特殊教育研究學刊
民 96，32 卷 4 期，85-109 頁

學前兒童色塊測驗之編製

林月仙
樹德科技大學幼保系助理教授

本研究旨在編製一套用來評量學前兒童聽覺理解能力的色塊測驗，並探討其信效度。研究對象包括預試樣本、常模樣本和身心障礙樣本共 2,257 名 3 至 5 歲兒童，其中常模樣本係根據台灣地區北、中、南、東之學前兒童人口比率，每一年齡組各取 500 名。本研究之主要結果如下：（1）每一年齡組色塊測驗各有 29 題，其中 22 題是共同試題，另外 7 題是各年齡組不同之試題；（2）學前兒童色塊測驗有良好的內部一致性和重測信度；（3）本測驗三歲組和四歲組資料與四因素模式適配度佳，五歲組資料則較適合三因素模式；（4）本測驗與「學前兒童語言能力的測驗」（張欣戊，1991）有中度相關；（5）學前兒童在不同因素試題之通過率，隨指令長度和複雜度之增加而遞減；（6）學前兒童在色塊測驗之表現會因其年齡、性別、居住地區和是否為身心障礙者，而有所差異；（7）色塊測驗所測量的聽覺理解能力，受時代變遷之影響不大。

關鍵詞：色塊測驗、學前兒童、聽覺理解

本研究承國科會專題研究計畫補助（計畫編號：NSC94-2413-H-366-004，NSC95-2413-H-366-002），感謝國立高雄師範大學曾進興教授和吳裕益教授，在測驗編製及論文撰寫方面，提供許多寶貴的意見；以及研究進行中所有提供協助者，包括測驗審題者、協助常模資料蒐集的北區聯絡人台北市龍山國小林興兆校長、中華國際幼兒文教學會邱慶明副理事長和吳宗賢理事，中區聯絡人清水國小附設幼稚園林良薰老師、提供受測幼兒的各園所長、施測的研究生和老師們，以及研究助理沈淑燕小姐、林曉蕙小姐、潘瑀 瑄小姐和邱奐宗先生。

色塊測驗（Token Test）是歐美國家評量成人和兒童聽覺理解功能時，常用的工具。此測驗係由義大利學者 De Renzi 和 Vignolo（1962）於期刊發表構想，並以臨床案例說明 Token Test 比其他評量工具能更敏銳地偵測失語症病人之輕度理解缺陷。由於 De Renzi 和 Vignolo 未將測驗內容與施測程序標準化，因此，陸續有許多研究人員參考 Token Test 的架構設計測驗，應用於失語症患者之聽覺理解評估與研究，四十多年來，色塊測驗在語言障礙群體的評量中應用極為廣泛，版本也相當多。

De Renzi 和 Vignolo（1962）所設計的 Token Test，包括大小兩種尺寸的圓形和長方形，每一尺寸形狀均有紅、黃、綠、藍、白等五種顏色，共 20 個色塊（token），受試者依施測者說出的動作和大小、顏色、形狀等屬性，操作這些色塊。全測驗包括五個分測驗，分測驗一、三、五僅呈現大的色塊，分測驗二和四則呈現全部色塊。分測驗一的指令相當簡單，每一指令中僅有顏色和形狀兩種屬性，例如：「拿黃色長方形」；分測驗二加入「大小」屬性，例如：「拿小的白色長方形」；分測驗三要求受試者拿兩個色塊，例如：「拿紅色圓形和綠色長方形」；分測驗四和分測驗三指令相似，只是增加了「大小」屬性，例如：「拿大的白色圓形和小的綠色長方形」。前四個分測驗藉著逐漸增加色塊屬性提高指令難度，每一個分測驗的動詞、句型和語句長度相同，受試者只需將注意集中在色塊屬性即可正確完成指令；分測驗五藉由加入介係詞、連接詞和副詞等文法詞彙變化指令的複雜度，例如：「把紅色圓形放在綠色長方形上面」、「如果有黑色圓形，就拿紅色長方形」、「指黃色圓形之前，先拿紅色長方形」等等。相較於前四個分測驗，分測驗五的指令較為多樣化，受試者需對整個指令全面理解方能正確作答，所需的注意、聽覺記憶及統整能力較高。

De Renzi 和 Vignolo（1962）發表 Token Test 之後，許多研究人員以原版的 Token Test 為基礎發展色塊測驗，改編情形可分為下列六項：（1）改變評分方式，例如 McNeil 和 Prescott（1978）將 0、1 二元計分，改為 1 至 15 多元計分；Orgass 和 Poeck（1966）計算受試者的錯誤數。（2）改變動詞，Boller 和 Vignolo（1966）將「拿起」（pick up）改為「碰觸」（touch）。（3）改變色塊的形狀，Spreen 和 Benton（1969, 引自 Spellacy & Spreen, 1969: 392）把「長方形」改為「正方形」。（4）改變色塊的顏色，例如 De Renzi（1979, 引自 Davis, 1983: 192）將「藍色」改為「黑色」；McNeil 和 Prescott 以「黑色」取代「黃色」。（5）增減題數或分測驗數，例如 McNeil 和 Prescott 的 Revised Token Test（以下簡稱 RTT）有 10 個分測驗 100 題；Spreen 和 Benton 的色塊測驗有 6 個分測驗 39 題。（6）改變施測方式，以電腦螢幕呈現色塊，藉由電腦語音播出事先錄製的指令，受試者聽到指令後以滑鼠或觸控螢幕點選色塊（D'Arcy & Connolly,1999; Eberwein, Pratt, McNeil, Fossett, Szuminsky, & Doyle, 2007; Sapir & Merkin, 2004, 引自 Cohen-Mimran & Sapir, 2007: 173），或以眼神移動（eye movements）的方式作答（Hallowell, Wertz, & Kruse, 2002）。

大多數版本色塊測驗都維持原版 Token Test 的五個分測驗，前四個分測驗各有 10 題，以同等難度詞彙組成相同句型和句長的指令，第五分測驗加入文法詞彙變化指令複雜度，僅在題數和指令內容稍加變化。少部分研究增減分測驗數和題數，並探討其信效度，例如，Spreen 和 Benton（1969，引自 Spellacy & Spreen, 1969: 392）將其他版本用來確認受試者認識顏色和形狀的 7 題準備測驗（pretest），列為第一分測驗，第二至第六分測驗題型和多數版本的一至五分測驗相同；不同的是題數，二至五分測驗各有 4 題，第六分測驗有 16 題，

總題數僅有 39 題。Spellacy 和 Spreen（1969）以 103 名腦傷的失語症和非失語症成人為施測對象，分析其內部一致性 α 係數為.92，能正確區辨 79%的失語症患者；接著使用與總分相關高的前 16 題分析，得到的 α 係數為.87，正確區辨失語症患者百分比為 84%，因此，Spellacy 和 Spreen 的結論是減少題數之後的色塊測驗仍具不錯的信效度。另外，McNeil 和 Prescott（1978）發展的 RTT 包括十個分測驗，每一分測驗各有 10 題共 100 題，並正式出版發行，且廣泛應用於失語症患者聽覺理解能力之評量與研究。其後，Arvedson, McNeil 和 West（1985）又從 RTT 的九個分測驗（第九分測驗除外），分別選取前 5 題，結合第九分測驗的 10 題，組成 55 題的簡版 RTT，名為 Five-Item Revised Token Test，Arvedson 等人研究結果，發現受試者之簡版 RTT 得分和標準版(100 題）得分的相關係數為 0.99；Park, McNeil 和 Tompkins （2000）進一步檢視簡版 RTT 之重測、評量者間及評量者內信度，Hula, Doyle, McNeil 和 Mikolic （2006）則探討簡版 RTT 的效度，研究結果顯示簡版 RTT 具有良好的信效度，其施測時間大約僅需標準版的一半。

應用色塊測驗評量兒童聽覺理解能力始於 1960 年代末期（Noll & Berry, 1969; Noll, 1970;均引自 McNeil & Prescott, 1978: 6）。到了 1970 年代晚期，DiSimoni（1978）將色塊測驗應用於兒童的施測，名為 Token Test for Children（以下簡稱 TTFC），並建立 3 至 12 歲兒童之年齡常模和年級常模，且正式出版發行。由於 TTFC 的施測方式、題數、指令內容與大多數版本的 Token Test 相似，因此，以兒童為對象的研究，有使用 Token Test 者（例如：Kurnatowski, Putyński, Lapienis, & Kowalska, 2006; Shiota, Koeda, & Takeshita, 2000），也有使用 TTFC 者（例如：Keith & Engineer, 1991; Yliherva, Olsén, & Järvelin, 2001）。

兒童聽覺理解障礙之本質可從語言學和心理學的觀點來探討（曾進興，1996）。語言學的觀點重視兒童語言知識結構之不足，例如：音韻（phonology）、語意（semantics）、語法（syntax）、語用（pragmatics）等語言表徵和規則運算（Lahey, 1988; Rescorla & Lee, 2000）。心理學觀點則強調兒童獲得和使用語言時心理歷程的缺陷，包括注意力、聽覺記憶、編碼、轉碼、提取、統整等（Kirk, McCarthy, & Kirk, 1968; Myklebust, 1954, 引自 Blosser, Weidner, & Dinero, 1976: 291）。兒童聽覺理解障礙評量的方式也因此有語言學和心理學兩種取向，依據語言學觀點設計的兒童聽覺理解能力評量，國內常見的有：以詞彙理解為主的「修訂畢保德圖畫詞彙測驗」（Peabody Picture Vocabulary Test-Revised, PPVT-R，陸莉、劉鴻香，1994），評量句法和語意理解的「修訂學前兒童語言障礙評量表」（林寶貴、杞昭安、李芃娟、林玉霞、曹純瓊、黃玉枝、黃桂君、錡寶香、韓福榮等人，2007）和「學前兒童語言能力的測驗」（張欣戊，1991）。

心理學取向的評量，因聽覺理解能力易受語句脈絡（context）、常識及詞彙量多寡之影響（Buck, 2001; Lahey, 1988; Paul, 1990），受試者在理解歷程中，會應用常識和語句脈絡猜測，彌補其不注意或記憶廣度不足所漏失的訊息（Miller & Paul, 1995; Paul, 1990）。若要評量受試者聽覺理解歷程之注意、聽覺記憶和統整能力，需盡可能降低評量材料的語言知識難度，減少語句中不必要的語言線索，例如：魏氏兒童智力量表之記憶廣度測驗，藉由受試者順向與逆向複誦施測者口述的數字，評量其記憶廣度和執行功能（executive function），但數字和口語性質不同，兩者間僅有中度相關，數字記憶廣度測驗不足以評量受試者之聽覺理解缺陷（Cohen-Mimran & Sapir, 2007）。另外，複誦除了理解還涉及表達，不適合用來評量沒有

口語或口語量極少之受試者。色塊測驗藉由遵從指令操作色塊，評量受試者之聽覺理解與執行功能，不受表達能力之限制，而且色塊測驗將語言知識減至最低，僅使用圓形、方形、大、小、紅、黃、綠、藍、白等簡單的詞彙成分，組成結構簡潔的常用語句，降低語言知識在聽覺理解歷程中所產生的影響，藉由逐步增加詞彙數延伸語句長度，所增加的詞彙都是受試者易於理解的，因而可將不同句長指令通過率的差異，歸因於受試者之注意和聽覺記憶。此外，色塊測驗也藉著增加介係詞、連接詞、和副詞等變化指令的複雜度，受試者若要正確完成此類指令，需對整個指令完全接收與理解（De Renzi & Vignolo, 1962）。因此，色塊測驗除了可用來評量失語症病患之聽覺理解能力，也適合用來評量腦傷、輕微腦神經發展失能（minor neurodevelopmental dysfunction，簡稱 MND）、語言障礙和學習障礙等輕度障礙兒童聽覺理解歷程之注意、記憶和統整能力（曾進興，1996; Eberwein et al., 2007; Silverman, Raskin, Davidson, & Bloom, 1977; Yliherva et al., 2001）。

整理西文文獻發現色塊測驗應用於兒童或青少年之研究可歸為四類。第一類以語言障礙、輕微腦神經發展失能和閱讀障礙者為對象，探討其聽覺理解能力是否顯著低於一般發展之同儕。研究結果相當一致，語言障礙、輕微腦神經發展失能和閱讀障礙者在色塊測驗之平均得分，皆顯著低於一般兒童（Blosser et al., 1976; Cohen-Mimran & Sapir, 2007; Kurnatowski et al., 2006; Yliherva et al., 2001）。

第二類研究以閱讀障礙或學習障礙者為對象，分析色塊測驗與其他測驗得分之關連性，或比較不同測驗間的區辨敏銳度。結果顯示學習（或閱讀）障礙兒童之色塊測驗得分與魏氏智力量表總分、語文量表、數字記憶廣度分量表間具有中度相關，相關係數介於.33至.61 之間（Cohen-Mimran & Sapir, 2007; Silverman et al., 1977; Wiig, Lapointe, & Semel, 1977）；色塊測驗分數和「西北句法篩檢測驗」（Northwestern Syntax Screening Test）總分相關為.63（Cartwright & Lass, 1974, 引自 DiSimoni, 1978: 3）；學習障礙青少年的色塊測驗得分與「西北句法篩檢測驗」之表達分量表得分相關為.44（Wiig et al., 1977）；Shiota 等人（2000）使用色塊測驗、修訂魏氏兒童智力量表和 Kaufman Assessment Battery for Children（簡稱 K-ABC），探討閱讀障礙兒童的認知特質，結果發現色塊測驗最能偵測閱讀障礙兒童之閱讀理解缺陷。

第三類研究以一般兒童為對象，比較不同性別、不同年齡的受試者在色塊測驗得分之差異。在性別差異方面，研究結果並不一致，Blosser 等人（1976）對 5.5 至 8 歲兒童施測，結果顯示女生色塊測驗平均得分高於男生；但 DiSimoni（1978）研究發現 5 至 12 歲組男女生平均得分未有顯著差異。在年齡方面，Silverman 等人（1977）的研究發現年齡與色塊測驗得分之間有正相關；Orgass 和 Poeck（1966）的研究建議 15 歲以下應分別建立常模，15 歲以上則可以忽略年齡因素。Noll 於 1970 年（引自 Silverman et al., 1977: 105）對幼稚園至小學六年級學童施測（每一年級 25 人），發現受試者之色塊測驗得分會隨年齡遞增，但三年級之後，不同年級受試者之得分平均數就沒有顯著差異了。DiSimoni（1978）的研究並未進行不同年齡間的差異檢定，但在 TTFC 的使用手冊將各年齡組得分與 Noll 的研究資料並列繪製折線圖，兩個研究折線圖趨勢相當一致。

第四類研究以色塊測驗做為特殊兒童介入方案成效的評估工具。例如：Keith 和 Engineer（1991）使用色塊測驗檢視注意力缺陷過動（ADHD）兒童服用利他能（Methylphenidate）時，其聽覺理解能力是否優於未服藥期間，結果顯示受試者服用利他能

期間，色塊測驗得分高於未服藥期間；Keith 和 Engineer 進一步探討受試者服藥期間，在不同分測驗得分進步情形，發現受試者在指令較為簡短的前三個分測驗增加之分數，不如指令較長且較為複雜的第四和第五分測驗。

據研究者所知，台灣的少數醫療院所（例如：台北榮民總醫院、嘉南療養院）有購置 RTT（McNeil & Prescott, 1978），但無本土化常模可供對照，使用得很少。近年來，中山醫學大學的巴甫仁（2007）嘗試將 Eberwein 等人（2007）發展的 Computerized Revised Token Test（簡稱 CRTT）中文化，目前該研究尚在進行中，仍未發行。以色塊測驗評量兒童聽覺理解能力者，僅有「聽覺理解測驗」（曾進興，1996），該測驗屬於「兒童認知功能綜合測驗」中語言測驗的一部份，建有幼稚園大班至國小三年級之年級常模，主要用來篩選及診斷 5 至 8 歲之學習障礙與發展遲緩兒童（陳振宇、謝淑蘭、成戎珠、黃朝慶、洪碧霞、櫻井正二郎、吳裕益、邱上真、陳小娟、曾進興等人，1996）。「聽覺理解測驗」包括三個分測驗，共 29 題，分測驗一和三呈現全部色塊，分測驗二呈現 10 個大的色塊。分測驗一有 9 題，主要在瞭解受試者是否認識色塊的顏色、形狀和大小，例如：「用手指黑色的」、「用手指方塊」、「用手指大的方塊」；分測驗二有 10 題，參考不同版本色塊測驗的分測驗一、三、五，及 RTT 分測驗七和九的指令設計題目，例如：「用手指白色圓形」、「先用手指黑色方塊，再指紅色圓形」、「把綠色方塊放在紅色圓形左邊」。分測驗三也是 10 題，指令類型含括各版本色塊測驗的分測驗二和四，以及 RTT 分測驗六、八和十之題型，例如：「用手指小的綠色圓形」、「把小的白色圓形放在大的綠色方塊右邊」、「如果有大的綠色方塊的話，就把小的紅色圓形放在它上面」。

與其他版本色塊測驗相較，「聽覺理解測驗」（曾進興，1996）有下列五項特色：(1) 將確認受試者瞭解顏色、形狀、大小的基準線測試納入計分。(2) 各版本色塊測驗之前四個分測驗各有 10 題句型和動作相同，色塊屬性相異的指令，「聽覺理解測驗」編製者考量 5 至 8 歲兒童耐心與專注時間有限，句型重複的題目沒有必要那麼多，乃各取兩題，並將各版本色塊測驗加入文法詞彙異質性較高的分測驗題目，配合色塊呈現，安排於分測驗二和三；(3) 將拿兩個色塊的指令，修正為有先後順序；(4) 分測驗二和三各增加一題其他版本沒有的題目，分別為「右手拿起黑色方塊，左手指黑色圓形」、「右手指小的白色圓形，左手握起拳頭」，旨在瞭解受試者是否被色塊有關的指令制約，因而降低注意力。(5) 聽覺理解測驗於使用手冊中明訂色塊間距，其他版本只提供色塊相對位置，並未標明色塊間的距離。

由於「聽覺理解測驗」（曾進興，1996）屬「兒童認知功能綜合測驗」中語言測驗的三個分測驗之一，除了提供年級常模外，其他相關分析均併入語言測驗中，無法看到當時有關該測驗的研究結果。最近有少數研究使用「聽覺理解測驗」進行研究，例如：余玻莉（2007）的研究指出學前身心障礙兒童之「聽覺理解測驗」得分，與「修訂畢保德圖畫詞彙測驗」（陸莉、劉鴻香，1994）的相關為.64，與「修訂學前兒童語言障礙評量表」（林寶貴等，2007）之口語表達、語言理解、表達詞彙之相關，依序為.77、.85 和.61。陳璽琳（2006）使用曾進興（1996）編製的「語詞理解測驗」和「聽覺理解測驗」探討外籍配偶子女之聽覺理解能力，結果指出外籍配偶子女之「語詞理解測驗」平均得分略低於常模，但未達顯著水準，「聽覺理解測驗」平均得分則顯著低於一般同齡兒童。李珊珊（2007）也使用「語詞理解測驗」和「聽覺理解測驗」，探討音韻異常兒童和一般兒童聽覺理解能力之差異情形，結果顯示音

韻異常兒童的「語詞理解測驗」得分平均數略低於常模，但未達顯著水準，「聽覺理解測驗」得分平均數，則顯著低於一般同齡兒童。可以說在區辨輕微聽覺理解能力差異上，「聽覺理解測驗」之敏銳度優於「語詞理解測驗」，與國外研究（Boller & Vignolo, 1966; De Renzi & Vignolo, 1962; Shiota et al., 2000）結果相呼應。

接受語言治療之學前兒童，絕大多數是因為他們有明確的障礙（例如：智能障礙、聽覺障礙、自閉症等），其口語理解與表達能力明顯落後於一般同齡兒童（Rescorla & Lee, 2000）。輕度障礙或發展遲緩兒童在日常生活情境下，當他沒有接收完整的語言訊息時，會藉助理解策略（comprehension strategies），參照情境或語句脈絡等相關線索，做出正確的反應（Miller & Paul, 1995; Paul, 1990），導致家長或教師高估其聽覺理解能力，以致未能早期發現他們聽覺理解困難的問題。直到進入小學，面對較多抽象、去脈絡的（decontextualized）認知情境，才顯現聽覺理解欠佳及連帶的學習問題。前述文獻顯示色塊測驗可敏銳地偵測受試者之輕微理解缺陷，若能善加應用，應可及早發現兒童聽覺理解之異常現象，提供早期介入服務，減少因聽覺理解能力欠佳所衍生之學習與生活適應問題。

國內僅有「聽覺理解測驗」（曾進興，1996）以色塊測驗的模式，評量兒童聽覺理解之注意、聽覺記憶和統整能力。該測驗編製者考量兒童的耐心與專注時間較成人短暫，將多數版本色塊測驗中，句型重複出現的前四個分測驗各取兩題代表，結合分測驗五部分題目，新增兩題其他版本沒有的題目，組成「聽覺理解測驗」。其題目出現順序、題數、測驗結構、施測對象與歐美各版本均有所不同，理應探討其信效度，但囿於它是「兒童認知功能綜合測驗」中「語言測驗」的分測驗之一，除了常模之外，當時有關信效度的研究都併入「語言測驗」中。此外，因「聽覺理解測驗」適用對象為 5 至 8 歲兒童，但我國特殊教育法第九條明訂「特殊教育學生之入學年齡向下延伸至三歲」，發展五歲以下兒童適用的色塊測驗，做為早期發現兒童輕微聽覺理解問題的評量工具，有其必要性；再就常模的新近性來看，「聽覺理解測驗」的常模建立距今十多年，有加以討論與更新的空間。基於前述文獻與背景，本研究之目的在編製一套 3 至 5 歲兒童適用的色塊測驗，並探討其信效度。

一個好的標準化測驗除了要有樣本較大之全國常模，當不同背景變項受試者之得分有顯著差異時，則應提供分組常模（subgroup norm, Groth-Marnat, 2003）。由於色塊測驗相關文獻僅探討不同年齡和性別兒童之得分差異，受試者之色塊測驗得分是否受居住地區影響，未見相關研究。就兒童語言能力之城鄉差異來看，文獻顯示都會區兒童語言能力優於鄉下地區（陸莉、劉鴻香，1994；Bornstein & Cote, 2005），可能與都會區的母親比鄉下地區的母親知道如何和孩子說話，比較常和孩子共讀童書與說故事（Williams, Soetjiningsih & Williams, 2000）有關。色塊測驗評量受試者的聽覺理解是語言能力向度之一，受試者之色塊測驗得分是否如同語言能力有著城鄉差異，是本研究考量城鄉分組常模建立需要性之參考依據。前述文獻顯示不同性別兒童在色塊測驗得分，有的女生高於男生（Blosser et al., 1976），有的無顯著差異（DiSimoni, 1978），以及 Keegstra, Knijff, Post 和 Goorhuis-Brouwer（2007）的研究結果指出有語言問題的男生多於女生，可能與女生的語言學習能力優於男生有關。本研究藉由比較不同性別受試者得分差異情形，決定是否建立不同性別之分組常模。

不同年代同齡受試者測驗得分是否有所差異？色塊測驗相關文獻未加探究。魏氏幼兒智力量表修訂版（Wechsler Preschool and Primary Scale of Intelligence-Revised，以下簡稱 WPPSI-R）之「比較研究」，指出 1989 年的 6 歲至 7 歲幼童無論是全量表、語文分測驗或作業分測驗，均傾向於比 1967 年的同齡者有更好的表現（陳榮華、陳心怡，2000）。本研究編製的學前兒童色塊測驗，五歲組試題修訂自「聽覺理解測驗」（曾進興，1996），保留相同的題數、內容和計分方式，本研究五歲組受試者在色塊測驗之得分，是否和 WPPSI-R 的研究一樣，異於「聽覺理解測驗」之同齡受試者？本研究藉此探討色塊測驗常模適用期間之長短。

研究方法

一、研究設計與研究假設

本研究包括兩部分，一是編製「學前兒童色塊測驗」，另一則在探討該測驗之信效度。效度研究方面，研究者提出 5 項假設：（1）受試者在學前兒童色塊測驗表現之實證資料，與研究者根據指令內容所提出的四因素理論模式適配度佳。（2）受試者之學前兒童色塊測驗及「學前兒童語言能力的測驗」（張欣戊，1991）得分有中度相關。（3）學前兒童在色塊測驗共同試題的通過率隨年齡遞增，五歲組優於四歲組，四歲組優於三歲組。（4）研究對象在色塊測驗之通過率隨指令的長度與複雜度增加而遞減。（5）智能障礙、自閉症、發展遲緩和語言障礙等身心障礙兒童，因生理機能與心智功能的限制，其色塊測驗得分顯著低於一般兒童。在常模建立之相關研究方面，提出 2 項假設：（1）根據前述文獻，本研究假設女生色塊測驗得分高於男生；另外，前述文獻顯示都會區兒童語言能力優於鄉下地區，就台灣社會生態來看，都會區兒童接觸的語言環境較豐富，其色塊測驗得分應優於非都會區兒童。（2）本研究五歲組常模樣本色塊測驗得分平均數，與 11 年前用來建立「聽覺理解測驗」（曾進興，1996）大班常模樣本之平均得分有顯著差異。

二、研究對象

本研究對象包括預試樣本、常模樣本和身心障礙樣本，分別說明如下：

（一）預試樣本

本研究預試樣本取自高雄市和高雄縣 20 家公私立幼托園所，包括三歲組 201 名，四歲組 200 名，五歲組 219 名，共 620 名學前一般兒童，其中男生 311 人，女生 309 人。

（二）常模樣本

本研究建立三、四、五歲常模，為使所選取之樣本具有代表性，每一年齡層 500 人，抽樣時有考量性別和居住地區之分布。在性別方面，盡可能使男女樣本數相當，但因台灣地區幼兒男女比率約為 110：100，故所蒐集的資料也有男生略多於女生的現象。在地區方面，研究者根據內政部戶政司 2006 年各縣市三、四、五歲人數，計算北、中、南、東四區人口比率，分別為 44.05%、26.76%、26.64%、2.52%，再根據人口比率換算各區樣本人數，都會區與非都會區各半，因東部地區人數少，無符合本研究都會區定義者（如表一註），僅有非都會區樣本。常模樣本選自公私立幼托園所，北區有台北市、台北縣、新竹縣、新竹市和桃園縣的 20 家，中區有台中縣、台中市和南投縣的 13 家、南區有高雄市、高雄縣和台南縣的 25 家，東區取樣少僅有台東縣的 1 家，常模樣本人數分配，及各年齡組月齡平均數、標準差如表一。

此外，本研究分別從三、四、五歲組常模樣本中各取 30 名，間隔兩週重測，以探討學前兒童色塊測驗之重測信度；同時也從各組常模樣本中，選取 102 至 159 名受試者，另施以「學前兒童語言能力的測驗」（張欣戊，1991），以探討本測驗之效標關連效度。重測信度和效標關連效度使用的兩群樣本並未重疊。

表一　各年齡組常模樣本人數分布與年齡描述統計（單位：月）

地區別/年齡		三歲組		四歲組		五歲組		合計
		男	女	男	女	男	女	
北部	都會區	62	34	56	54	48	62	316
	非都會區	67	57	59	51	65	45	344
中部	都會區	38	28	32	35	35	31	199
	非都會區	31	37	34	33	41	27	203
南部	都會區	41	24	40	26	39	27	197
	非都會區	42	25	34	32	37	29	199
東部	都會區	0	0	0	0	0	0	0
	非都會區	11	3	7	7	7	7	42
合　計		292	208	262	238	272	228	1500
月齡平均數		44.51	45.19	55.88	56.19	67.78	67.41	
（標準差）		3.66	3.74	4.19	4.23	4.66	4.50	

註：都會區指直轄市、省轄市及人口數超過 30 萬的縣轄市，其餘屬非都會區。

（三）身心障礙樣本

智能障礙、聽覺障礙、自閉症、特定型語言障礙（specific language impairment，簡稱 SLI）和發展遲緩等身心障礙兒童，因生理機能和心智功能的限制，其聽覺理解能力通常比一般同齡兒童低（Lahey, 1988; Miller & Fletcher, 2005; Rescorla & Lee, 2000）。為瞭解學前兒童色塊測驗是否能有效區辨一般與前述五類身心障礙兒童，本研究將就讀幼托園所的發展遲緩、智能障礙、自閉症、聽覺障礙和語言障礙兒童納入施測對象。因三歲和四歲組前述五類身心障礙樣本總數分別僅有 22 和 36 人，各類障礙人數幾乎都少於 10 人，五歲組聽覺障礙人數也僅有 6 人，均不適合進行比較。因此，本研究用來檢視色塊測驗能否能區辨一般兒童和身心障礙兒童，使用的身心障礙樣本為五歲組發展遲緩、智能障礙、自閉症和語言障礙兒童，分別有 69、20、25 和 23 人。

三、研究工具

本研究使用的工具包括「學前兒童色塊測驗」，及「學前兒童語言能力的測驗」（張欣戊，1991），分別說明如下：

（一）學前兒童色塊測驗

學前兒童色塊測驗係本文作者與「聽覺理解測驗」作者，參考 Token Test 之相關文獻，設計用來評量 3 至 5 歲兒童聽覺理解歷程之注意、聽覺記憶及統整能力。本測驗包括三個分測驗：（1）基準線測試（呈現大的圓形和方形），受試者依施測者之指示，「指」某一顏色或形狀的色塊；（2）測驗一（呈現大的圓形和方形），受試者依指示「指」某一顏色形狀的色塊、依序指出兩個不同顏色形狀的色塊、將某一顏色形狀的色塊放在某一顏色形狀色塊的某個位置；（3）測驗二（呈現全部色塊），測驗內容與形式和測驗一相似，只是加入 10 個小的色塊，題目增加「大小」屬性，指令變得較為複雜。

配合學齡前兒童聽覺理解能力發展，三個年齡組試題句型類似，內容略有差異。其中，五歲組題目修訂自「聽覺理解測驗」（曾進興，1996），保留相同的題數、內容、計分方式，修訂的部分有下列六項：（1）修正壓克力色塊的厚度與大小，讓受試者更容易拿取與移動；（2）增加方塊位置圖，將色塊位置標準化，方便施測者使用；（3）調整題目順序，

減少擺放與收拾色塊次數，使施測程序更為流暢；（4）將基準線測試顏色之後「的」刪除，避免受試者因等待「的」之後的名詞而延宕或未做答；（5）將「方塊」修正為「方形」，減少受試者藉由最後一個字猜測的可能性；（6）增加暖身活動，讓受試者適應本測驗施測方式，避免低估受試者能力。

研究者將色塊測驗內容依指令句型分為四部分，分別評量受試者對簡單指令一、二和複雜指令一、二之聽覺理解能力。（1）簡單指令一，題型如：「用手指紅色」、「用手指方形」、「用手指紅色圓形」、「用手指大的方形」，三、四、五歲組分別有 13、11、11 題；（2）簡單指令二，比「簡單指令一」多一些屬性或條件，和「簡單指令一」相同的是只要一個動作即可完成，例如：「用手指小的綠色方形」、「不要指黃色方形，只要指黃色圓形」，三、四、五歲組分別有 7、5、5 題；（3）複雜指令一，題目型式為：「先指……再指……」，例如：「先用手指小的黃色圓形，再指大的綠色方形」，相較於簡單指令，複雜指令一需要記憶兩倍的色塊屬性，以及先後順序，三、四、五歲組均各有 4 題；（4）複雜指令二，題目型式有二：一為「把……放在……的（方位）」，例如：「把小的白色圓形放在大的綠色方形旁邊」，受試者除了需記住兩種色塊的屬性之外，尚須記住方位詞；另一題型為：「左手……，右手……」，例如：「右手拿起黑色方形，左手指黑色圓形」，需要左右兩手做不同的動作，考量三歲組兒童尚未具備左右概念，三歲組沒有這類題型，三、四、五歲組複雜指令二分別有 5、9、9 題。

本測驗以個別施測方式進行，施測者依序逐一念題目給受試者聽，受試者根據聽到的指令操作色塊，題目呈現後 10 秒鐘內正確作答者得 1 分，其餘為 0 分，全對得 29 分，全錯為 0 分，本測驗之信效度詳見本文研究結果。

（二）學前兒童語言能力的測驗

此測驗是張欣戊（1991）所設計，用來評量國內兩歲半到六歲學齡前兒童之國語句法理解能力，全測驗包括六個分測驗，分別針對 2 歲半、3 歲、3 歲半、4 歲、5 歲及 6 歲兒童進行測試，藉由個別施測篩檢語言遲緩之可能個案，施測過程為施測者先述說一個句子，受試者指出四張圖畫中與句意相符的一張。該測驗編製者根據分析結果指出，「學前兒童語言能力的測驗」能有效區辨不同年齡之受試者。本研究使用測驗二（3 歲）、測驗四（4 歲）和測驗五（5 歲）三個分測驗，分別評量常模樣本中每一年齡層至少 100 人，分析語言知識脈絡低的「學前兒童色塊測驗」和強調國語句法理解的「學前兒童語言能力的測驗」之關連性，探討本測驗之效標關連效度。

四、研究程序

（一）編製預試題目、施測材料和施測說明

1.編擬預試題目

前述文獻指出，色塊測驗比其他語言理解測驗能更敏銳地偵測出受試者的輕度理解障礙（Boller & Vignolo, 1966; De Renzi & Vignolo, 1962; DiSimoni, 1978; Shiota et al., 2000），然而，多數版本色塊測驗均包括五個分測驗 61 題，RTT（McNeil & Prescott, 1978）則有十個分測驗 100 題，皆不適合直接翻譯用來施測專注時間較成人短暫的學前兒童。曾進興（1996）編製的「聽覺理解測驗」僅有三個分測驗 29 題，含括了上述各版本色塊測驗之各分測驗題型，但該測驗施測對象最小為五歲兒童，能否直接用來評量五歲以下兒童？為避免難度過高的疑慮，研究者乃增加 10 題較為簡單的題目，到高雄市凱旋國小附設幼稚園、私立聖文幼稚園和冠冠托兒所，施測三、四、五歲組兒童各 10 人，根據幼兒反應調整題目後，敦請兩位兒童語言病理學學者、一位測驗學者及兩位年資 15 年以上的資深語言治療師審查題目

之後，參考審題專家意見修正成為預試題目。為進行跨年代比較，五歲組試題與「聽覺理解測驗」的題數和施測內容相同，僅有部分修正如第 8 頁和第 9 頁所述；四歲組試題和五歲組相同的有 27 題，另以 2 題較簡單的指令取代較複雜的指令；三歲組有 19 題和五歲組相同，另外 10 題則是較為簡單的指令。

2.設計施測材料

本測驗材料包括記錄紙、壓克力色塊和色塊位置圖。測驗記錄紙主要參考曾進興（1996）所編製的「聽覺理解測驗」記錄紙，施測者除了記錄受試者是否正確外，尚須詳細記錄錯誤的部分，包括大小、顏色、形狀、位置、先後順序、是否重述題目等。壓克力色塊的設計，考量三歲兒童對顏色的認知，使用紅、黃、綠、黑、白五種顏色，擔任語言治療師的審題專家之一，就臨床經驗提出幼兒對紅色和黃色可能會混淆，但實際施測並未發現此一混淆現象，因此維持這五種顏色。在壓克力的厚度與大小方面，「聽覺理解測驗」使用的色塊，厚度為 0.2 公分，大的圓形直徑和方形邊長均為 4 公分，小的圓形直徑和方形邊長均為 2 公分，受試者不易拿取與移動，研究者將色塊厚度修正為 0.4 公分，大的圓形直徑和方形邊長為 2.5 公分（1 吋），小的圓形直徑和方形邊長為 1.9 公分（3/4 吋），方便受試者拿取或移動色塊。各版本色塊測驗均在使用說明中列出色塊相對位置圖，施測者根據使用說明排列色塊，研究者考量初次使用者不易記住色塊位置與掌控色塊間距，因此本研究設計了色塊位置圖，以淡黃色西卡紙為底，藉以凸顯五種色塊的顏色。另外，因部分受試者認得顏色名稱國字，導致受測時分心或從中獲得線索，因而以各顏色英文字第一個字母標示。

3.撰寫使用說明

本測驗說明除詳盡敘述施測前、中、後注意事項，以及受試者表現記錄與計分方式外。比較特別的部分是本測驗起初多數題目只要受試者「用手指……」，之後則有「用手拿……」或「把……放在……（位置）」等指令，學前兒童可能一時無法轉換導致誤聽，因而低估其能力，因此安排暖身活動：施測者事先準備一些不同顏色和形狀的糖果（或磁鐵/積木），讓受試者練習「用手指……」、「用手拿……」「把……放在……（位置）」等指令。

（二）預試、試題分析與修正題目

本研究以高雄市和高雄縣 620 名 3 至 5 歲兒童為預試對象，先以古典測驗理論（Classical Test Theory, CTT）進行試題與測驗分析，結果發現三、四、五歲組試題內部一致性 Cronbach's α 係數，分別為.88、.82 和.80，表示各年齡組試題具良好的內部一致性。就各年齡組通過率而言，三、四、五歲組受試者在基準線測試之平均通過率，分別為.95、.98 和.99，標準差均很小（分別為.08、.06 和.02），可說絕大多數預試樣本均能正確作答，即使是三歲組兒童也具備基本的形狀（圓形和方形）和顏色（紅、黃、綠、黑和白）概念。三個年齡層受試者在 19 個共同試題的通過率，一致呈現隨年齡遞增的現象。就受試者答題分析結果來看，三歲組試題有一名三歲半兒童全對，表示該年齡層可增加一兩題較難的題目；四歲和五歲組兒童之整體通過率不低，但無任何受試者完全答對。

為瞭解題目難度與受試者能力的適配情形，並調整三歲組試題，研究者乃將三個年齡層受試者答題資料合併，使用題目反應理論（Item Response Theory, IRT）軟體 BILOGMG 3.0 之單參數 logistic 模式，同時估計題目參數及能力參數。單參數模式假定各題之鑑別度相同，擬似猜測率為 0，僅需估計題目難度，答對題數相同者，其能力估計值也相同。首先進行模式適合度檢定，結果呈現本測驗單參數模式適合度不錯，接著進行題目與能力參數估計，結果顯示以本測驗評估 3 至 5 歲兒童聽覺

理解能力，能力值在-2個標準差至平均數者，其能力估計標準誤平均數在0.3左右，比能力值在平均數以上者低，表示本研究所編製的「學前兒童色塊測驗」，適合用來評估聽覺理解能力在-2個標準差至平均數之學前兒童。另外，考量題目架構，並參考以三歲組為參照之題目難度值，測驗一以2題難度稍高的題目取代難度偏低試題，測驗二也以1題難度稍高的題目取代難度偏低題目，修正後的三歲組試題有22題和五歲組相同，僅有7題是比較簡單的題目；四歲組和五歲組正式測驗題目與預試相同。

（三）選取樣本、施測人員訓練與正式施測

1.選取樣本

本研究依北、中、南、東學齡前兒童人口比率計算樣本數後，從各縣市教育局和社會局網站蒐尋各園所電話住址選取幼托園所，藉由電話聯繫，或透過國小校長（國小附幼）、幼教巡迴輔導教師及中華國際幼兒文教學會理監事，向園所長說明施測程序與目的，徵求園所長同意後，再商請幼托園所教師轉發家長同意書，家長同意施測之幼兒乃成為本研究樣本，為確保有足夠可用的常模樣本，各區取樣人數均比預定的常模樣本數多30至50人。

2.施測人員訓練

本測驗以個別施測方式蒐集資料，所需樣本人數眾多，施測人員除了研究者和助理外，也商請幼保、幼教或聽語所研究生、語言治療師、幼托園所教師協助施測。因不易將施測人員集中，研究者乃配合施測人員的時間，到各園所以小組或個別方式說明施測程序與計分方式，接著由園所隨機選取一名幼兒，研究者現場示範完整施測程序，接受訓練的施測人員和研究者分別紀錄與評分後，逐題核對紀錄，並討論不一致的部分，直到施測人員確實明白施測程序與計分方式為止，並發給每一位施測人員「測驗使用說明」，隨時可以參閱。每場施測說明需1至1.5小時。

3.正式施測

本研究正式施測時間從民國95年11月至96年5月，施測記錄紙收集後，經研究者逐一檢查，淘汰記錄不完整或年齡與記錄紙不相符資料，由研究助理輸入成為SPSS資料，研究者以SPSS描述統計再檢核輸入資料之正確性後，始成為分析的原始資料。

（四）資料分析

除了編製學前兒童色塊測驗外，探討其信效度也是本研究目的之一，信度分析方法有三：（1）使用Cronbach's α係數探討本測驗之內部一致性；（2）以Pearson積差相關分析重測信度；（3）以古典測驗理論為依據，使用公式$\sigma_E = \sigma_X\sqrt{1-\rho_{XX'}}$計算測量標準誤。效度研究的分析方法有五：（1）以結構方程模式（structural equation model, SEM）軟體Amos 6.0進行驗證性因素分析（confirmatory factor analysis, CFA）；（2）以Pearson積差相關分析效標關連效度；（3）使用描述統計和單因子變異數分析，探討不同年齡受試者在共同試題之通過率與得分差異；（4）用重複量數變異數分析探討相同年齡受試者在不同題型之通過率差異；（5）以一個樣本平均數差異z檢定，比較身心障礙樣本和常模樣本在本測驗得分之差異。常模建立之相關研究分析方法有二：（1）使用獨立樣本二因子變異數分析，探討不同性別、居住地區的樣本得分之差異；（2）以兩個樣本平均數差異z檢定，比較本研究五歲組和「聽覺理解測驗」（曾進興，1996）幼稚園大班常模樣本得分差異。

結果與討論

以下分別就信度、效度和常模建立之相關研究三部分，說明本研究結果，並與相關文獻進行討論。

一、信度研究

學前兒童色塊測驗的信度包括內部一致性、重測信度、測量標準誤三種，分析結果摘要如表二，分別說明如下：

（一）內部一致性

本研究以常模樣本分析學前兒童色塊測驗之內部一致性，從表二可看出三、四、五歲組試題之 Cronbach's α 係數，依序為.91、.88 和.86，表示本研究所編製的學前兒童色塊測驗之內部一致性相當良好。

（二）重測信度

本研究分別從三、四、五歲組常模樣本中各取 30 名，間隔兩週重測，三、四、五歲組受試兩次測驗得分之 Pearson 積差相關係數分別為.86、.85 和.84，表示受試者兩次測驗得分一致性相當高，亦即本測驗間隔兩週之重測信度良好。

表二　學前兒童色塊測驗各類信度一覽表

年齡組（標準差）	α 係數	重測信度	測量標準誤	
			原始分數	T 分數
三歲組（σ=6.43）	.91	.86	1.93	3.00
四歲組（σ=5.58）	.88	.85	1.93	3.46
五歲組（σ=5.09）	.86	.84	1.90	3.74

（三）測量標準誤

本研究將各年齡組試題內部一致性信度係數（$\rho_{XX'}$）和常模樣本得分標準差（σ_X），代入公式$\sigma_E=\sigma_X\sqrt{1-\rho_{XX'}}$，計算出三、四、五歲組試題之原始分數測量標準誤，依序為 1.93、1.93 和 1.90；如果以 10 為標準差代入公式，則可計算出 T 分數測量標準誤，如表二所示，三、四、五歲組分別為 3.00、3.46 和 3.74。測量標準誤在使用上是以信賴區間的概念，說明受試真分數可能的散佈範圍，以三歲組受試者為例，原始分數 20 分者，其真分數有 95%的可能性會落在（20±1.96×1.93）之間；T 分數為 40 者，真正的 T 分數有 95%的可能性會落在（40±1.96×3.00）之間。測驗的信度愈高，其測量標準誤愈小，真分數的散佈範圍越小，測量就越準確。一般而言，信度係數介於.80 至.90 之間屬良好，.90 以上則是非常好（McCauley, 2001）。本研究所發展的色塊測驗各年齡組試題內部一致性信度係數均高於.80，其中三歲組還超過.90，根據信度所計算的測量標準誤當然也同樣是可接受的。

二、效度研究

（一）本測驗因素結構之驗證性因素分析

本研究參考各版本色塊測驗和「聽覺理解測驗」（曾進興，1996）編擬測驗題目，用來評量受試者對簡單指令一、二和複雜指令一、二之聽覺理解能力。因單題信度低且得分不符合常態分配，本研究參考 Bandalos 和 Finney（2001，引自吳裕益，2007:1）的建議，將受試者在 9 種題型得分加總成為「群組分數」（parcel scores），做為驗證性因素分析最下層的指標變項（the lowest-order indicator variables），探討這 9 類題目之因素結構，是否符合研究者編製測驗時，預定要評量的 4 類聽覺理解能力。圖一至圖三分別是三、四、五歲組標準化解的因素結構模式圖，由附件一的模式適合度指標摘要表，可看出三、四、五歲組之四因素模式自由度（df），分別比虛無模式（null model）少 14、15 和 15，當自由度為 14 和 15 時，χ^2 值分別為 29.14 和 30.58 即達.01 統計顯著水準，本研究四因素模式 χ^2 值比虛無模式分別少了 1538.2、1416.6 和 1226.8，可說本研究實證資料與四因素模式適合度，顯著優於虛無模式。但五歲組受試者對簡單指令一、二之聽覺理解能力的兩個潛在變項相關為 1，未能符合 Bagozzi 和 Yi（1988，引自吳裕益，2005: 14）

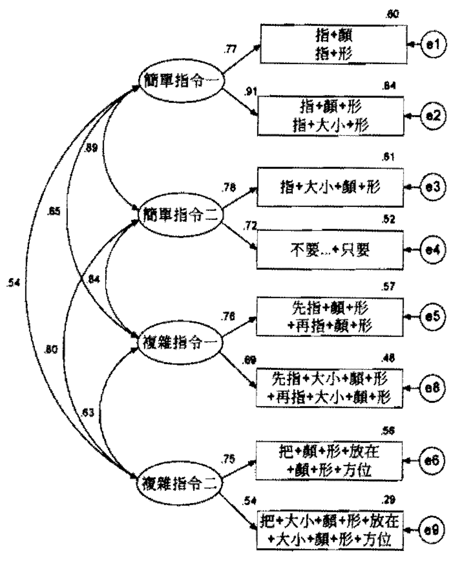
提出的模式基本適配標準，乃將這兩個潛在變項合併，修正為圖四的三因素模式，分析發現三因素模式比虛無模式自由度少 12，χ^2 值少了 1198.0，比自由度 12 時，達.01 統計顯著水準的 26.22 多了許多，表示三因素模式仍顯著優於虛無模式，且符合 Bagozzi 和 Yi 提出的「參數間相關絕對值未太接近 1」之基本適配標準，可說本研究五歲組之實證資料較適合三因素模式。由圖一、圖二和圖四可看出各潛在因素（受試者對簡單指令一、二和複雜指令一、二的聽覺理解能力），對觀察變項（受試者在 9 類題目之表現）的因素負荷量介於.54 至.91 之間，能解釋觀察變項變異量的.29 至.84；各潛在因素間相關介於.54 至.89 之間。四個潛在因素分別代表受試對簡單指令一、二和複雜指令一、二之聽覺理解能力，彼此間有中至高的相關是合理的。

（二）效標關連效度

本研究從三、四、五歲組常模樣本中，分別選取 159、117 和 102 名受試者，施以「學前兒童語言能力的測驗」（張欣戊，1991），使用 Pearson 積差相關探討受試者在本測驗與「學前兒童語言能力的測驗」得分之關連性，結果顯示三、四、五歲組受試者在兩測驗得分之相關依序為.52、.65 和.53（p 均小於.001），屬中度相關。

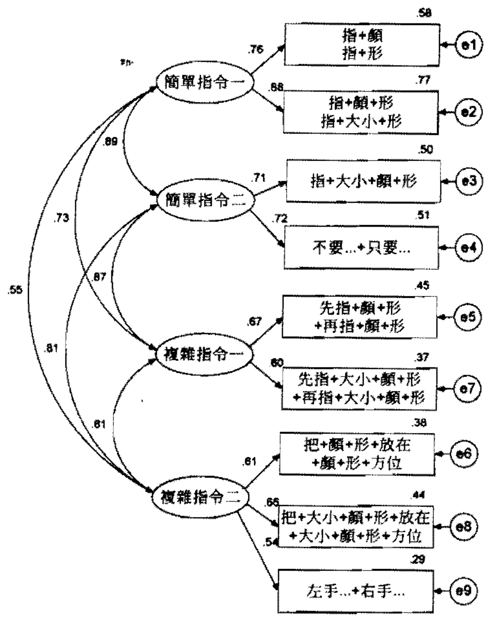
（三）不同年齡組受試者在共同試題得分差異分析

受試者在本測驗之通過率是否隨年齡增長而遞增？由於三個年齡組試題略有不同，僅能就共同試題之通過率進行比較。本研究以單因子變異數分析比較三個年齡組受試，在 22 題共同試題之得分平均數差異，分析結果顯示三組平均數有顯著差異（$F_{(2,1497)}=111.29$，$p<.001$）），關連強度量數（strength of association measures）η^2 為.13，表示年齡差異對受試者在共同試題得分變異的解釋量估 13%，接近大效果量（η^2 之低、中、高值依序為.01、.06 和.14，吳裕益，2003: 8）。接著以 Scheffe 法進行事後比較，結果為五歲組（M=17.46, SD=3.69）顯著高於四歲組（M=14.99, SD=4.37），四歲組又顯著高於三歲組（M=13.37, SD=4.94）。

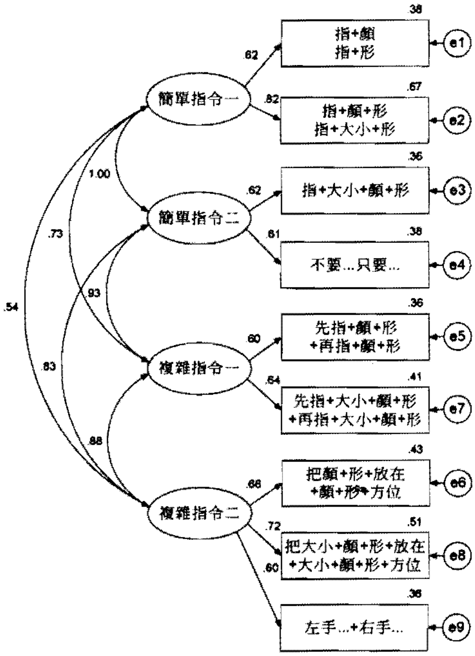


註：1.三歲組沒有「左右」方位詞相關題目。
2.潛在變項表示受試對四類指令之聽覺理解能力。

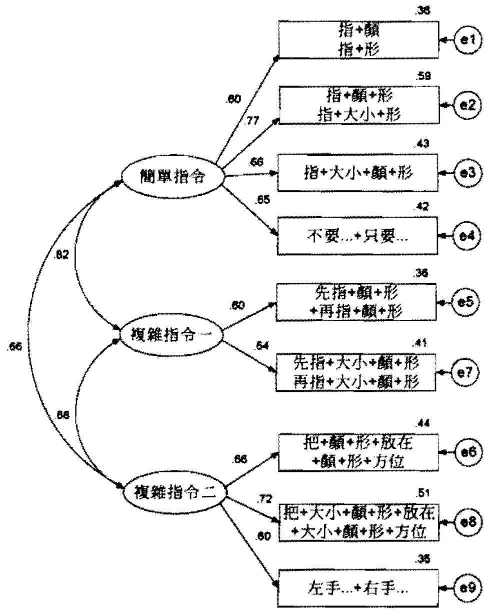
圖一 三歲組四因素結構模式圖



圖二 四歲組四因素結構模式圖



圖三 五歲組四因素結構模式圖



圖四 五歲組三因素結構模式圖

（四）相同年齡組受試者在不同因素試題通過率差異分析

相同年齡之受試者在測量不同潛在能力的試題得分是否有所差異？因各潛在因素試題題數不等，乃將各因素試題得分除以題數轉換成平均通過率後，再以重複量數變異數分析進行比較，由表三可看出相同年齡組受試者在不同因素試題之通過率有顯著差異，η^2 分別為.668、.674 和.672，需進一步比較哪些因素試題間通過率差異達顯著水準。為使多次比較的總 α 值不超過.05，本研究使用 Bonferroni 法調整事後比較之顯著水準，結果顯示三歲組和四歲組，通過率由高至低依序為簡單指令一、簡單指令二、複雜指令一、複雜指令二，彼此間差異均達統計顯著水準。五歲組根據前述驗證性因素分析僅將試題分為簡單指令、複雜指令一和複雜指令二，分析結果和三、四歲組相同，通過率由高至低依序為簡單指令、複雜指令一、複雜指令二，彼此間差異也均達顯著水準。

（五）身心障礙樣本與常模樣本色塊測驗得分差異分析

表四是身心障礙樣本和常模樣本色塊測驗得分平均數和標準差、z 檢定摘要表、及各組樣本在不同因素題目之平均通過率。由表四可看出發展遲緩、智能障礙、自閉症和語言障礙兒童之色塊測驗得分，分別比常模低 4.77、8.56、5.23 和 5.43 分，均顯著低於常模樣本。另外，從常模及身心障礙樣本在各因素試題平均通過率，可看出前述四類身心障礙樣本在各因素試題之通過率均比常模樣本低。研究者分別以 z 檢定比較四類身心障礙樣本和常模樣本在各因素試題平均通過率之差異情形，結果顯示四類身心障礙樣本在三類因素試題的平均通過率都顯著低於常模樣本（p 均小於.001），也就是說三類因素試題均能有效區辨學前一般兒童與發展遲緩、智能障礙、自閉症和語言障礙等身心障礙兒童。

表三　相同年齡組受試者在不同因素試題得分之描述統計與平均數差異分析摘要表

年齡組 試題因素別	平均 通過率	標準 差	*df*	*F*	*p*	η^2	事後比較
三歲組（*N*=500）							
簡單指令一	.82	.24	3,1497	1002.67***	.000	.668	簡單一>簡單二>複雜一>複雜二
簡單指令二	.52	.31					
複雜指令一	.32	.33					
複雜指令二	.17	.23					
四歲組（*N*=500）							
簡單指令一	.90	.19	3,1497	1031.58***	.000	.674	簡單一>簡單二>複雜一>複雜二
簡單指令二	.65	.32					
複雜指令一	.46	.33					
複雜指令二	.25	.22					
五歲組（*N*=500）							
簡單指令	.89	.15	2,998	1023.11***	.000	.672	簡單>複雜一>複雜二
複雜指令一	.59	.32					
複雜指令二	.36	.27					

註：1.***$p < .001$

2.事後比較以 Bonferroni 法調整顯著水準。

表四　身心障礙樣本和常模樣本色塊測驗得分描述統計和 *z* 檢定摘要表

組　別	測驗得分		Z 檢定		各因素試題平均通過率		
	M	*SD*	*z*	*p*	簡單	複雜一	複雜二
常　模(*N*=500)	19.91	5.09			.89	.59	.36
發展遲緩（*N*=69）	15.14	4.84	-7.78***	.000	.79	.37	.12
智能障礙（*N*=20）	11.35	5.84	-7.52***	.000	.61	.16	.10
自 閉 症（*N*=25）	14.68	5.71	-5.14***	.000	.78	.31	.11
語言障礙（*N*=23）	14.48	3.74	-5.12***	.000	.73	.38	.14

***$p < .001$

三、常模建立之相關研究

（一）不同性別、居住地區受試者得分之差異分析

本研究以二因子變異數分析探討常模樣本色塊測驗得分，是否因性別和居住地區而有所差異，以及性別和居住地區兩個因子間是否有交互作用。表五是不同居住地區男女生得分描述統計與二因子變異數分析摘要表，由表可看出性別和居住地區兩因子均無交互作用（三個年齡層 p 都大於.05），因此僅需分別探討性別和居住地區之主要效果。就性別而言，表五資料顯示三歲組無論是都會區或非都會區，女生平均數皆略高於男生，但未達顯著水準；四歲組和五歲組也一致呈現女生平均數高於男生，且達顯著水準，但關連強度量數 η^2 都很

小，分別為.009 和.010，性別不同僅能解釋受試者色塊測驗得分變異的 0.9%至 1%，幾乎可以忽略不管，因此本研究之常模對照表不需分男女。就居住地區來看，三個年齡層一致呈現都會區平均數高於非都會區，且均達顯著水準，三、四、五歲組之 p 值分別為.012、.001 和.000，關連強度量數 η^2 依序為.013、.023 和.075，三歲組和四歲組受試者之居住地區和色塊測驗得分關連度低，不需分別建立常模，但對五歲組受試者而言，不同居住地區能解釋其色塊測驗得分變異的 7.5%，兩者間關連強度屬中等，因此，需建立都會區與非都會區之分組常模對照表。

表五　不同性別、居住地區受試者色塊測驗得分平均數差異二因子變異數分析摘要表

年齡	地區別	性別	描述統計			二因子變異數分析			
			N	M	SD	變異來源	F	p	η^2
三	都會區	男	141	16.73	6.57	性別	2.068	.151	.004
		女	86	17.95	6.15	地區	6.335*	.012	.013
	非都會區	男	151	15.64	6.47	性別*地區	0.421	.517	.001
		女	122	16.10	6.28				
四	都會區	男	128	17.30	5.03	性別	4.357*	.037	.009
		女	115	18.78	5.06	地區	11.761**	.001	.023
	非都會區	男	134	16.06	6.19	性別*地區	0.823	.365	.002
		女	123	16.64	5.60				
五	都會區	男	122	21.10	4.42	性別	4.807*	.029	.010
		女	120	21.70	4.26	地區	40.002***	.000	.075
	非都會區	男	150	17.97	5.46	性別*地區	0.672	.413	.001
		女	108	19.29	5.09				

註：1.都會區指直轄市、省轄市及人口數超過 30 萬的縣轄市，其餘屬非都會區。
2. $*p < .05$, $**p < .01$, $***p < .001$

（二）五歲組常模樣本與「聽覺理解測驗」大班常模樣本得分平均數差異分析

本研究假設受試者之色塊測驗得分，與 11 年前用來建立「聽覺理解測驗」（曾進興，1996）大班常模樣本之平均得分有顯著差異。為進行研究假設檢定，五歲組的施測內容、題數、計分方式與「聽覺理解測驗」相同，僅做部分修正如第 8 頁和第 9 頁所述。由於「聽覺理解測驗」屬「兒童認知功能綜合測驗」中語言測驗的一部份，除了常模外，無論是信效度研究或測驗結果的解釋，均併入「語言測驗」中，因而無法取得單獨以「聽覺理解測驗」所進行的相關分析。在無原始資料情況下，研究者依據該測驗直線轉換常模對照表，計算與本研究五歲組年齡相當的幼稚園大班之平均數及標準差，分別為 20.50 和 6.50，再與本研究五歲組常模樣本平均數（M=19.91，SD=5.09）比較，結果發現相隔 11 年的兩個常模樣本平均數沒有顯著差異（z=1.35，p>.05）。

四、討論

色塊測驗版本相當多，通常有四或五個內容同質的分測驗，及一個內容異質的分測驗，大多數研究以總分或通過率呈現受試者在測

驗上的表現，鮮少探討色塊測驗之因素結構。本研究嘗試依指令之語句結構，將試題分為 9 類（三歲組 8 類），以四因素模式進行驗證性因素分析，結果顯示三歲組和四歲組的模式適合度很好，唯五歲組模式之「簡單指令一」和「簡單指令二」兩個潛在因素相關係數為 1，不符方程結構模式之基本適配標準，因而將這兩個因素合併為「簡單指令」，以三因素模式分析，結果發現三因素模式適合度顯著優於虛無模式，可以說對五歲組受試者而言，學前兒童色塊測驗中用來評量受試者對簡單指令一、二聽覺理解能力的 4 類題目，所測量的特質共同性（communality）高，兩個潛在變項可合併為一--受試者對「簡單指令」之聽覺理解能力。

本研究分析顯示各年齡組受試者之學前兒童色塊測驗得分，和「學前兒童語言能力的測驗」(張欣戊，1991)有中度相關，與 Cartwright 和 Lass(1974, 引自 DiSimoni, 1978: 3)及 Wiig 等人(1977)的研究結果相似，色塊測驗和「西北句法篩檢測驗」得分有中度相關。學前兒童色塊測驗及「學前兒童語言能力的測驗」都是以語句呈現口語刺激，受試者操作色塊或指認圖片作答，評量其聽覺理解能力，色塊測驗主張降低語言知識脈絡，用以評量受試者聽覺理解歷程之注意、記憶和統整能力，「學前兒童語言能力的測驗」則強調句法和語意的理解，兩者間有相同也有相異之處，受試者在兩種測驗之得分有中度相關應是合理的。

本研究結果呈現受試者在共同試題之平均得分隨年齡增加而提高，五歲組顯著高於四歲組，四歲組又顯著高於三歲組，與 Silverman 等人（1977）以學習障礙兒童所進行的研究結果相呼應，年齡和色塊測驗得分間有顯著的正相關（$\gamma=.42, p<.01$），也與 Noll（1970, 引自 Silverman, et al., 1977: 105）的研究結果相同，小學三年級之前，受試者的色塊測驗得分隨年齡遞增，可說本測驗對不同年齡之學前兒童具有相當不錯的區辨力。

研究者根據驗證性因素分析結果，比較同齡受試者在不同因素試題之平均通過率差異，結果顯示三個年齡組受試者在不同因素試題之通過率，呈現相同的組型：受試者對於用來測量「複雜指令二」聽覺理解能力，有上、下、左、右、旁邊等方位詞的試題，通過率最低；需同時記住兩種色塊屬性和先後順序的「複雜指令一」因素試題通過率次低。Spellacy 和 Spreen（1969）的研究結果也指出失語症患者對這兩類題型感到比較困難，但未比較哪一類較困難。本研究根據分析結果推論，「複雜指令二」因素試題通過率最低的原因，可能是受試者的聽覺記憶需同時儲存兩個色塊屬性和方位詞，「複雜指令一」的句型均為「先指……再指……」，受試者只需依色塊屬性出現的順序操作色塊，先後順序則可自動化運作，降低注意、聽覺記憶和統整負荷，處理的正確率也隨之提高；若將用來評量「複雜指令一」潛在因素的題目修正為「指……之前，先指……」和「先指……再指……」隨機出現，受試者所需的注意、聽覺記憶和統整處理負荷增加，其通過率是否依然高於測量「複雜指令二」因素的試題？未來研究可進一步探討。

智能障礙、聽覺障礙、自閉症、特定型語言障礙和發展遲緩等身心障礙兒童，因生理機能和心智功能的限制，其聽覺理解能力通常比一般同齡兒童低（Lahey, 1988; Miller & Fletcher, 2005; Rescorla & Lee, 2000）。本研究三歲和四歲組身心障礙樣本分別僅有 22 和 36 人，五歲組聽覺障礙兒童也只有 6 人，考量樣本數小抽樣誤差大（Agresti & Finlay, 1999），故僅將五歲組智能障礙、自閉症、語言障礙和發展遲緩四類身心障礙樣本之色塊測驗得分平均數，分別與常模樣本進行差異檢定，結果顯示四類身心障礙樣本平均得分都顯著低於

常模樣本。此外，Boller 和 Vignolo（1966）的研究指出內容異質的分測驗五，用來區辨失語症與非失語症者的敏感度高於前四個同質性高的分測驗，因而建議基於實用理由可單獨使用分測驗五。本研究結果發現「簡單指令」、「複雜指令一」和「複雜指令二」三種因素試題，均能有效區辨學前一般兒童與前述四類身心障礙兒童。若就身心障礙樣本和常模樣本在各因素試題平均通過率的差距來看，「複雜指令一」和「複雜指令二」的差距大於「簡單指令」之差距，與 Boller 和 Vignolo 的研究結果相似。

本研究三歲和四歲組身心障礙樣本人數分別僅有 22 和 36 名，係因就讀幼托園所的三、四歲兒童，較少領有身心障礙手冊或持有發展遲緩之醫師診斷證明者，可能是領有手冊的三、四歲兒童，其障礙程度較為嚴重，大多在早療中心接受療育，到了五歲才到幼托園所就讀；障礙程度較輕或發展遲緩者，家長抱持觀望態度等待孩子迎頭趕上，尚未接受評估與診斷。未來若累積較多的三歲和四歲身心障礙樣本人數，可再檢視身心障礙樣本與常模得分之差異組型。

色塊測驗得分是否因性別和居住地區不同而有所差異？由前述文獻得知，男女性之色塊測驗得分並無一致的優劣。本研究蒐集較大樣本（每一年齡組 500 人）進行分析，結果顯示三個年齡組無論是都會區或非都會區，均呈現女生略高於男生的現象，三歲組未達顯著水準，四歲和五歲組女生顯著高於男生，但關連強度量數 η^2 很小，不需建立男女生之分組常模。另外，本研究分析結果也發現，三個年齡組不分男女，都會區受試者之色塊測驗得分平均數均高於非都會區，與探討兒童語言能力是否有城鄉差異之相關研究結果（陸莉、劉鴻香，1994；Bornstein & Cote, 2005）一致，其原因或許如 Williams 等人（2000）的研究所述，都會區的母親比鄉下地區的母親知道如何和孩子說話，比較常和孩子共讀童書與說故事。另外，本研究發現都會區和非都會區樣本之色塊測驗得分差距隨年齡遞增，此現象與張欣戊（1991）發展「學前兒童語言能力的測驗」之研究結果相同，張欣戊將受試者得分之地區差距隨年齡遞增的現象，歸因於北高兩地兒童對國語熟習程度不同，面對「更深入涉及國語句法特殊性」之試題，差異就越明顯。學前兒童色塊測驗三個年齡層有 22 題共同試題，另外 7 題則隨年齡增加而提高試題難度，都會區和非都會區受試者之色塊測驗得分差距隨年齡增長而加大，其原因或許如同張欣戊的推論，都會區兒童無論是學校或家裡，大多以國語為主要的語言，在日常生活中經常聽到較長且複雜的語句；非都會區兒童家中成員的主要語言，有些是閩南語、客家話，或併用些許國語，在生活環境中聽到的國語語句較為簡短。本測驗以國語施測，當指令較為簡短時，不同居住地區的受試者正確率差異不大，指令較長且複雜時，則顯現不同居住地區受試者通過率之差異。

語言是人類生活中重要的溝通工具。為達成溝通目的，人們使用的語詞或語詞所代表的意義，會隨著社會變遷而有所變化。近一、二十年來，學校教育重視美語和鄉土語言的學習，加上科技的進步與傳播媒體的蓬勃發展，社會上出現許多新語詞，也有許多舊語詞被賦予不同的意義。主張降低語言知識脈絡的色塊測驗，受試者之得分是否因時代變遷而有所差異？前述分析結果發現，本研究五歲組常模樣本色塊測驗平均得分，與 11 年前用來建立「聽覺理解測驗」（曾進興，1996）大班常模樣本平均得分沒有顯著差異，與魏氏幼兒智力量表修訂版（WPPSI-R）之「比較研究」（陳榮華、陳心怡，2000）結果不同。可能是本研究與「聽覺理解測驗」常模建立時間相距 11 年，不如 WPPSI 和 WPPSI-R 相隔 22 年之社會變遷大；

也可能是色塊測驗使用顏色、形狀、大小、介係詞、方位詞等不因時代變遷而變化的「封閉詞」（closed class，曾進興譯，2006），組成由簡而繁的指令，測量受試者聽覺理解歷程之注意、聽覺記憶和統整能力，也較不受時代變遷影響所致，本研究結果顯示色塊測驗所建立的常模適用期間較長。

結　論

一、主要研究結果

本研究旨在編製一套 3 至 5 歲兒童適用的色塊測驗，並探討其信效度。主要研究結果有七項：（1）配合學前兒童聽覺理解能力發展，三個年齡組試題內容略有差異，每一年齡組色塊測驗各有 29 題，其中 22 題是共同試題，另外 7 題是各年齡組不同之試題。（2）信度研究結果顯示學前兒童色塊測驗有良好而穩定的內部一致性和重測信度。（3）學前兒童色塊測驗之因素結構，藉由驗證性因素分析發現三歲組和四歲組之實徵資料與四因素模式適配良好，五歲組資料則與三因素模式有較佳的適配度。（4）效標關連效度研究結果顯示，學前兒童色塊測驗與強調句法和語意理解的「學前兒童語言能力的測驗」（張欣戊，1991）有中度相關。（5）三個年齡組受試者在不同因素試題之通過率，均隨指令長度和複雜度之增加而遞減。語句簡短屬「簡單指令一」因素之試題通過率最高，加入大小屬性，一個動作即可完成的「簡單指令二」因素試題次之，需同時記住兩種色塊屬性和先後順序的「複雜指令一」再次之，通過率最低的是指令中含有兩種色塊屬性和方位詞的「複雜指令二」。（6）學前兒童在色塊測驗之表現會因其年齡、性別、居住地區和是否為身心障礙者，而有所差異。受試者在共同試題之得分平均數隨年齡增加而提高，五歲組顯著高於四歲組，四歲組又顯著高於三歲組；各年齡組女生之色塊測驗得分平均數略高於男生，四歲組和五歲組達統計顯著水準；都會區受試者之色塊測驗得分顯著高於非都會區；學前身心障礙兒童之色塊測驗得分顯著低於一般兒童。（7）本研究五歲組常模樣本與聽覺理解測驗（曾進興，1996）幼稚園大班常模樣本之平均數沒有顯著差異，可以說色塊測驗所測量的聽覺理解歷程之注意、聽覺記憶和統整能力，受時代變遷影響不大，所建立的常模適用期間較長。

二、研究限制

本研究因部分園所未登錄或不方便提供家長學歷、職業和收入等資料，探討不同性別和居住地區受試者色塊測驗得分差異時，未將社經地位納入自變項，在應用與解釋上需審慎。另外，本研究三歲和四歲組身心障礙樣本人數分別僅有 22 和 36 名，未進行三歲和四歲組身心障礙樣本和常模樣本得分差異檢定，研究者目前持續蒐集三歲組和四歲組身心障礙樣本施測資料，預定各年齡組智能障礙、聽覺障礙、自閉症、特定型語言障礙和發展遲緩等聽覺理解能力欠佳之身心障礙兒童各 30 人以上，再檢視身心障礙樣本與常模樣本之差異組型。

三、建議

（一）對測驗應用的建議

前述文獻指出色塊測驗可敏銳地偵測受試者之輕微理解缺陷，本研究也發現研究對象在色塊測驗之表現會因為其年齡、居住地區與是否為身心障礙者，而有所差異，若能善加應用，應可及早發現兒童聽覺理解之異常現象，提供早期介入服務，減少因聽覺理解歷程之注意、聽覺記憶和統整能力欠佳所衍生的學習與適應問題。在使用上，研究者建議施測人員應

詳讀使用說明，瞭解施測程序、紀錄與計分方式，方可進行施測。

測驗結果的解釋與應用上，可使用常模對照表將原始分數轉換為百分等級和常態轉換 T 分數，如果受試者之百分等級在 10 以下，家長和幼教老師應多留意他是否經常需要師長重述指令或併用動作手勢才能理解，答案如果是肯定的，在日常生活中應讓受試者學習「注意」聽，藉由練習增加「聽覺記憶廣度」與「統整能力」。此外，本研究根據信度和標準差計算各年齡組原始分數和 T 分數測量標準誤，提供使用者以信賴區間的概念，解釋受試者真分數的散佈範圍。

由於三歲組和四歲組常模樣本在複雜指令一、二的通過率低於 0.5，當受試是身心障礙或疑似發展遲緩兒童時，三歲組只需使用簡單指令一和二，四歲組則使用簡單指令一、二和複雜指令一，並透過等化（equating）研究所得之題目參數，估計其聽覺理解能力值（如下一段所述，本測驗正式推廣使用時會提供施測不同題目時，估計受試者能力值的電腦程式）。

（二）對未來研究的建議

本研究考量學前兒童口語理解能力發展迅速，不同年齡組測驗有共同試題，也有各年齡組特有的試題，未來研究可透過共同試題進行等化研究，分析「學前兒童色塊測驗」所有題目在各年齡組的題目參數與受試者能力估計值，對於聽覺理解能力欠佳之學前兒童，選擇難度較低的題目施測時，可藉由等化研究所估計的題目參數，估計其聽覺理解能力值。也可針對同一群受試者，每隔半年施測一次，進行縱貫研究，瞭解學前兒童聽覺理解歷程之注意、聽覺記憶和統整能力的成長軌線，並追蹤本研究受試者就讀國小階段之學習與生活適應資料，探討學前兒童色塊測驗之預測效度。

學前兒童色塊測驗施測歷程中，施測者需詳細記錄受試者錯誤的部分，本研究限於篇幅，未探討一般與身心障礙兒童之錯誤類型，未來可進一步分析與比較一般及身心障礙兒童的錯誤類型。也可以和醫療院所的語言治療師合作，探討接受語言治療的學前身心障礙兒童在色塊測驗的反應類型，瞭解語言治療師使用色塊測驗評量個案時有哪些不方便需要改進，以及語言治療師使用色塊測驗評量個案之後，如何應用這些資料設計介入方案，探討學前兒童色塊測驗之臨床效度。

色塊測驗主要是用來評量聽覺理解歷程之注意、聽覺記憶和統整能力，文獻及本研究結果均顯示，色塊測驗和其他用來評量聽覺理解能力或聽覺記憶的測驗間有中度相關，這些測驗間的共同性如何？未來研究可嘗試使用色塊測驗、詞彙理解測驗、句法理解測驗，以及修訂魏氏幼兒智力量表的理解測驗、詞彙測驗、句子測驗（評量廣度記憶）等，對同一群受試者施測，進行因素分析探討各類「聽覺理解測驗」之因素結構。

學前兒童色塊測驗藉由施測者口述指令，受試者操作色塊，評量受試者的聽覺理解能力，比較適合用來評量輕度障礙或發展遲緩兒童，未來研究可嘗試為障礙程度較嚴重，無法使用色塊測驗、詞彙理解測驗或句法理解測驗評量者設計情境式理解測驗，提供語言治療師和早療相關人員評量重度障礙兒童之聽覺理解能力。也可以增加指令長度和複雜度，發展學齡兒童適用的色塊測驗，探討學習障礙兒童在聽覺理解歷程之注意、聽覺記憶和統整能力。

參考文獻

巴甫仁（2007）：**台灣版 C-RTT**。私人通訊。

余玻莉（2007）：**國內常用的三種語言評量工具在鑑定語言發展遲緩兒童的信效度之比較研究**。國立高雄師範大學聽力學與語

言治療研究所碩士論文（未出版）。
吳裕益（2003）：**實驗效果之評估**。國立高雄師範大學特殊教育學系（未出版）。
吳裕益（2005）：**線性結構模式的理論與應用**。國立高雄師範大學特殊教育學系（未出版）。
吳裕益（2007）：**驗證性因素分析各分量表題目分群的優缺點**。國立高雄師範大學特殊教育學系（未出版）。
李珊珊（2007）：**國小語障資源班音韻異常兒童構音及語言能力研究**。國立高雄師範大學聽力學與語言治療研究所碩士論文（未出版）。
林寶貴、杞昭安、李芃娟、林玉霞、曹純瓊、黃玉枝、黃桂君、錡寶香、韓福榮（2007）：**修訂學前兒童語言障礙評量表第四階段執行成果報告**。國立臺灣師範大學特殊教育學系。
張欣戊（1991）：**學前兒童語言能力的測驗**。行政院國家科學委員會專題研究計畫成果報告（NSC 79-0301-H-002-067）。國立台灣大學心理系。
陳振宇、謝淑蘭、成戎珠、黃朝慶、洪碧霞、櫻井正二郎、吳裕益、邱上真、陳小娟、曾進興（1996）：**兒童認知功能綜合測驗使用手冊**。教育部社教司。
陳榮華、陳心怡（2000）：**魏氏幼兒智力量表修訂版（中文版）指導手冊**。台北：中國行為科學社。
陳璽琳（2006）：**外籍配偶子女口語能力研究**。樹德科技大學幼兒保育系碩士論文（未出版）。
陸莉、劉鴻香（1994）：**修訂畢保德圖畫詞彙測驗指導手冊**。台北：心理。
曾進興（1996）：**兒童認知功能綜合測驗之七--語言測驗使用手冊**。教育部社教司。
曾進興（譯）（2006）：**兒童語言發展**（R. Cattell 著。Children's Language Consensus and Controversy）。台北：心理。
Agresti, A., & Finlay, B. (1999). *Statistical methods for the social sciences* (3rd Ed.). New Jersey: Prentice Hall.
Arvedson, C. J., McNeil, M. R., & West, T. (1985). Prediction of Revised Token Test overall, subtest, and linguistic unit scores by two shortened versions. *Clinical Aphasiology, 15*, 57-63.
Blosser, J. L., Weidner, W. E., & Dinero, T. (1976). The effect of rate-controlled speech on the auditory receptive scores of children with normal and disordered language abilities. *The Journal of Special Education, 10*, 291-298.
Boller, F., & Vignolo, L. A. (1966). Latent sensory aphasia in hemisphere-damaged patients: an experimental study with the Token Test. *Brain, 89*, 815-830.
Bornstein, M. H. & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. *Infancy, 7*, 299-316.
Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University press.
Cohen-Mimran, R., & Sapir, S. (2007). Deficits in working memory in young adults with reading disabilities. *Journal of Communication Disorders, 40*, 168-183.
D'Arcy, R. C. N., & Connolly, J. F. (1999). An event-related brain potential study of receptive speech comprehension using a modified Token Test. *Neuropsychologia 37*, 1477-1489.
Davis, G. A. (1983). *A survey of adult aphasia*. Englewood Cliffs, New Jersey: Pren-

tice-Hall.

De Renzi, E., & Vignolo, L. A. (1962). The token test: A sensitive test to detect receptive disturbance in aphasia. *Brain, 85*, 665-678.

DiSimoni, F. (1978). *The Token Test for children manual.* Austin, Texas: Pro-ed.

Eberwein, C. A., Pratt, S. R., McNeil, M. R., Fossett, T. R. D., Szuminsky, N. J. & Doyle, P. J. (2007). Auditory performance characteristics of the Computerized Revised Token Test (CRTT). *Journal of Speech, Language, and Hearing Research, 50*, 865-877.

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). Hoboken, New Jersey: John Wiley & Sons.

Hallowell, B., Wertz, R. T., & Kruse, H. (2002). Using eye movement responses to index auditory comprehension: An adaptation of the Revised Token Test. *Aphasiology, 16*, 587–594.

Hula, W., Doyle, P. J., McNeil, M. R., & Mikolic, J. M. (2006). Rasch modeling of Revised Token Test performance: Validity and sensitivity to change. *Journal of Speech, Language, and Hearing Research, 49*, 27–46.

Keegstra, A. L., Knijff, W. A., Post, W. J. & Goorhuis-Brouwer, S. M. (2007). Children with language problems in a speech and hearing clinical: Background variables and extent of language problems. *International Journal of Pediatric Otorhinolarygology, 71*, 815-821.

Keith, R. W., & Engineer, P. (1991). Effects of methylphenidate on the auditory processing abilities of children with attention deficit hyperactivity disorder. *Journal of Learning Disabilities, 24*, 630-636.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *Illinois Test of Psycholinguistic Abilities examiner's manual.* Urbana: University of Illinois Press.

Kurnatowski, P., Putyński, L., Lapienis, M., & Kowalska, B. (2006). Neurocognitive abilities in children with adenotonsillar hypertrophy. *International Journal of Pediatric Otorhinolaryngology, 70*, 419-424.

Lahey, M. (1988). *Language disorders and language development.* New York: Macmillan.

McCauley, R. J. (2001). *Assessment of language disorder in children.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

McNeil, M. R., & Prescott, T. E. (1978). *Revised Token Test.* Austin, Texas: Pro-ed.

Miller, J. F., & Fletcher, P. (2005). Developmental theory and language disorders: Background issues. In P. Fletcher & J. F. Miller (Eds.), *Developmental theory and language disorders* (pp. 1-9). Amsterdam: John Benjamins.

Miller, J. F., & Paul, R. (1995). *The clinical assessment of language comprehension.* Baltimore, Maryland: Paul H. Brookers.

Orgass B., & Poeck, K. (1966). Clinical validation of a new test for aphasia: an experimental study of the Token Test. *Cortex, 2*, 222-243.

Park, G. H., McNeil, M. R., & Tompkins, C. A. (2000). Reliability of the Five-Item Revised Token Test for individuals with aphasia. *Aphasiology, 14*, 527-535.

Paul, R. (1990). Comprehension strategies: World knowledge and the development of sentence comprehension. *Topics in Language Disorders, 10*(3), 63-75.

Rescorla, L. A., & Lee, E. C. (2000). Language impairment in young children. In T. L. Layton, E. R. Crais, & L. R. Watson (Eds.), *Handbook of early language impairment in children: Nature* (pp. 1-38), Albany, NY: Delmar Thomson Learning.

Shiota, M., Koeda, T., & Takeshita, K. (2000). Cognitive and neurophysiological evaluation of Japanese dyslexia. *Brain & Development, 22*, 421-426.

Silverman, I., Raskin, L. M., Davidson, J. L., & Bloom, A. S. (1977). Relationships among Token Test, age, and WISC Score for children with learning problems. *Journal of Learning Disabilities, 10*. 104-107.

Spellacy, F. J., & Spreen, O. (1969). A short form of the Token Test. *Cortex, 5*, 390-397.

Wiig, E. H., Lapointe, C., & Semel, E. M. (1977). Relationships among language processing and production abilities of learning disabled adolescents. *Journal of Learning Disabilities, 10*. 292-299.

Williams, P. D., Soetjiningsih, H., & Williams, A. R. (2000) Balinese mothers' developmental timetables for young children. *Western Journal of Nursing Research, 22*, 717-735.

Yliherva, A., Olsén, P., & Järvelin, M.-R. (2001). Linguistic skills in relation to neurological findings at 8 years of age in children born preterm. *Log Phon Vocol, 26*, 66-75.

附件　學前兒童色塊測驗驗證性因素分析模式適合度指標

年齡組	模式適合度指標								
模式別	x^2	df	p	x^2/df	GFI	AGFI	NFI	IFI	RMSEA
三歲組（N=500）									
四因素模式	20.3	14	.121	1.447	.990	.975	.987	.996	.030
虛無模式	1558.5	28	.000	55.661	.266	.334	.000	.000	.331
四歲組（N=500）									
四因素模式	47.3	21	.001	2.251	.980	.956	.968	.982	.050
虛無模式	1463.9	36	.000	40.665	.455	.319	.000	.000	.282
五歲組（N=500）									
四因素模式	17.4	21	.685	.830	.992	.984	.986	1.000	.000
三因素模式	46.2	24	.004	1.923	.980	.962	.963	.982	.043
虛無模式	1244.2	36	.000	34.560	.360	.391	.000	.000	.259

說明：觀察資料與理論模式之適配與否，通常使用 x^2 檢定，但 x^2 值易受人數影響，當樣本數很大時，很容易達顯著水準，以致輕易拒絕觀察資料與理論模式適配之假定。因此，評鑑模式適合度時尚須參考其他指標，例如：適配度指數（goodness-of-fit index, GFI）、調整後適配度指數（adjusted goodness-of-fit index, AGFI）、正規化適配度指數（normed fit index, NFI）、增值適配度指數（incremental fit index, IFI），這些指數均意指一個理論模式可以解釋實際資料共變數結構的比率，其值愈接近 1，適配度愈佳，大於.90 就可說適配度極佳，由上表可看出三個年齡層的四個模式之模式適合度都高於.90。此外，均方根近似誤差（root mean square error of approximation, RMSEA）是指每個自由度差距量數（measure of discrepancy per degree of freedom），RMSEA 不大於.05 是「適配度良好」（good fit）；.05 至.08 屬於「適配度尚佳」（fair fit），由上表也可看出四個模式與觀察資料之適配度均達「良好」之標準。

Bulletin of Special Education
2007, 32(4), 85-109

Development of the Token Test for Chinese-Speaking Preschoolers

Lin Yueh-Hsien
Assistant Professor, Dept. of Early Childhood Care and Education, Shu-Te University

ABSTRACT

The study was aimed at developing a test for assessing preschool children's auditory comprehension skills based on the Token Test concept originally proposed by De Renzi and Vignolo in the 1960s. This test was especially designed to probe the information processing skills underlining auditory comprehension. This article reviews the procedures in establishing a national norm and also reports its reliability and validity. A total of 2,257 children aged from 3 to 5 were proportionally sampled from the four major geographical regions in Taiwan participated in the study for the use in the pilot phase, norm establishing, and validity checking. The formal test thus developed contains 29 items for each age group, including a common set of 22 items across the three age groups and another set of 7 items unique to each age group. The test was found to have good internal consistency and test-retest reliability. Factor analysis revealed the good fit for the four-factor model with the 3- and 4-year-old groups, yet the three-factor model with the 5-year-old group. As for the criterion-related validity, the test was moderately correlated with an existent auditory comprehension test designed to probe the syntactic and semantic aspects of the language. In addition, the passing rate decreased with the increasing length and complexity of test items. Age, gender, residential area, and disability were critical factors accounting for the test performance. Finally, the overall abilities tapped by the revised test with the current preschooler cohort are not substantially different from the cohort sampled for a previous version almost one decade years ago, indicating that the information processing skills tapped by the test are rather immune to the cultural changes across different generations.

Keywords: Token Test, preschoolers, auditory comprehension