

研究論文

Mokken Scale 量表之應用： 以長期變遷的語言學習 和使用態度題組為例*

廖培珊** 江振東*** 林定香****

摘要

調查問卷常使用計分量表來測度受訪者的潛藏特質，然而以常見的葛特曼量表為例，題目間難易程度之嚴格假定，缺乏隨機性的概念與測量誤差之考量等，使其應用受到限制。被視為項目反應理論之無母數模型的 Mokken 量表，具單一向度、局部獨立性、單調性等的特質，並考量回答類別的反應機率，因此可用以檢測含數個題目的題組是否可組成測量同一潛藏概念之量表。

本研究介紹 Mokken 量表模型之基本假設與特性，並以「台灣社

-
- * 本研究為科技部（前國家科學委員會）補助整合型前瞻計畫「調查計畫串連
加值方法與應用」子計畫三（NSC100-2420-H-001-018-MY2）之部分研究
成果，所使用資料係採自科技部資助之「台灣地區社會變遷基本調查」計畫
歷年成果（DOI: 10.6141/TW-SRDA-tscs），資料由中央研究院學術調查研究
資料庫釋出。
 - ** 通訊作者，中央研究院人文社會科學研究中心調查研究專題中心副研究員。
台北市南港區 11529 研究院路二段 128 號，電話：(02)27898183，傳真：
(02)27854160，E-mail: psliao@gate.sinica.edu.tw。
 - *** 國立政治大學統計學系副教授。
 - **** 國立台北大學統計學系教授。

會變遷基本調查」2004 年至 2014 年共三期的資料檔中有關語言學習及使用態度之題組為例，利用 Mokken 量表的條件特性來檢視量表組成，並將估算出之量表分數與個人背景和語言使用之變項進行分析。研究結果指出，以逐項條件檢驗的方式來看，包含三個题目的語言學習及使用態度題組整體而言符合 Mokken 量表之特性；邏輯斯迴歸分析則可看出個人背景及家中使用語言與潛藏態度的關係，均會因調查年度的不同而有差異，本研究並就 Mokken 量表模型的特性與應用分析結果提出討論。

關鍵詞：Mokken 量表、葛特曼量表、項目反應理論、潛藏特質、語言學習

**Application of Mokken Scale:
An Example of Attitudes toward Language Learning and
Use with Longitudinal Data**

Pei-shan Liao* Jeng-Tung Chiang Ting-Hsiang Lin*****

ABSTRACT

Cumulative scales have been widely used in questionnaires to measure respondents' latent traits, such as personality or attitudes toward certain events. Different scaling models have been developed to reveal the disparity among respondents, in which Guttman-type scales are commonly used. However, their assumption of a unique combination of response patterns and lack of stochastic implications of the scalogram indices have limited their application to empirical studies. The Mokken scale, which can also be seen as a non-parametric item nonresponse model (NIRT), has been developed to relax the rigid assumptions of the Guttman scale while maintaining its requirements of

* Associate Research Fellow, Center for Survey Research, RCHSS, Academia Sinica

** Associate Professor, Department of Statistics, National Chengchi University

*** Professor, Department of Statistics, National Taipei University

uni-dimensionality, local independence, and monotonicity. In addition, the Mokken scale takes into account probabilities of responses to question items, which enable the examination of whether these items are measures of the same latent construct.

This study describes the basic assumptions and characteristics of the Mokken scale, and demonstrates its application using empirical data. Question items which measure the latent attitudes about language learning and use from 2004-2014 Taiwan Social Change Survey (TSCS) are used as an example to examine the scalability of the items and the scale that consists of those items. Furthermore, logistic regression of the latent trait using Mokken scale scores on demographic characteristics and a variable associated with language used at home are performed. Our results indicate that the association between the examined variables and the latent trait varies depending on the survey year. Discussion on the conditions of the Mokken scale model and findings based on empirical data are provided.

Keywords: Mokken scale, Guttman scale, Item Response Theory, latent trait, language learning

一、前言

量表為調查問卷中常見的形式，用以測度受訪者的潛藏特質 (latent trait)，例如人格特質或對特定問題之態度等。由於這些潛藏變項無法直接觀察，為瞭解受訪者在特質或態度上程度的差異，學者設計出不同形式的加總或累計的方式來予以測度給分；葛特曼量表 (Guttman scale) 即為其中一例。建構葛特曼量表時，其量表模型要求組成量表的題目彼此間需符合特定之難易程度排列，亦即題目之排序僅為單一

向度；回答組合的數量與內容亦需依據題目難易的排列而受到限制，因此對於非特定回答組合所允許之誤差值甚小，若超出範圍則不被認定為葛特曼量表。然而，前述條件並不容易達成，因此在以問卷調查所獲得之訪問資料中，常出現資料結構不符合原始量表設計的狀況，使得研究者無法以實證資料確認問卷題目彼此間之強弱差異。此外，葛特曼量表並未考量因測量誤差或抽樣誤差而出現的錯誤回答；用以檢測量表符合程度之評量標準也無法依據所涉及之題目或樣本數量來做統計上的調整。當量表被判定為不合適的葛特曼量表時，研究者通常也很難以替代性的量表模型提供一個明確的量化描述，因而為學者們所詬病（McCutcheon 1987; Clogg and Sawyer 1981）。

此外，調查所得之原始資料通常為多類別或順序尺度之測量，但在檢測題組之量表特性時，卻常被簡化為二元計分（Coromina and Saris 2012; Liao and Tu 2006）；或者並不考慮其所具備之量表特性，或是題目間可能的難易順序，而直接採用類似李克特氏量尺（Likert-type scale），加總各題的回答後做為量表分數，視為連續尺度來進行分析（王衛東 2010）。為了改善前述葛特曼量表模型的缺失，學者們發展出替代的統計模型來估計參數，例如潛藏結構分析（latent structure analysis）的不同潛藏類別模型或 Mokken 量表（Mokken 1971），可用來改善葛特曼量表對於單一向度（uni-dimensionality）的嚴格假定。這些量表模型皆著眼於機率的觀點，來對回答反應進行評估；因此，當受訪者的特性較強時，對於接受程度較高的題目，持同意看法的機率會高於不同意的看法。同理，當受訪者的特性較弱時，對於接受程度較低的題目，持不同意看法的機率則會高於同意的看法。

由於考量了回答機率，相對於葛特曼量表的二元計分方式，Mokken 量表模型在進行分析檢測時，較能夠保有題目在測量上的多

元計分特性；同時，Mokken 量表可考量題目間既定的難易程度，進而以機率的概念計算量表分數。舉例而言，當被認定的答案出現的機率值越大時，表示該題目相較於其他題目來得容易，因此在加值計分上其加權分數較小；反之，當出現機率之數值越小時，表示該題目相較其他題目來得困難，因此在加值計分上其加權分數較大。因此，我們可透過 Mokken 量表多元計分的模型配置，來尋求合理的加值計分標準與量表分數，無須受到葛特曼量表對特定回答組合僅可有特定量表分數的限制。

由於其較複雜的統計模型和運算軟體之需求，Mokken 量表在研究上的運用並未廣泛取代既有的葛特曼量表模型，多侷限在公共衛生領域來探討量表題組內容的適用性 (Maindal et al. 2009; Nitschke et al. 2009; Shenkin et al. 2014)，於近期才有一些應用於社會科學的研究 (Coromina and Saris 2012; Loner 2016; Tai and Treas 2013; van der Veer et al. 2013)；而國內的相關研究則是付之闕如，僅有部分探討量表特性的學位論文 (如江怡萱 2014；高千涵 2012；黃雅雯 2012；楊茵筑 2010)。同時，多數研究採用 Mokken 量表主要用來檢視量表題目之表現，以決定量表組成，但經常是以簡化後之二元計分處理回答類別 (Coromina and Saris 2012; Loner 2016; Maindal et al. 2009; Nitschke et al. 2009; Shenkin et al. 2014; Tai and Treas 2013; van der Veer et al. 2013)，無法真實反映出問卷量表原有的多元計分 (polytomous responses) 特性，殊為可惜。

以國內各項具代表性的大型學術調查而言，常會因議題的不同，而使得調查後所獲得之資料類型未必通過葛特曼量表模型的檢測，進而無法適切地估計各類回答類型之分數。即使調查資料符合量表條件，使用者亦多以二元計分方式來處理 (Liao and Tu 2006)，同時僅

限於回答類型之討論，而未必採用量表的計分特性。利用 Mokken 量表模型除可改善葛特曼量表模型對於單一向度的嚴格假定外，也可由機率觀點給予量表分數。本研究以歷年社會變遷調查中，測量同一潛藏概念，且題目間接受度具有高低差異的一組量表為例，利用 Mokken 量表模型來估算其潛藏結構之機率分數，檢視不同年度的資料是否均符合量表特性。同時，亦納入問卷中的相關變項進行分析，來展現 Mokken 量表分數的應用效果。

本研究所考慮之潛藏模型：Mokken 量表，可被視為項目反應理論（item response theory, IRT）的無母數模型（nonparametric model）（Molenaar 1997），同時亦屬於放寬葛特曼量表假定之量表模型。以下各節將簡要說明葛特曼量表模型，接續介紹測量該量表中之潛藏特質的 Mokken 量表模型，並說明其與項目反應理論之差異。最後利用實證資料進行研究與分析結果，藉此評估 Mokken 量表模型進行實證資料分析的效用。

二、量表模型

（一）葛特曼量表模型與項目反應理論

常應用於社會科學的葛特曼量表（Guttman 1944）屬於計分量表，以一組依據難易度或接受度來排序的題目，來定位受訪者的態度，並希望特定量表分數僅反映出單一類型的回答組合。假設一組題目共有 I 題，雖然題目之間的難易程度排列順序可能有 $I!$ 種，但在葛特曼量表模型的架構下僅允許一種排列方式，因此葛特曼量表分析係針對一組依循難易程度所排序的題目，根據受訪者之回答反應來量測其特質（或稱為潛藏能力 θ ）的方法。由於每個題目都是為了量測同一種潛藏

特質所設計，此種潛藏特質必須包含在題目的假設之內，以符合單一向度的假設（Guttman 1950）。在此前提之下，研究者便可根據每一位受訪者對所有題目的作答總得分來進行其潛藏能力估計，藉以推論所有受訪者的能力強弱。由於題目之間具有難易程度而所量測的潛藏能力或特質只有一個，因此可以在實數線上，將各個題目依照接受難易的程度與每一位受訪者能力的強弱，分別排序出來。由此可見葛特曼量表模型決定性（deterministic）的性質，此設計可以確保對一個量表總得分相同的受試者而言，僅有一種獨特的回答類型（Clogg and Sawyer 1981）。

假設量表各題目的接受程度由容易至困難依序排列，從葛特曼量表模型的架構可知，當受訪者同意某一個題目，但其後續題目不同意時，可視為轉折點。受訪者至轉折點回答同意的題數越多，就表示其潛藏特質越強，因此葛特曼量表亦是一個累加性的量表。通常研究者會以二元計分的方式，將題目作答同意的情況記為得分 1，而題目作答不同意的情况則記為得分 0。當受訪者的總得分相同時，就表示他們的作答反應類型是一致的。因此除了這些固定的回答型式之外，如果其他的回答類型也反映出相同的總得分，這些回答類型便被視為錯誤回答，稱為葛特曼量表的錯誤類型（Guttman errors）。

這些獨特類型的論點亦隱含了單調同質性模型（Monotone Homogeneity Model, MHM）的假定（Green 1956; Loevinger 1948）。所謂單調同質性模型即表示對於任何題目 i ，「同意」該題目的受訪者其總得分應該比「不同意」該題目的受試者之總得分還高，其中 $i=1, \dots, I$ 。當所有作答反應類型皆沒有出現任何葛特曼量表的錯誤類型時，可稱此量表為完美的葛特曼量表（perfect Guttman scale）。如果觀察到的作答類型有過多葛特曼量表的錯誤類型出現，則該量表會被視為不符合

葛特曼量表模型。

然而前述葛特曼量表的假定過於嚴苛，且未考量隨機及測量誤差，使得實務上並不容易滿足各項條件。因此，後續學者從古典機率論的觀點，提出不同的模型，來討論個人 (person) 與題目 (item) 參數的推估可如何改善，包含項目反應理論 (IRT) 的 Rasch 模型 (Andrich 1985) 及 Mokken 量表模型 (Molenaar 1997) 等。由於 Mokken 量表亦被視為 IRT 的無母數模型 (Molenaar 1997; Sijtsma and Molenaar 2002)，許多特性與 IRT 模型相同，因此在討論 Mokken 量表之前，先說明 IRT 模型的基本特性，以及常見的 Rasch 模型。

項目反應理論主要是用來描述題目特性與受測者的特質如何同時影響其回答反應的一種模式 (余民寧 2009)，其中敘述題目特性包括題目難度、題目鑑別度、與題目可猜測度，而受測者的潛藏特質通常指潛藏能力、潛在認知或潛在意向。其基本假設有二：(1) 潛藏變項具備單一向度之特性；(2) 局部獨立性 (local independence)；意即每位受試者對於某一題目的回答僅與其潛藏特質有所關聯，題目與題目間的答題反應彼此之間則沒有任何關聯。此外並假設所涉及之潛藏變項為連續尺度。根據其所包含的試題特性參數個數可以區分為單參數的 Rasch 模型、二參數模式及三參數模式 (Birnbaum 1968)，也可以依據計分形態來分，分成二元計分 (dichotomous) 與多元計分 (polytomous) 模型。

項目反應理論可以根據題目的特性，透過試題反應函數 (item response function, IRF) 來表現受試者在某一題目答對的機率與其潛藏特質之間的對應關係。若題目的數量共有 X_1, \dots, X_I 項，而受試者對於題目 i 的回答值為 $x_i \in \{0, 1, \dots, m\}$ ，則試題反應函數可以定義為一個機率函數 $P(X_i = x_i | \theta)$ 。對於多元計分的題目，若題目的回答共有 $m+1$

種類別，則每一項具有 $m+1$ 類排序回答的題目可拆解成 m 個試題階段 (item step)，可以 $X_i \geq x$ 表示，而 $P(X_i \geq x | \theta)$ 稱為試題階段反應函數 (item step response functions)。因為對所有的 θ 而言， $P(X_i \geq 0 | \theta) = 1$ ，故第 i 題與 θ 的特徵關係可由 m 個試題階段反應函數來呈現： $P(X_i \geq 1 | \theta)$, $P(X_i \geq 2 | \theta)$, ..., $P(X_i \geq m | \theta)$ 分別表示潛藏能力為 θ 的受試者，第 i 題回答反應大於等於 1, 2, ..., 或 m 的機率。IRT 可針對每一個題目產生試題特徵曲線 (item characteristic curve, ICC)，將 IRF 的函數關係表達在二維的平面圖形上，其中橫軸座標軸呈現的是受試者的潛藏能力，而縱軸座標軸則為受試者答對該題的機率，亦即 ICC 為 IRF 的函數圖形。在多數的情況下，試題特徵曲線呈現一個 S 型的邏輯斯函數 (logistic function)；依據模型的不同，同一個量表的不同題目可能呈現平行 (不相交) 或不平行 (相交) 的特徵曲線。

在常見的 IRT 模型中，Rasch 模型屬於單參數模型 (Rasch 1960)，題目答對的機率同時受到受試者能力特質與題目難易度所影響 (王文中 2004)；而受試者的整體表現就是各題目答對機率的聯合機率。藉由這個模型與聯合機率的觀念，研究者可根據受試者在各個題目的答題反應來估計受試者的潛藏特質以及題目特性，同時回答不同題目的受試者其潛藏能力也可以互相比較。換言之，每個受試者跟每個題目均被單一的難度參數所限定，此種不變性 (invariance) 是 Rasch 模型與葛特曼量表模型的共通點，前者也因此被用來做為後者的機率模型之探討 (Andrich 1985)。在 Rasch 模式中，鑑別度的高低是一個需要考慮的問題；但在葛特曼量表中，並沒有這個問題，因為葛特曼量表具有完全鑑別度。

雖然 Rasch 模型可分別估算題目難度與受試者能力，且原始分數是潛藏特質 θ 的充分統計量 (sufficient statistic) (王文中 2004: 651)，

類似於葛特曼量表模型以量表分數做為充分統計量，來估計受試者的定位；因此二者在原始分數與潛藏特質的一對一關係容易理解。但是 Rasch 模型的假設過於嚴苛，且適用於題目數量多的量表（van Schuur 2003），對於以數量少的題目來測量單一概念的問卷而言，模型配適度（model fit）並不佳，因此本研究考量其他學者同樣從機率觀點所發展出 Mokken 量表，來處理葛特曼量表類型之模型配適度與其他相關的問題。

（二）Mokken 量表

Mokken 量表首先由 Rob J. Mokken 在 1971 年所提出，屬於心理量化的模型（Sijtsma and Molenaar 2002），可依照題目（外顯變項）的難易程度來量測受試者不同的潛藏特質或能力。Mokken 量表對於題組架構的要求與葛特曼量表雷同，亦同樣可應用於二元計分與多元計分（Mokken and Lewis 1982; Sijtsma and Molenaar 2002）。葛特曼量表強調的是每一個題目中之影響變項相互間是有絕對的次序關係，也可以說相互間具有層級順序關係。所以當受試者的態度分數相同時，其行為表現必定相同，故所對應到受試者的潛藏特質應是一致的；若是在態度分數相同時還出現其他不同的態度組合時，就被視為前述的葛特曼錯誤。然而 Mokken 量表並不認為同意後面較難被接受的題目時，就代表前面較容易被接受的題目也必然會被同意；亦即題目彼此間的對應關係未必是絕對的層級順序關係，可以透過機率大小來衡量；換言之，當受試者同意後面較難的題目時，前面較容易的題目被同意的機會亦增大，但不見得完全都是同意，只是同意的機率較後者大。

影響機率的因素包含題目的難易程度以及受試者潛藏能力 θ ，也就是說題目越容易且受試者潛藏能力越強，則答對（同意）的機率就

會相對變大；反之若題目越困難且受試者潛藏能力越弱，則答對（同意）的機率相對地就變小了。至於題目的難易程度及受試者潛藏能力如何估計？一般來說，Mokken 量表是依據所有受試者對每一個題目答對（同意）的比例做為衡量題目難易程度的標準，機率值越大就代表該題目越容易，機率值越小就代表該題目越困難，研究者只要根據機率值大小排序，就可以瞭解每一個題目的難易程度；若重新依據難易程度來排序題目時，建構出來的量表就可以在特定條件下成為葛特曼量表。另外，依據每一個受試者對所有題目答對（同意）的總數做為衡量潛藏能力 θ 的標準，總數越大就代表該受試者潛藏能力 θ 越強。此外，為了方便繪圖及潛藏能力 θ 的比較，可以將每一個受試者的潛藏能力 θ 進行標準化或者是研究者針對受試者在不同總題數之下，賦予 θ 合理的定義。一般來說， θ 的值為任意實數，也可以經由轉換變為全部正值或負值。

由項目反應理論的觀點來看，Mokken 量表被視為無母數的 IRT 模型（NIRT）（Molenaar 1997），其基本假定除了 IRT 原有的單一向度（uni-dimensionality）與局部獨立性（local independence）外，還包含單調性（monotonicity）及非相交性（nonintersecting item response functions）等性質。在單一向度的條件下，量表中的所有題目全都聚焦於同一潛藏特質來做設計。局部獨立性指的是在給定 θ 的條件下，兩兩題目間的反應相互獨立，亦即 $P(X_i=x_i, X_j=x_j | \theta) = P(X_i=x_i | \theta)P(X_j=x_j | \theta)$, $i \neq j$ 。單調性指的是所有試題反應函數 $P_i(\theta)$ 是 θ 的非遞減函數；就任意 $\theta_a < \theta_b$ 而言， $P(X_i \geq x | \theta_a) \leq P(X_i \geq x | \theta_b)$, $i=1, \dots, I, x=0, \dots, m$ 恆成立。而非相交性指的是試題反應函數彼此間不相交，亦即無論受訪者的潛藏特質或能力為何，較難的題目的答對（同意）的機率一定不會高於較簡單的題目，因此就所有受訪者而言，所有題目在難易程度的認知

上完全相同 (Sijtsma and Molenaar 2002)。此外，Mokken 量表模型的試題特徵曲線 (ICC) 呈現為分段線性函數 (piecewise linear function)，並具備單調同質性的特質，可視為 IRT 的特例 (高野等 2014)。

在這些假設前提下，衍生出兩個無母數的項目反應理論模型：單調同質性模型 (MHM) 及雙重單調性模型 (Double Monotonicity Model, DMM)。單調同質性模型的基本假設符合 Mokken 量表的其中三項：單一向度、局部獨立性、單調性；MHM 用來描述一組同質性 (單一向度) 題目的回答反應，且其 IRFs 與潛藏特質的關係為單調性，意謂受試者在 θ 上的順序可以由答題總分 X_+ 來決定，因此 MHM 是在順序尺度 (ordinal scale) 上測量受試者的 IRT 模型 (Sijtsma and Junker 1996; Wind 2015)。一個配適度佳的 MHM 表示研究者雖然無法如有母數的 IRT 模型般可獲取對於 θ 的估計值，卻可以利用受試者總分的高低，將他們的潛藏特質 θ 排序。在實務上，MHM 可用以檢視並建構適當的試題或問卷題組。

雙重單調性模型則包含兩種單調的性質，亦即除了 MHM 的三項假設外，亦符合第四項的非相交性假設，其對各題目所描繪出之 ICC 呈平行；因此被視為 MHM 的特例，具有題目順序不變 (Invariant Item Ordering, IIO) 的特性。如果包含了 k 個二元計分的題目可以被編號排序，且對於每個 θ 而言，試題反應函數符合 $P_1(\theta) \leq P_2(\theta) \dots \leq P_k(\theta)$ 的條件，那麼這些題目就具備了題目順序不變性。在應用上，研究者可以依據受試者對於個別試題表現來比較表現的好壞；而這些題目的順序在同一母體中的次群體之間同樣具有不變性。若資料性質符合 DMM，不僅可將受試者的潛藏特質排序，還可以利用題目的正向答案比例之高低，對於題目難易度進行排序 (Sijtsma and Junker 1996; Wind 2015)。

如前所述，一般的 MHM 可用以檢測題組內容的適當性，這些用來檢測題組是否滿足 Mokken 量表條件的指標，包含三個量表係數 (scalability coefficients)，或稱為同質性係數 (coefficient of homogeneity) (Loevinger 1948)： H_{ij} 、 H_i 、 H ，其中 H_{ij} 代表測量兩兩配對題目 i 、 j 之間的同質性 (homogeneity) 程度或結合性 (association)，亦即題目間的正向關係； H_i 代表測量特定的第 i 題目可能與其他題目配置在一起的同質性整合效果，因此它會聚集所有與 i 題目有關的配對 H_{ij} 之訊息，其中 $j=1, 2, \dots, I$ ，但是 $j \neq i$ ，其數值越大表示 i 題目與其餘題目的正向關係越強。 H 代表測量整體量表題目配置在一起的同質性效果，因此其會聚集所有個別題目 H_i 的訊息，其中 $i=1, 2, \dots, I$ ，數值越大表示所有題目彼此的同質性越高，也就是正向關係越強。關於三個係數的詳細公式可參見相關文獻 (Loevinger 1948; Sijtsma and Molenaar 2002)。

量表係數蘊含著兩個重要的特性；首先，對於所有題目 i 、 j 而言，量表係數 H_{ij} 、 H_i 、 H 滿足以下的關係式：

$$\begin{aligned} \min_j H_{ij} &\leq H_i \leq \max_j H_{ij} \\ \min_i H_i &\leq H \leq \max_i H_i \\ \min_{i,j} H_{ij} &\leq H \leq \max_{i,j} H_{ij} \end{aligned}$$

其次，在 MHM 模型中，量表係數的範圍跟機率的定義一樣，均在 0 與 1 之間。因此，當 $H=1$ ，表示沒有任何的葛特曼錯誤，此時題組及唯一個完美的葛特曼量表；當 $H=0$ ，則表示題目間沒有任何關聯性。Mokken (1971) 根據經驗法則，建議對於所有 $i=1, 2, \dots, I$ ， H_i 至少必須 ≥ 0.3 ；當 $0.3 \leq H < 0.4$ ，此時的整體試題量表是一個弱的量

表（鑑別力較差）；當 $0.4 \leq H < 0.5$ ，此時的整體量表是一個居中的量表（鑑別力普通）；當 $H \geq 0.5$ ，此時的整體量表是一個強的量表（鑑別力較好）。

在進行 Mokken 量表計分之前，必須先確認組成量表的這些題目至少符合 MHM 的條件。研究者可利用幾個步驟來檢測組成量表的題目是否合乎單調同質性的條件（Sijtsma and Molenaar 2002），以決定其是否可被確認為 Mokken 量表。首先要檢視各個題目的平均分數來確認題目間的難易順序；亦即被接受或同意的程度（item popularities）；接著檢測題目彼此間 H_{ij} 值是否均大於零，並檢視個別題目的量表係數 H_i ，以確認個別題目至少需具備正向的鑑別力（ $H_i > 0$ ），同時也計算出整體量表的 H 值，來決定這些題目所組成的量表是否可被視為一個可接受量表（ $H \geq 0.3$ ）。最後則須檢視試題反應函數（IRFs）是否具備非遞減之單調性特質；當題目的測量屬於多元計分（polytomous）的情況時，則需於前述步驟中考慮對於個別題目在不同回答類別的機率，亦即由某一回答類別轉換到其他回答類別的機率，稱為試題階段反應函數（Item step response functions, ISRFs）。

三、研究方法

本研究將以歷年社會變遷調查中對於語言學習和使用態度的一組題目為例，利用三個年度的調查資料先檢視量表題組是否符合 Mokken 量表模型的基本假設，於符合假設的條件下進行 Mokken 量表計分。此外，並以歷年問卷均包含的相關基本變項來討論其對於語言學習和使用態度的關聯性，同時呈現歷年的變遷趨勢。本研究亦針對分析結果與 Mokken 量表之應用提出討論。

（一）資料來源

分析資料取自於「台灣社會變遷基本調查」（以下簡稱「社會變遷調查」），此一調查計畫為國內歷史悠久之大型學術調查，以台灣地區設有戶籍，年齡在 18 歲以上的本國籍國民為研究母體，選取具有全國代表性之樣本來進行面對面問卷訪談。本研究所使用的資料為社會變遷調查四期五次（2004 年）（章英華、傅仰止 2006）、五期五次（2009 年）（傅仰止、杜素豪 2010）及六期五次（2014 年）（傅仰止等 2015）三個年度的宗教文化組問卷資料。在抽樣設計上，這三個年度均是以分層多階段等比例隨機抽樣抽樣法（stratified multi-stage probability proportional to size sampling）來選取受訪對象，其中第一階段的鄉鎮市區分層之依據，因社會人口與經濟結構的變遷，在 2005 年有所更新（侯佩君等 2008）外，其餘階段的抽樣設計在三個年度中均相同。三筆調查資料完訪的問卷數量分別為 2004 年為 1,881 份、2009 年為 1,927 份及 2014 年的 1,934 份，完訪率依序為 56%、49.3% 及 58%。

在所選取的調查資料檔中，有關語言學習和使用的題組問題為：「在台灣，國語、台語和客家語是主要的日常用語，原住民各族也有自己的語言。目前對於各種語言的學習和使用有著以下的一些看法，請問您贊不贊成以下的看法？」，其中包含三個題目，¹ 分別為：(a) 我們應該在家裡教小孩故鄉的方言；(b) 學校的老師可以用國語以外的其他日常語言教課（不只在鄉土教學課堂中）；及 (c) 學校老師可

1 相同的題組亦包含於 1999 年社會變遷調查三期四次的文化價值組問卷中，然因考量本研究使用之其他變項並未列於其中，亦或是題目或選項的內容有明顯差異，故僅分析 2004 年至 2014 年的資料。

以用台語教課。原始選項為四點量尺，從 1 至 4 分別為「很贊成」、「有點贊成」、「不太贊成」、「很不贊成」，後續分析調整為 0 到 3，分別與「很不贊成」至「很贊成」的四種反應類型作對應。此外，我們也選取了受訪者的年齡、性別、教育程度，以及在家中與父母交談的語言類型等變項，藉以瞭解相關變項與語言學習和使用態度間的可能關係。針對前者，我們先依循前述說明的步驟來檢測題目是否符合 MHM 的條件，藉由 R 程式語言中的 mokken 套件 (van der Ark 2016)，來判斷這些題目組成 Mokken 量表的適合性。其次，我們利用 Mokken 量表模型計算出的語言學習使用量表分數的高低將受訪者區分為高低分兩群組，再使用 SAS 9.3 的 PROC LOGISTIC 來探討語言學習及使用態度與這些變項的關聯性。相關的 R 與 SAS 程式，請見附錄 1。

(二) 分析模型

在進行 Mokken 量表計分之前，必須先確認組成量表的這些題目符合 MHM 的條件。由於此一題組僅包含三個題目，因此本研究採用前述的檢驗步驟，先確認這三個題目並未違反 MHM 量表之條件。此外，由於量表分數的計算需採用每一題目的回答，因此若量表中任一題目含有無效之回答（包含不知道、拒答、遺漏值等），則該筆資料將不被採計於量表分數中。

針對 Mokken 量表分數的應用，在單調同質性模型 (MHM) 的條件下，學者們證明了當選項為二元計分（亦即 $m=1$ ）的情況下 (Grayson 1988; Huynh 1994; Hemker et al. 1996)，就任意實數 a 且 $u > v$ ，

$$P(\theta > a | X_+ = u) \geq P(\theta > a | X_+ = v)$$

恆成立；也就是說題組的答題總分 X_+ 越高，潛藏特質 θ 較大的可能

性也越高。因此潛藏特質 θ 的大小，大致可以經由答題總分 X_+ 來反應；換言之， X_+ 越大，相對的 θ 也會越大。遺憾的是，在多元計分（亦即 $m > 1$ ）的情況下，上述不等式並不成立。這麼一來 θ 的大小就無法透過 X_+ 的數值來反應，因此 θ 與其他解釋變項之間的關聯性，也就無法藉由建構一個 X_+ 與解釋變項間的線性迴歸模型來進行探討。不過 Van der Ark and Bergama（2010）證明了一個較弱的特性如下：就任意實數 a 及 b 而言，

$$P(\theta > a | X_+ \geq b) \geq P(\theta > a | X_+ < b)$$

仍成立。也就是說在多元計分的情況下，雖然答題總分 X_+ 較高並不意謂潛藏特質 θ 較大的可能性也較高，但是如果將所有的受訪者依答題總分 X_+ 區分成兩群組；亦即高分群與低分群，則由 van der Ark and Bergama（2010）的不等式可知，高分群擁有較高潛藏特質 θ 的可能性會高於低分群；反之，低分群擁有較高潛藏特質 θ 的可能性則會低於高分群。因此，本研究將利用 Mokken 量表所計算出來的分數，將受訪者區分為高分群與低分群兩個群組，分別代表潛藏特質較強或較弱的群組，藉以反映受訪者的語言學習及使用態度的強弱。如此一來，我們便可以利用一個邏輯斯迴歸模型，來檢視年齡、性別、教育程度、與父母交談所使用之語言等變項，與語言學習及使用態度強弱傾向的組成比例之間的關聯性，同時瞭解受訪者強弱傾向的組成比例。此外，由於三筆資料均是各年度的代表性樣本，且為長期重複性調查（longitudinal repeated survey），因此我們也將藉由三個年度的資料，來探討、推估社會整體對於語言學習及使用態度強弱傾向的組成比例，是否會隨著時間的改變而有所變化。

四、分析結果

(一) 量表題目之單調同質性模型 (MHM) 條件檢測

由於所分析之語言學習及使用態度之題組僅包含三個題目，且三個年度的測量方式（包含題目與選項內容）均相同，因此我們以四個步驟來檢驗各年度資料是否滿足單調同質性模型 (MHM) 之基本假設（單一向度、局部獨立性、單調性）(Sijtsma and Molenaar 2002)。首先檢視各題目的原始回答反應。令三個年度的資料中，第 k 年的受訪者對於題目 x_i 選擇第 j 個回答類別（數值）的觀測數量為 n_{ijk} ，其中 $i=1, 2, 3, j=0, 1, 2, 3$ ，且 $k=2004, 2009, 2014$ ，則

n_{ijk} : k 年時，題目 i 回答數值為 j 的觀測數量

n_{i+k} : k 年時，題目 i 的有效觀測數量

$p_{ijk}=n_{ijk}/n_{i+k}$: k 年時，題目 i 回答數值為 j 的回答比例

$p_{i(\geq j)k} = \sum_{j'=1}^3 p_{ij'k}$: k 年時，題目 i 回答數值大於等於 j 的回答比例。

由於 $p_{i(\geq 0)k} = \sum_{j'=0}^3 p_{ij'k} = 1$ 恆成立，因此在後續的討論中，我們將不再針對 $j=0$ 的情況做說明。

表 1 的中間四行為各年度受訪者對於各個題目之回答反應的分布資料，其中每個細格中的三個數字由上至下分別為 n_{ijk} 、 p_{ijk} 及 $p_{i(\geq j)k}$ 。由這些數據可以發現，無論是哪一年度，且不論回答反應 j 為 1、2 或 3，都可以觀察到 $p_{1(\geq j)k} \geq p_{2(\geq j)k} \geq p_{3(\geq j)k}$ 的情況。舉例來說，就 2004 年而言，當回答反應 j 為 1 時， $p_{1(\geq 1)2004} = 0.9935$ ， $p_{2(\geq 1)2004} = 0.9487$ ， $p_{3(\geq 1)2004} = 0.9135$ ，因此 $p_{1(\geq 1)2004} > p_{2(\geq 1)2004} > p_{3(\geq 1)2004}$ 。此外，由右邊第二行的平均分數也可以得知，無論是單一年度的資料 ($\bar{X}_{ik} = \frac{1}{n_{i+k}} \sum_{j=0}^3 j n_{ijk}$ ；

亦即 k 年題目 i 的平均回答分數)，或是三個年度整合的資料 ($\bar{X}_i = \frac{1}{n_{i++}} \sum_k \sum_{j=0}^3 j n_{ijk}$ ；亦即題目 i 的總平均回答分數)，第 1 題的平均數值最高，第 3 題則是最低。綜合前述兩項觀察，大致上可以發現第 1 題似乎最容易取得受訪者的認同；相對而言，第 3 題則最不容易獲得認同。

其次，我們需要檢視兩兩題目間的共變異數是否均為正值。表 2 的第 2 至 4 欄列出各年度任意兩個題目之間的量表係數 H_{ij} 的估計值及標準差，由於 H_{ij} 的估計值皆為正值，因此並沒有違反題目彼此間的共變異數都必須為正值的要求。

此外，表 2 也列出的是個別题目的量表係數 H_i 。在 MHM 的模型假設前提之下，題目 i 的量表係數 H_i 至少須大於 0.3，才會被視為具有足夠的判別能力 (discrimination power)。由於所有题目的總量表係數 H 必定介於 H_i 的最小值與最大值之間，因此 H 至少也需要大於 0.3，才可視為一個可以接受的量表；若量表係數大於 0.5，則可視為相當不錯的量表。綜合表 1 及表 2 的數據可以推論，由這三個题目所

表 1 語言學習及使用態度之回答分佈：2004、2009、2014 年度

年度	回答	0	1	2	3	平均數	樣本數
	題目						
2004	1. 應該在家教故鄉方言	12 (0.65%) (100%)	62 (3.34%) (99.35%)	511 (27.53%) (96.01%)	1271 (68.48%) (68.48%)	2.64	1856
	2. 老師可用國語以外的日常語言教課	94 (5.13%) (100%)	279 (15.23%) (94.87%)	713 (38.92%) (79.64%)	746 (40.72%) (40.72%)	2.15	1832
	3. 老師可以用台語教課	159 (8.65%) (100%)	366 (19.90%) (91.35%)	678 (36.87%) (71.45%)	636 (34.58%) (34.58%)	1.97	1839

表 1 (續)

年度	回答 題目	0	1	2	3	平均數	樣本數
2009	1. 應該在家教 故鄉方言	6 (0.31%) (100%)	56 (2.93%) (99.69%)	537 (28.09%) (96.76%)	1313 (68.67%) (68.67%)	2.65	1,912
	2. 老師可用國 語以外的日 常語言教課	77 (4.04%) (100%)	275 (14.43%) (95.96%)	664 (34.84%) (81.53%)	890 (46.69%) (46.69%)	2.24	1,906
	3. 老師可以用 台語教課	118 (6.17%) (100%)	405 (21.18%) (93.83%)	661 (34.57%) (72.65%)	728 (38.08%) (38.08%)	2.05	1,912
2014	1. 應該在家教 故鄉方言	5 (0.27%) (100%)	82 (4.37%) (99.73%)	561 (29.88%) (95.36%)	1229 (65.48%) (65.48%)	2.61	1,877
	2. 老師可用國 語以外的日 常語言教課	82 (4.41%) (100%)	267 (14.37%) (95.59%)	714 (38.43%) (81.22%)	795 (42.79%) (42.79%)	2.20	1,858
	3. 老師可以用 台語教課	121 (6.52%) (100%)	331 (17.85%) (93.48%)	720 (38.81%) (75.63%)	683 (36.82%) (36.82%)	2.06	1,855
總 樣 本	1. 應該在家教 故鄉方言	23 (0.41%) (100%)	200 (3.54%) (99.59%)	1609 (28.50%) (96.05%)	3813 (67.55%) (67.55%)	2.63	5,645
	2. 老師可用國 語以外的日 常語言教課	253 (4.52%) (100%)	821 (14.67%) (95.48%)	2091 (37.37%) (80.81%)	2431 (43.44%) (43.44%)	2.20	5,596
	3. 老師可以用 台語教課	398 (7.10%) (100%)	1102 (19.66%) (92.90%)	2059 (36.73%) (73.24%)	2047 (36.51%) (36.51%)	2.03	5,606

註：中間四行每一細格中的三個數字由上至下分別為 n_{ijk} 、 p_{ijk} 及 $p_{i(\geq)jk}$ 。回答數值分別為 0：很不贊成、1：不贊成、2：贊成、3：很贊成。

表 2 各年度題目之 Mokken 量表係數

year	H_{12} (se)	H_{13} (se)	H_{23} (se)	H_1 (se)	H_2 (se)	H_3 (se)	H (se)
2004	0.442 (0.029)	0.359 (0.028)	0.754 (0.018)	0.398 (0.026)	0.644 (0.016)	0.604 (0.017)	0.563 (0.017)
2009	0.431 (0.029)	0.327 (0.027)	0.784 (0.018)	0.376 (0.026)	0.659 (0.016)	0.611 (0.016)	0.566 (0.017)
2014	0.438 (0.028)	0.352 (0.027)	0.759 (0.017)	0.394 (0.025)	0.641 (0.015)	0.604 (0.017)	0.560 (0.017)
All	0.436 (0.016)	0.346 (0.016)	0.766 (0.010)	0.389 (0.015)	0.648 (0.009)	0.607 (0.010)	0.563 (0.010)

建構的量表應該沒有違反單一向度的假設，並具有足夠的鑑別力。

至於單調性之假設，圖 1(a) 至 (c) 的左圖分別呈現第 1 題至第 3 題的試題階段反應函數 (ISRFs) $P(X_i \geq j | \theta)$, $i=1, 2, 3$ 的估計值，其中圖內的三曲線，由上至下分別是 $P(X_i \geq 1 | \theta)$, $P(X_i \geq 2 | \theta)$, $P(X_i \geq 3 | \theta)$ ；圖 1(a) 至 (c) 的右圖則是試題反應函數 (IRFs) 的估計值 (Sijtsma and Molenaar 2002)。由這些圖形來看，可以得知第 2、3 題並未違反單調性的假設；但是第 1 題似乎有違反單調性的可能，亦即在 θ 值較小處， $P(X_1 \geq 3 | \theta)$ 似乎並未呈現單調遞增的現象。不過檢視實際分析數據可以發現 (結果未呈現)，違反單調性假設處，其 crit 值為 40，依據 Molenaar and Sijtsma (2000) 的建議，當 crit 值大於 80 時才需要留意，小於 40 時，基本上可以視為是隨機誤差所導致，因此第 1 題應該也可以視為符合假設。所謂的 crit 值，是 Molenaar and Sijtsma (2000) 所建議用以判斷單調性假設是否違反的一個工具，這是綜合多項違反單調性假設指標的一個加權平均值，其值越大代表單調性假設的違反情事越形嚴重。

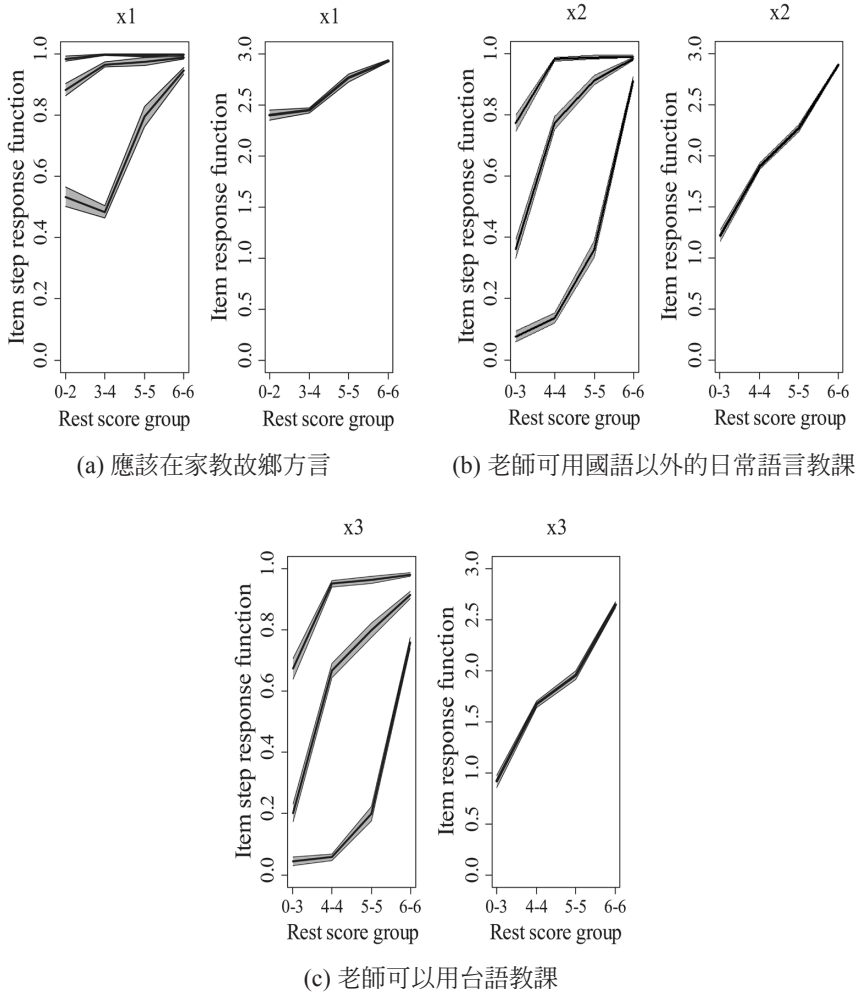


圖 1 單調性檢定結果

綜合上述的檢測，可確認本研究所採用的語言學習及使用態度題組，符合 Mokken 量表模型之 MHM 的基本假設。在此前提下，我們便可以將受訪者依答題總分來區分為高、低分兩群組，這是因為 $P(\theta > a$

$|X_+ \geq b) \geq P(\theta > a | X_+ < b)$ 這一不等式告訴我們，相較於答題總分較低的群組，答題總分較高的族群具有較高潛藏特質的可能性也越高。也就是說，高分群較可能具有較高的潛藏特質，而低分群則較不可能具有較高的潛藏特質。因此儘管潛藏特質是無法觀測得到，不過我們卻可以透過答題總分將受訪者區分為高低分的群組，藉以間接反映受訪者潛藏特質的強或弱。由於我們的最終目的是想要瞭解哪種特徵的受訪者較可能具有較強的潛藏特質，因此透過高低分群的區別，我們便能夠間接探討受訪者個人變項及家中主要語言使用，與語言學習及使用態度的關係，以及歷年的變遷。本文中，我們依據答題總分，將分數區分為高分群 ($X_+ \geq 6$) 及低分群 ($X_+ < 6$)，所占比例分別為 77.85% 及 22.15%；並以低分群做為對照組，進行二元邏輯斯迴歸分析。嚴格來說，此處我們採用 6 分做為高低分群的分界線，並沒有任何的理論依據。不過就這組資料而言，由於改採其他分數作為分群界線所得出來的結果並沒有明顯差異，因此在後續的討論中，我們將依以 6 分做為高低分群的分界線所得到的分析結果，進行說明。

（二）Mokken 量表分數的應用：語言學習及使用態度之變遷

由於本研究所使用的「社會變遷調查」資料，在歷年來均是以相同的抽樣設計與資料蒐集方式來獲得具有全國代表性樣本的訪問資料，因此這三期資料的樣本結構十分類似，茲將各變項的次數分布整理於附錄 2，並簡要說明如下。

在受訪者性別的部分，無論是哪一個年度的資料，都是男性的比例（約為 51%）稍高於女性；年齡的分布上，則是僅 2009 年並無 19 歲（含）以下的受訪者，其餘年齡層的分布趨勢在三個年度均相似。受訪者的教育程度分布在三個年度則有明顯差異；小學以下教育程度

者，在 2004 年為 28.1%，到了 2014 年則降至 18.5%。另一方面，教育程度為專科以上的受訪者，則從 2004 年的 31.0%，至 2014 年增加為 45.9%，整體的教育程度分布隨著調查時間越接近現代而越高，也反映出近代高等教育擴張的結果。

除了個人基本變項外，亦選用了受訪者在家中和父母交談主要使用的語言，來瞭解其與個人對語言學習及使用態度的關係，該變項的次數分配亦列於附錄 1。從歷年的資料可知台語仍是最多人在家中和父母交談所使用的語言，但是在比例上則呈現下降的趨勢（2004 年為 73.2%，2014 年為 63.3%）；相對地，使用語言為國語的比例則有上升的趨勢。客家話的使用比例在這些年之間並無變動，其他語言的使用比例在不同的調查年度則互有高低。

在邏輯斯迴歸的模型中，依變項如前所述，是將 Mokken 量表的答題總分以 6 分為切點分成高低兩群；自變項的部分，調查年度、性別、年齡、教育程度、家中使用語言等變項，均以虛擬變項的方式帶入，其中年齡重新分組為 30 歲（含）以下、31-50 歲，及 51 歲（含）以上，家中使用語言則是將原住民語、家鄉話及其他歸於同一類。各變項的參照組分別為：調查年度為 2004 年、女性、30 歲（含）以下、小學（含）以下，及在家和父母交談主要使用語言為國語。任一變項含有不知道／忘記了、拒答等無反應之資料，皆視為遺漏值而從分析中排除。此外，分析模型考慮調查年度及家中使用語言與個人背景變項之間的交互作用效果，除嘗試瞭解可能的調節效果（moderating effect）外，亦為了使模型盡可能達到較佳的配適度（goodness-of-fit），即 Hosmer-Lemeshow 檢驗結果之 p 值需大於 0.05。

邏輯斯迴歸模型的分析結果列於表 3，這個模型的適合度檢定的 p 值為 0.1041，達到配適度標準。在所檢視的變項當中，性別並無顯

表 3 語言學習及使用態度的潛藏特質迴歸分析結果

變 項	係數	標準誤
調查年：2009	0.4790	0.3505
調查年：2014	1.0429	0.3495
性別	0.3336	0.0686
年齡：31-50 歲	32.2329***	0.1675
年齡：51 歲以上	57.1001***	0.2023
交互作用：調查年 2009* 年齡 31-50 歲	1.0754	0.2384
交互作用：調查年 2009* 年齡 51 歲以上	2.3044	0.2788
交互作用：調查年 2014* 年齡 31-50 歲	17.0959***	0.2360
交互作用：調查年 2014* 年齡 51 歲以上	10.8172**	0.2767
語言：台語	75.3931***	0.1684
語言：客家語	1.2987	0.2422
語言：原住民、家鄉話、其他	9.7029**	0.2501
交互作用：調查年 2009* 語言—台語	0.0248	0.2303
交互作用：調查年 2009* 語言—客家語	2.2899	0.3483
交互作用：調查年 2009* 語言—原住民語、家鄉話、其他	4.4161*	0.3238
交互作用：調查年 2014* 語言—台語	10.8994**	0.2275
交互作用：調查年 2014* 語言—客家語	0.0779	0.3554
交互作用：調查年 2014* 語言—原住民語、家鄉話、其他	6.2750*	0.3722
教育：國中	6.8347***	0.2200
教育：高中	29.9696***	0.1970
教育：專科以上	25.8454***	0.2078
交互作用：調查年 2009* 教育—國中	0.7539	0.3217
交互作用：調查年 2009* 教育—高中	7.5658***	0.2847
交互作用：調查年 2009* 教育—專科以上	1.4656	0.2894
交互作用：調查年 2014* 教育—國中	2.5307	0.3305
交互作用：調查年 2014* 教育—高中	6.2808*	0.2827
交互作用：調查年 2014* 教育—專科以上	4.6379*	0.2906

* $p < .05$; ** $p < .01$; *** $p < .001$ 。

著效果；年齡、教育程度、家中使用語言除了有顯著的主要效果外，這些變項與調查年度的交互作用效果亦達統計顯著水準，因此接續以交互作用效果為主來詮釋分析結果，並以圖形表示以利說明。以調查年度與年齡的交互作用效果而言，由於依據性別、家中使用語言、教育程度的任意組合所繪製的圖形結構完全相同，因此圖 2 僅呈現給定性別、家中使用語言、教育程度的數值均為 1 的情況下，不同年齡組在各期資料對於語言學習及使用態度之 Mokken 量表分數在 6 分及以上的高分群所占有的比例來做說明。三期資料全都顯示出年輕族群中高分群所占有的比例最高，亦即年輕受訪者贊成所提問類型的語言學習及使用方式的認同比例最高；中年族群次之，老年族群具有最低的認同比例。不過由於年輕族群的認同比例似乎略呈下降的趨勢，但中、老年族群反而呈現明顯的上升趨勢，因此三者之間的差異呈現逐漸縮

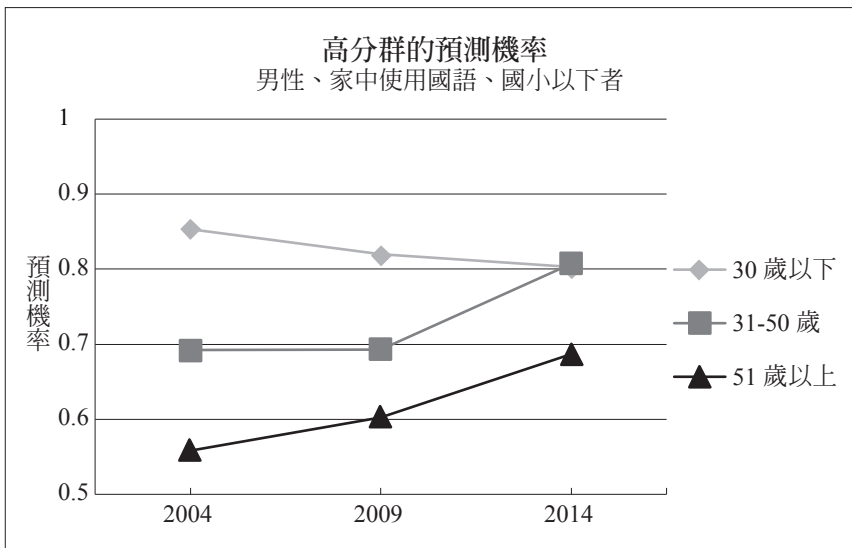


圖 2 調查年度與年齡組的交互作用效果

小的情況。

在調查年度與家中使用語言的交互作用效果部分，由於其他變數組合情形下的圖形結構並無不同，圖 3 僅呈現給定性別、年齡、教育程度數值均為 1 的情況下，不同的使用語言在各期資料對於語言學習及使用態度為高分群所占有的比例做說明。由圖 3 可以看出，三期資料均顯示出在家與父母交談使用台語的受訪者，贊成此類語言學習及使用方式的認同程度最高，但是似乎隨著調查年度越近期而略為呈現下降趨勢。使用客家語的族群在贊成態度上的認同比例則似乎略呈上升再下降的趨勢，而使用國語的族群似乎略呈下降的趨勢，不過這兩個族群的變化趨勢並不明顯；反倒是使用其他語言的族群下降情況相當明顯。

同理，在調查年度與教育程度的交互作用效果上，其他變數組合

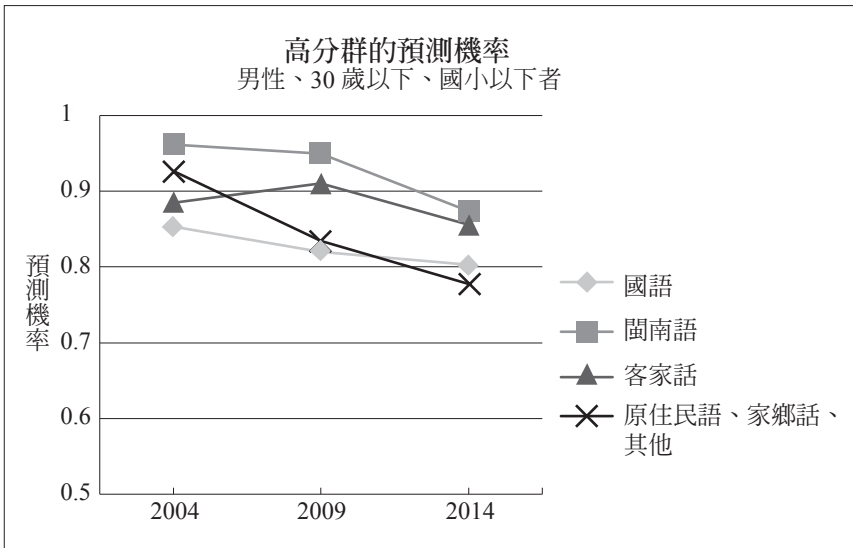


圖 3 調查年度與家中使用語言的交互作用效果

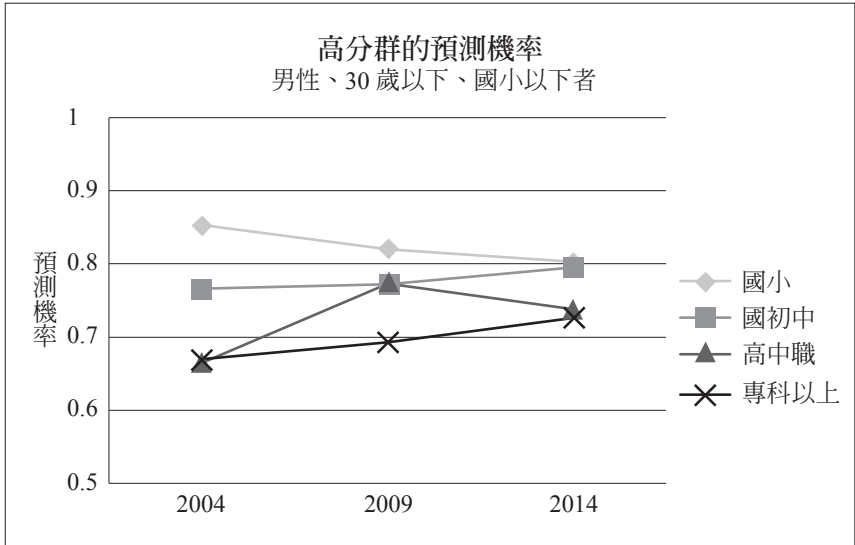


圖 4 調查年度與教育程度的交互作用效果

情形下的圖形結構也相似，因此我們僅以圖 4 來做說明。在給定性別、年齡、家中使用語言數值均為 1 的情況下，由圖 4 可以發現，這三期的資料都顯示出教育程度為小學及以下學歷者認同所提問的語言學習及使用方式的比例圖 4，而專科以上教育程度者的認同比例則是最低。不過由於小學以下教育程度者的認同比例呈現略微下降的趨勢差距，但是專科以上教育程度者反而呈現明顯的上升趨勢，因此教育程度差異的不同反應，有隨著調查年度的不同而呈現逐漸縮小的情況。

五、結論

一般符合葛特曼量表形式的問卷資料多以二元計分的方式處理，且未從機率觀點考慮回答類型的組合，如此一來容易因模型條件的限

制而使分析策略變得僵化，也忽略了可能的測量誤差。利用項目反應理論的無母數模型所發展出來的 Mokken 量表模型，藉由機率值的大小來反映題目間既定的難易程度，並同時保有多元計分的特性，故其具備單一向度、局部獨立性，與單調性的基本假設。雖然 Mokken 量表的發展已久，並且在社會科學的應用也日趨多見，但就量表模型之概念與應用而言，國內仍屬罕見。本研究介紹 Mokken 量表模型之內涵，並利用「台灣社會變遷基本調查」的歷年資料檔中有關語言學習及使用態度之量表，以保留原始多元計分的方式，進行模型檢測與量表分數之實證應用。

從量表題組的模型檢測過程可知，在考量回答機率的條件下，僅包含三個題目的量表仍可能出現不符合檢測條件的狀況。雖然針對我們所分析的資料來說，依據 Molenaar and Sijtsma (2000) 的建議，可以合理地將其視為隨機誤差，不過當量表題組包含數量較多的題目時，極可能使得檢測結果出現更多不符合假設條件的情形。模型假設吻合與否，是量表分析結果可靠與否的重要關鍵之一，因此需提醒讀者，當題組中的題目數量增加時，複雜性也必定提高，進而影響檢測結果，故在使用上需更為謹慎。

在實證分析的結果方面，我們依據 Mokken 量表的答題總分將受訪者區分為高低兩群，進行邏輯斯迴歸分析，來瞭解個人基本變項及家中使用語言與 Mokken 量表分數之關聯性，並檢視不同調查年度的變化趨勢。分析結果指出，個人背景及家中使用語言與潛藏態度的關係則會因調查年度的不同而有差異。此處將受訪者依據答題總分區分為高低兩群的做法，與直接視題組為一李克特量表並加總其總分再做分群的常見方式，似乎沒甚麼不同；然而要強調的是，本研究所分析的題組為經過驗證後符合 MHM 模型的一個 Mokken 量表，因此可以

確認答題總分的確能夠做為一項合理的分析工具，而非直觀地認定量表為一李克特量表，逕自加總總分來進行分析，此為本研究的貢獻之一。此一結果也間接提供李克特量表問卷之使用者，進行量表分析前或可先進行 MHM 的模型檢驗，以確保使用總分作為分析工具的適當性。

整體而言，Mokken 量表除具備傳統葛特曼量表的單一向度特性，以及潛藏結構之機率概念外，亦同時考量题目的二元計分與多元計分模式，在實證研究上有其應用價值。本研究以 Mokken 量表中的單調同質性模型（MHM）做為檢測量表組成之各题目的條件，並且是以多元計分的方式進行檢測，不同於過往研究以二元計分的方式來分析（如 Coromina and Saris 2012; Loner 2016; Tai and Treas 2013），此為本研究的另一項貢獻，也為測驗題組的多元計分模式展現更多的可能性。

由於在實務上問卷题目是以多元計分來設計為主回答類別，未來可考慮加入雙重單調性模型（DMM）的檢測條件，更深入探究 Mokken 量表的應用。因 DMM 實屬於 MHM 的特例，在實證分析上更容易遇到基本假設無法滿足的情形（Sijtsma and Molenaar 2002），因此可預期相關研究的困難度也可能較高。本研究在分析過程確實也遇到此一困境，經檢測後發現所分析探討的題組顯然違反了 DMM 模型的第四個假設，亦即題目順序不變（IIO）的模型假設。不過由於高低分群的分群作法只須滿足 MHM 的模型假設即可，因此本文的分析結果並不會因為無法滿足 IIO 的模型假設，而有所改變。由於本研究所採用之題組僅包含了三個题目，分析結果可推論的範圍也有所侷限，後續可藉由檢測更多题目所組成的量表，來探究量表題組滿足 DMM 假設的可能性。

附錄 1 分析程式

R

```
> library(xlsx)
> x=read.xlsx("D:/MSA/R/msa14.xlsx", sheetIndex=1, header=TRUE)
> xx1=cbind(x[,1],x[,2],x[,3])
> local({pkg<-select.list(sort(.packages(all.available=TRUE))),graphics=
  TRUE)
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
Loading required package: poLCA
Loading required package: scatterplot3d
Loading required package: MASS
> coefH(xx1)
> aisp(xx1)
> summary(check.iio(xx1))
> summary(check.monotonicity(xx1))
> plot(check.monotonicity(xx1),item=1)
> plot(check.monotonicity(xx1),item=2)
> plot(check.monotonicity(xx1),item=3)
> summary(check.pmatrix(xx1))
> plot(check.pmatrix(xx1),pmatrix='ppp',item=1)
> plot(check.pmatrix(xx1),pmatrix='ppp',item=2)
> plot(check.pmatrix(xx1),pmatrix='ppp',item=3)
> plot(check.pmatrix(xx1),pmatrix='pmm',item=1)
```

```
> plot(check.pmatrix(xx1),pmatrix='pmm',item=2)
> plot(check.pmatrix(xx1),pmatrix='pmm',item=3)
> check.reliability(xx1)
> plot(check.restscore(xx1),item=1)
> plot(check.restscore(xx1),item=2)
> plot(check.restscore(xx1),item=3)
```

SAS

```
PROC LOGISTIC PLOTS (ONLY MAXPOINTS=NONE)=(EFFECT(X
=(year) SLICEBY=edu));
```

```
CLASS year(REF='2004') gender(REF='1') agegroup(REF='1')
language(REF='1') edu(REF='1')/PARAM=REF;
```

```
MODEL sumgroup(EVENT='1')=year gender year|agegroup
year|language year|edu/AGGREGATE SCALE=NONE LACKFIT;
```

```
RUN;
```


附錄 2 分析變項之次數分配

變 項		卷 別		四期五次 2004		五期五次 2009		六期五次 2014	
		次數	百分比	次數	百分比	次數	百分比		
性別	男	969	51.5	982	51.0	987	51.0		
	女	912	48.5	945	49.0	947	49.0		
年齡	19 歲（含）以下	42	2.2	0	0	42	2.2		
	20-29 歲	387	20.6	389	20.2	347	17.9		
	30-39 歲	364	19.4	338	17.5	359	18.6		
	40-49 歲	386	20.5	404	21.0	334	17.3		
	50-59 歲	323	17.2	372	19.3	386	20.0		
	60-69 歲	192	10.2	219	11.4	253	13.1		
	70-79 歲	131	7.0	138	7.2	145	7.4		
	80 歲（含）以上	56	3.0	67	3.5	68	3.6		
	不知道／忘記了	0	0	0	0	0	0		
	拒答	0	0	0	0	0	0		
教育程度	小學以下	528	28.1	397	20.6	357	18.5		
	國初中	255	13.6	239	12.4	209	10.8		
	高中職	507	27	514	26.7	476	24.6		
	專科以上	584	31	777	40.3	888	45.9		
	其他	6	0.3	0	0	3	0.2		
	不知道／忘記了	0	0	0	0	0	0		
	拒答	1	0.1	0	0	1	0.1		
在家中與 父母交談 使用語言	國語	251	13.3	394	20.4	500	25.9		
	台語	1376	73.2	1242	64.5	1225	63.3		
	客家話	126	6.7	129	6.7	127	6.6		
	原住民語	2	0.1	20	1.0	5	0.3		
	家鄉話	39	2.1	33	1.7	23	1.2		
	其他	84	4.5	109	5.7	54	2.8		
	不知道	2	0.1	0	0	0	0		
	拒答	1	0.1	0	0	0	0		
總樣本數（N）		1881	100.0	1927	100.0	1934	100.0		

資料來源：「台灣社會變遷基本調查計畫」第六期第五次調查計畫執行報告，附錄四（傅仰止等 2015）；教育程度為重新整併後之次數分配。

參考文獻

- 王文中，2004，〈Rasch 測量理論與其在教育和心理之應用〉。《教育與心理研究》27(4): 637-694。
- 王衛東，2010，〈東亞的社會距離與全球化〉。《教學與研究》6: 51-56。
- 江怡萱，2014，〈摩根量表分析：多元計分試題下受試者潛在特質排序之相關探討〉。
台北：國立政治大學統計學系碩士論文。
- 余民寧，2009，〈試題反應理論（IRT）及其應用〉。台北：心理出版社。
- 侯佩君、杜素豪、廖培珊、洪永泰、章英華，2008，〈台灣鄉鎮市區類型之研究：「台灣社會變遷基本調查」第五期計畫之抽樣分層效果分析〉。《調查研究—方法與應用》23: 7-32。
- 高千涵，2012，〈試題偏難或偏易情況下之 Mokken 尺度量表問項篩選分析〉。台北：國立政治大學統計學系碩士論文。
- 高野祐一、角田信太郎、村木正昭，2014，〈ノンパラメトリック項目反応理論のための数理最適化モデル〉。《数理解析研究所講究録》1879: 180-191。
- 章英華、傅仰止，2006，台灣社會變遷基本調查計畫 2004 第四期第五次：宗教文化組（C00112_2）【原始數據】（取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫 DOI: 10.6141/TW-SRDA-C00112_2-1，取用日期：2016年6月1日）
- 傅仰止、杜素豪，2010，台灣社會變遷基本調查計畫 2009 第五期第五次：宗教組（C00157_2）【原始數據】（取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫 DOI: 10.6141/TW-SRDA-C00157_2-1，取用日期：2016年6月1日）
- 傅仰止、章英華、杜素豪、廖培珊，2015，台灣社會變遷基本調查計畫 2014 第六期第五次：宗教文化組（C00310_2）【原始數據】（取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫 DOI: 10.6141/TW-SRDA-C00310_2-1，取用日期：2016年6月1日）
- 黃雅雯，2012，〈Mokken 尺度量表下潛在特質排序之正確率探討〉。台北：國立政治大學統計學系碩士論文。
- 楊茵筑，2010，〈探討含有缺失資料的 Mokken scale 之量表係數〉。台北：國立台北大學統計學系碩士論文。
- Andrich, David, 1985, "An Elaboration of Guttman Scaling with Rasch Models for Measurement." *Sociological Methodology* 15: 33-80.

- Birnbaum, Allan, 1968, "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability." Pp. 395-479 in *Statistical Theories of Mental Test Scores*, edited by Frederic M. Lord and Melvin R. Novick. Reading, MA: Addison-Wesley.
- Clogg, Clifford C., and Darwin O. Sawyer, 1981, "A Comparison of Alternative Models for Analyzing the Scalability of Response Patterns." *Sociological Methodology* 12: 240-280.
- Coromina, Lluís and Willem E. Saris, 2012, "Measurement of Supranationalism." *Survey Research Methods* 6(2): 77-86.
- Grayson, D. A., 1988, "Two-Group Classification in Latent Trait Theory: Scores with Monotone Likelihood Ratio." *Psychometrika* 53(3): 383-392.
- Green, Bert F., 1956, "A Method of Scalogram Analysis Using Summary Statistics." *Psychometrika* 21(1): 79-88.
- Guttman, Louis, 1944, "A Basis for Scaling Qualitative Data." *American Sociological Review* 9(2): 139-150.
- , 1950, "The Basis for Scalogram Analysis." Pp. 60-90 in *Measurement and Prediction*, edited by Samuel A. Stouffer, Louis Guttman, Edward A. Schuman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen. Princeton, NJ: Princeton University Press.
- Hemker, Bas T., Klaas Sijtsma, Ivo W. Molenaar, and Brian W. Junker, 1996, "Polytomous IRT Models and Monotone Likelihood Ratio of the Total Score." *Psychometrika* 61(4): 679-693.
- Huynh Huynh, 1994, "A New Proof for Monotone Likelihood Ratio for the Sum of Independent Bernoulli Random Variables." *Psychometrika* 59(1): 77-79.
- Liao, Pei-shan and Su-hao Tu, 2006, "Examining the Scalability of Intimacy Permissiveness Scale in Taiwan." *Social Indicators Research* 76(2): 207-232.
- Loevinger, Jane, 1948, "The Technic of Homogeneous Tests Compared with Some Aspects of 'Scale Analysis' and Factor Analysis." *Psychological Bulletin* 45(6): 507-529.
- Loner, Enzo, 2016, "A New Way of Looking at Old Things. An Application of Guttman Errors Analysis to the Study of Environmental Concern." *Quality and Quantity* 50(2): 823-847.
- Maindal, Helle Terkilsaen, Ineta Sokolowski, and Peter Vedsted, 2009, "Translation, Adaptation and Validation of the American Short Form Patient Activation Measure (PAM13) in a Danish Version." *BMC Public Health* 9: 209.
- McCutcheon, Allan L., 1987, *Latent Class Analysis*. Newbury Park, CA: Sage.
- Mokken, Robert J., 1971, *A Theory and Procedure of Scale Analysis with Applications in Political Research*. New York: De Gruyter.
- Mokken, Robert J., and Charles Lewis, 1982, "A Nonparametric Approach to the Analysis of

- Dichotomous Item Responses.” *Applied Psychological Measurement* 6: 417–430.
- Molenaar, Ivo W., 1997, “Nonparametric Models for Polytomous Responses,” Pp. 369–380 in *Handbook of Modern Item Response Theory*, edited by Wim J. van der Linden, and Ronald K. Hambleton. New York: Springer.
- Molenaar, Ivo W., Klass Sijtsma, and P. Boer, 2000, *MSP5 for Windows: A Program for Mokken Scale Analysis for Polytomous Items*. Groningen, Netherlands: ProGamma.
- Nitschke, Joachim, Michael Osterheider, and Andreas Mokros, 2009, “A Cumulative Scale of Severe Sexual Sadism.” *Sexual Abuse: A Journal of Research and Treatment* 21(3): 262–278.
- Rasch, George, 1960, *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen and Lydiche.
- Shenkin, Susan D., Roger Watson, Ken Laidlaw, John M. Starr, and Ian J. Deary, 2014, “The Attitudes to Ageing Questionnaire: Mokken Scaling Analysis.” *PLoS One* 9(6): e99100. doi: 10.1371/journal.pone.0099100.
- Sijtsma, Klaas, and Ivo W. Molenaar, 2002, *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage.
- Sijtsma, Klaas, and B. W. Junker, 1996, “A Survey of Theory and Methods of Invariant Item Ordering.” *British Journal of Mathematical and Statistical Psychology* 49(1): 79–105.
- Tai, Tsui-O, and Judith Treas, 2013, “Housework Task Hierarchies in 32 Countries.” *European Sociological Review* 29(4): 780–791.
- Van der Ark, L. Andries, 2016, *CRAN—Package Mokken*, <http://cran.r-project.org/web/packages/mokken/> (Date visited: June 12, 2016).
- Van der Ark, L. Andries, and Wichor P. Bergsma, 2010, “A Note on Stochastic Ordering of the Latent Trait Using the Sum of Polytomous Item Scores.” *Psychometrika* 75(2): 272–279.
- Van der Veer, Kees, Laurens Higler, Susan Woelders, Reidar Ommundsen, and Regina Pernicec, 2013, “Developing and Validating a Cross-national Cumulative Scale Measuring Attitudes toward Illegal Immigrants.” *International Journal of Social Research Methodology* 16(5): 429–443.
- Van Schuur, Wijbrandt H., 2003, “Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory.” *Political Analysis* 11(2): 139–163.
- Wind, Stefanie A., 2015, “Evaluating the Quality of Analytic Ratings with Mokken Scaling.” *Psychological Test and Assessment Modeling* 57(3): 423–444.