marketing Letters*, *9*(1), 21-35. doi: 10.1023/A:1007911903032
- Bradburn, N. (1969). *The structure of psychological well-being*. Chicago, IN: Aldine.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466. doi: 10.1037/0033-2909.105.3.456
- Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order factor structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287-321. doi: 10.1207/s15328007sem1302_7
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.
- Bühler, C. (1935). The curve of life as studied in biographies. *Journal of Applied Psychology*, *19*(4), 405-409. doi: 10.1037/h0054778
- Bühler, C., & Massarik, F. (1968). *The course of human life*. New York, NY: Springer.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*(4), 421-483. doi: 10.1177/109442819814004
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504. doi: 10.1080/10705510701301834
- Cheng, S. T., & Chan, A. C. M. (2005). Measuring psychological well-being in the Chinese. *Personality and Individual Differences*, *38*(6), 1307-1316. doi: 10.1016/j.paid.2004.08.013
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5

- Clarke, P. J., Marshall, V. W., Ryff, C. D., & Wheaton, B. (2001). Measuring psychological well-being in the Canadian study of health and aging. *International Psychogeriatrics, 13*(S1), 79-90. doi: 10.1017/S1041610202008013
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542-575. doi: 10.1037/0033-2909.95.3.542
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75. doi: 10.1207/s15327752jpa4901_13
- Erikson, E. H. (1959). *Identity and the life-cycle*. New York, NY: Norton.
- Fernandes, H. M., Vasconcelos-Raposo, J., & Teixeira, C. M. (2010). Preliminary analysis of the psychometric properties of Ryff's scales of psychological well-being in Portuguese adolescents. *The Spanish Journal of Psychology, 13*(2), 1032-1043. doi: 10.1017/S1138741600002675
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. doi: 10.2307/3151312
- Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 6*(3), 258-269. doi: 10.1037/1082-989X.6.3.258
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-144. doi: 10.1080/03610739208253916
- Jahoda, M. (1958). *Current concepts of positive mental health*. New York, NY: Basic Books.
- Joshanloo, M. (2014). Eastern conceptualizations of happiness: Fundamental differences with Western views. *Journal of Happiness Studies, 15*(2), 475-493. doi: 10.1007/s10902-013-9431-1
- Jung, C. G. (1933). *Modern man in search of a soul*. New York, NY: Harcourt.
- Kafka, G. J., & Kozma, A. (2002). The construct validity of Ryff's scales of psychological well-being (SPWB) and their relationship to measures of subjective well-being. *Social Indicators Research, 57*(2), 171-190. doi: 10.1023/A:1014451725204
- Keyes, C. L. M., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology, 82*(6), 1007-1022. doi: http://dx.doi.org/10.1037/0022-3514.82.6.1007
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

- Li, R. H. (2014). Reliability and validity of a shorter Chinese version for Ryff's psychological well-being scale. *Health Education Journal*, 73(4), 446-452. doi: 10.1177/0017896913485743
- Li, R. H., Kao, C. M., & Wu, Y. Y. (2015). Gender differences in psychological well-being: Tests of factorial invariance. *Quality of Life Research*, 24(11), 2577-2581. doi: 10.1007/s11136-015-0999-2
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.
- Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variable in SEM and MACS models. *Structural Equation Modeling*, 13(1), 59-72. doi: 10.1207/s15328007sem1301_3
- Maslow, A. H. (1968). *Toward a psychology of being* (2nd ed.). New York, NY: Von Nostrand.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568-592. doi: 10.1037/0021-9010.93.3.568
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. doi: 10.1007/BF02294825
- Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). New York, NY: Taylor & Francis Group.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.
- Neugarten, B. L. (1973). Personality change in late life: A developmental perspective. In C. Eisendorfer & M. P. Lawton (Eds.), *The psychology of adult development and aging* (pp. 311-335). Washington, DC: American Psychological Association.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203-212. doi: 10.1016/j.jrp.2006.02.001
- Richardson, C. G., Ratner, P. A., & Zumbo, B. D. (2009). Further support for multidimensionality within the Rosenberg self-esteem scale. *Current Psychology*, 28(2), 98-114. doi: 10.1007/s12144-009-9052-3
- Roger, C. R. (1961). *On becoming a person*. Boston, MA: Houghton Mifflin.
- Rosenberg, M. (1965). *Society and adolescent self-image*. Princeton, NJ: Princeton University Press.

- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52(1), 141-166. doi: 10.1146/annurev.psych.52.1.141
- Ryff, C. D. (1989a). Beyond Ponce de Leon and life satisfaction: New directions in quest of successful aging. *International Journal of Behavioral Development*, 12(1), 35-55. doi: 10.1177/016502548901200102
- Ryff, C. D. (1989b). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069-1081. doi: 10.1037/0022-3514.57.6.1069
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69(4), 719-727. doi: 10.1037/0022-3514.69.4.719
- Ryff, C. D., Lee, Y. H., Essex, M. J., & Schmutte, P. S. (1994). My children and me: Mid-life evaluations of grown children and of self. *Psychology and Aging*, 9(2), 195-205. doi: 10.1037/0882-7974.9.2.195
- Salmela-Aro, K., & Schoon, I. (2005). Introduction to the special section: Human development and well-being. *European Psychologist*, 10(4), 259-263. doi: 10.1027/1016-9040.10.4.259
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, 17(3), 343-358. doi: 10.1207/s15327906mbr1703_3
- Shmotkin, D. (2011). The pursuit of happiness: Alternative conceptions of subjective well-being. In L. W. Poon & J. Cohen-Mansfield (Eds.), *Understanding well-being in the oldest old* (pp. 27-45). New York, NY: Cambridge University Press.
- Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Ryff's scales of psychological well-being: Method, mode, and measurement effects. *Social Science Research*, 35(4), 1080-1102. doi: 10.1016/j.ssresearch.2005.07.004
- Stanton, J. M., Sinar, E., Balzer, W., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55(1), 167-194. doi: 10.1111/j.1744-6570.2002.tb00108.x
- Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1), 138-161. doi: 10.1037/0033-2909.134.1.138
- Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90.

- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, 43(4), 599-616. doi: 10.1007/s11135-007-9143-x
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70. doi: 10.1177/109442810031002
- Van Dierendonck, D. (2004). The construct validity of Ryff's scales of psychological well-being and its extension with spiritual well-being. *Personality and Individual Differences*, 36(3), 629-643. doi: 10.1016/S0191-8869(03)00122-3
- Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, 64(4), 678-691. doi: 10.1037/0022-3514.64.4.678
- Wu, C. H., & Yao, G. (2006). Analysis of factorial invariance across gender in the Taiwan version of the satisfaction with life scale. *Personality and Individual Differences*, 40(6), 1259-1268. doi: 10.1016/j.paid.2005.11.012

收件日期：104年06月29日

複審一日期：104年08月18日

複審二日期：104年09月22日

通過日期：104年10月19日



附錄一 18題心理幸福感量表及描述統計 (N = 1185)

	平均	標準差	偏態	峰度
正向關係 (PR)	13.41	2.23	-.10	.58
1.我喜愛與家人或朋友聊天和分享個人話題	4.55	0.98	-.42	.35
2.人們形容我是個肯付出的人，願意花時間在他人身上	4.28	0.90	-.05	.44
3.我知道我可以信任我的朋友，而他們也知道可以信任我	4.58	0.85	-.18	.16
自主 (AU)	12.20	2.40	.17	.19
4.我對自己的主張很有信心，即使與多數人的共識不同	3.96	0.94	.14	.11
5.我不是那種會屈服於社會壓力而表現出某些行為或思考方式的人	4.03	1.00	.07	-.04
6.我以自己認為重要的價值來評論自己，而非依照別人的標準	4.21	0.95	.01	.06
環境精熟 (EM)	12.93	2.24	.08	.64
7.整體而言，我覺得自己能掌管生活情境	4.28	0.85	-.11	.66
8.我能管理好日常生活中該做的事情	4.34	0.85	-.08	.66
9.我通常能管理好個人財務與事務	4.31	0.90	-.17	.35
個人成長 (PG)	14.62	2.33	-.29	.01
10.我認為有新的體驗來激發自己如何看待自我及世界很重要	4.78	0.87	-.18	-.34
11.對我而言，人生是持續學習、改變和成長的過程	4.96	0.86	-.33	-.24
12.我欣見自己看事情的觀點逐年改變且成熟	4.88	0.89	-.47	.36
生活目的 (PL)	13.10	2.57	-.09	-.13
13.我喜歡為未來訂定計畫，並努力實踐它	4.33	1.00	-.19	.02
14.我會積極完成已擬訂的計畫	4.19	0.97	-.08	-.00
15.我不是那一種對人生毫無目標的人	4.58	1.02	-.44	.35
自我接納 (SA)	12.27	2.42	-.07	.38
16.當我回顧過往，對於大多數事情的結果我感到滿意	3.93	0.95	-.19	.50
17.整體來說，我認為自己有自信且積極	4.05	0.95	-.09	.24
18.當我和身邊朋友相比時，我覺得自己還不錯	4.29	0.92	-.14	.37
總量表	78.52	10.64	.04	.54

註：所有偏態係數標準誤皆為 .07，所有峰度係數標準誤皆為 .14

Reliability, Validity, and Measurement Invariance of the Brief Chinese Version of Psychological Well-Being Scale among College Students

Ren-Hau Li

Chung-Shan Medical University

Min-Ning Yu

National Chengchi University

The two aims of this research were to obtain reliability and validity information of Li's (2014) 18-item brief Chinese version of the Psychological Well-Being Scale across samples, and to test its measurement invariance across time. Three samples of college students were recruited in the study. They were named as sample I, sample II, and common sample, in which 442, 349, and 394 students, respectively. Data were analyzed using structural equation modeling techniques. Results showed that theoretical framework of six-factor oblique model had good model-fits. Cronbach's α coefficients ranged from .73 to .87 for subscale scores, and was .91 for the total score; 2 ~ 3 weeks test-retest reliability coefficients of subscale scores ranged .59 ~ .65, and was .70 for the total score. Based on the 6-factor model, factor loadings were among .65 ~ .92 providing convergent validity, and interfactor correlations were ranging .30 ~ .72 offering discriminant validity. Besides, cross validation was supported. Furthermore, metric invariance and scalar invariance existed across groups in the measurement equivalence tests; therefore it ensured that the scale can be compared on its constructs across multiple measurement times, including comparisons in interfactor correlations and latent means.

Keywords: cross validation, measurement invariance, metric invariance, psychological well-being, scalar

invariance

